

Lecture Notes in Networks and Systems 729

Pandian Vasant ·

Mohammad Shamsul Arefin ·

Vladimir Panchenko · J. Joshua Thomas ·

Elias Munapo · Gerhard-Wilhelm Weber ·


Roman Rodriguez-Aguilar *Editors*

# Intelligent Computing and Optimization

Proceedings of the 6th International  
Conference on Intelligent Computing  
and Optimization 2023 (ICO2023),  
Volume 1

 Springer

## Series Editor

Janusz Kacprzyk , *Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland*

## Advisory Editors

Fernando Gomide, *Department of Computer Engineering and Automation—DCA, School of Electrical and Computer Engineering—FEEC, University of Campinas—UNICAMP, São Paulo, Brazil*

Okay Kaynak, *Department of Electrical and Electronic Engineering, Bogazici University, Istanbul, Türkiye*

Derong Liu, *Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, USA*

*Institute of Automation, Chinese Academy of Sciences, Beijing, China*

Witold Pedrycz, *Department of Electrical and Computer Engineering, University of Alberta, Alberta, Canada*

*Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland*

Marios M. Polycarpou, *Department of Electrical and Computer Engineering, KIOS Research Center for Intelligent Systems and Networks, University of Cyprus, Nicosia, Cyprus*

Imre J. Rudas, *Óbuda University, Budapest, Hungary*

Jun Wang, *Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong*

The series “Lecture Notes in Networks and Systems” publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

For proposals from Asia please contact Aninda Bose ([aninda.bose@springer.com](mailto:aninda.bose@springer.com)).


Pandian Vasant · Mohammad Shamsul Arefin ·  
Vladimir Panchenko · J. Joshua Thomas ·  
Elias Munapo · Gerhard-Wilhelm Weber ·  
Roman Rodriguez-Aguilar  
Editors

# Intelligent Computing and Optimization

Proceedings of the 6th International Conference  
on Intelligent Computing and Optimization  
2023 (ICO2023), Volume 1

*Editors*

Pandian Vasant  
Faculty of Electrical and Electronics  
Engineering, Modeling Evolutionary  
Algorithms Simulation and Artificial  
Intelligence  
Ton Duc Thang University  
Ho Chi Minh City, Vietnam


Vladimir Panchenko   
Laboratory of Non-traditional Energy  
Systems, Department of Theoretical  
and Applied Mechanics, Federal Scientific  
Agroengineering Center VIM  
Russian University of Transport  
Moscow, Russia

Elias Munapo  
School of Economics and Decision Sciences  
North West University  
Mmabatho, South Africa

Roman Rodriguez-Aguilar  
Facultad de Ciencias Económicas y  
Empresariales, School of Economic  
and Business Sciences  
Universidad Panamericana  
Mexico City, Mexico

Mohammad Shamsul Arefin  
Department of Computer Science  
Chittagong University of Engineering  
and Technology  
Chittagong, Bangladesh

J. Joshua Thomas  
Department of Computer Science  
UOW Malaysia KDU Penang University  
College  
George Town, Malaysia

Gerhard-Wilhelm Weber   
Faculty of Engineering Management  
Poznań University of Technology  
Poznan, Poland

ISSN 2367-3370

ISSN 2367-3389 (electronic)

Lecture Notes in Networks and Systems

ISBN 978-3-031-36245-3

ISBN 978-3-031-36246-0 (eBook)

<https://doi.org/10.1007/978-3-031-36246-0>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature  
Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG  
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

## Preface

The sixth edition of the *International Conference on Intelligent Computing and Optimization (ICO'2023)* was held during April 27–28, 2023, at G Hua Hin Resort and Mall, Hua Hin, Thailand. The objective of the international conference is to bring the global research scholars, experts and scientists in the research areas of intelligent computing and optimization from all over the world to share their knowledge and experiences on the current research achievements in these fields. This conference provides a golden opportunity for global research community to interact and share their novel research results, findings and innovative discoveries among their colleagues and friends. The proceedings of ICO'2023 is published by SPRINGER (in the book series *Lecture Notes in Networks and Systems*) and indexed by SCOPUS.

Almost 70 authors submitted their full papers for the 6th ICO'2023. They represent more than 30 countries, such as Australia, Bangladesh, Bhutan, Botswana, Brazil, Canada, China, Germany, Ghana, Hong Kong, India, Indonesia, Japan, Malaysia, Mauritius, Mexico, Nepal, the Philippines, Russia, Saudi Arabia, South Africa, Sri Lanka, Thailand, Turkey, Ukraine, UK, USA, Vietnam, Zimbabwe and others. This worldwide representation clearly demonstrates the growing interest of the global research community in our conference series. The organizing committee would like to sincerely thank all the authors and the reviewers for their wonderful contribution for this conference. The best and high-quality papers will be selected and reviewed by International Program Committee in order to publish the extended version of the paper in the international indexed journals by SCOPUS and ISI WoS.

This conference could not have been organized without the strong support and help from LNNS SPRINGER NATURE, Easy Chair, IFORS and the Committee of ICO'2023. We would like to sincerely thank *Prof. Roman Rodriguez-Aguilar* (Universidad Panamericana, Mexico) and *Prof. Mohammad Shamsul Arefin* (Daffodil International University, Bangladesh), *Prof. Elias Munapo* (North West University, South Africa) and *Prof. José Antonio Marmolejo Saucedo* (National Autonomous University of Mexico, Mexico) for their great help and support for this conference.

We also appreciate the wonderful guidance and support from *Dr. Sinan Melih Nigdeli* (Istanbul University—Cerrahpaşa, Turkey), *Dr. Marife Rosales* (Polytechnic University of the Philippines, Philippines), *Prof. Rustem Popa* (Dunarea de Jos University, Romania), *Prof. Igor Litvinchev* (Nuevo Leon State University, Mexico), *Dr. Alexander Setiawan* (Petra Christian University, Indonesia), *Dr. Kreangkri Ratchagit* (Maejo University, Thailand), *Dr. Ravindra Boojhawon* (University of Mauritius, Mauritius), *Prof. Mohammed Moshiul Hoque* (CUET, Bangladesh), *Er. Aditya Singh* (Lovely Professional University, India), *Dr. Dmitry Budnikov* (Federal Scientific Agroengineering Center VIM, Russia), *Dr. Deepanjali Shrestha* (Pokhara University, Nepal), *Dr. Nguyen Tan Cam* (University of Information Technology, Vietnam) and *Dr. Thanh Dang Trung* (Thu Dau Mot University, Vietnam). The ICO'2023 committee would like to sincerely thank all the authors, reviewers, **keynote speakers** (*Prof. Roman Rodriguez-Aguilar*,

*Prof. Kaushik Deb, Prof. Rolly Intan, Prof. Francis Miranda, Dr. Deepanjali Shrestha, Prof. Sunarin Chanta*), **plenary speakers** (*Prof. Celso C. Ribeiro, Prof. José Antonio Marmolejo, Dr. Tien Anh Tran*), session chairs and participants for their outstanding contribution to the success of the 6th ICO'2023 in Hua Hin, Thailand.

Finally, we would like to sincerely thank *Prof. Dr. Janusz Kacprzyk, Dr. Thomas Ditzinger, Dr. Holger Schaepe* and *Ms. Varsha Prabakaran* of **LNNS SPRINGER NATURE** for their great support, motivation and encouragement in making this event successful in the global stage.

April 2023

Dr. Pandian Vasant  
(Chair)

Prof. Dr. Gerhard-Wilhelm Weber  
Prof. Dr. Mohammad Shamsul Arefin  
Prof. Dr. Roman Rodriguez-Aguilar  
Dr. Vladimir Panchenko  
Prof. Dr. Elias Munapo  
Dr. J. Joshua Thomas

# Contents

## Digital Transformation, Image Analysis and Sensor Technology

Digitally-Enabled Dynamic Capabilities for Digital Transformation .....	3
<i>Leonardo Pitança Formoso, Gabriel Souza Pinho, and Selma Regina Martins Oliveira</i>	
Digitization of Feeding Processes in Pond Aquaculture Using a Cyber-Physical System for Analyzing Monitoring Data and Transmitting Information Using LoRaWAN Technology .....	17
<i>V. Kostin, N. Sokolova, V. Kochetkov, I. Yudaev, A. Silaev, and R. Borzin</i>	
Energy Efficient Routing Approaches in Wireless Sensor Networks: A Review .....	27
<i>Priyankaben K. Patel and Amrutbhai N. Patel</i>	
Robust Vehicle Speed Estimation Based on Vision Sensor Using YOLOv5 and DeepSORT .....	36
<i>Dea Angelia Kamil, Wahyono, and Agus Harjoko</i>	
Automatic Alignment of Aerospace Images Based on the Search for Characteristic Points .....	47
<i>A. Kolesenkov, B. Kostrov, V. Panchenko, and Yu Daus</i>	
Method for Plant Leaves Square Area Estimation Based on Digital Image Analysis .....	56
<i>Y. Proshkin, A. Smirnov, D. Burynin, and V. Panchenko</i>	
Digital Revolution Through Computational Intelligence: Innovative Applications and Trends .....	66
<i>Ramsagar Yadav, Mukhdeep Singh Manshahia, and M. P. Chaudhary</i>	
Compressive Sensing and Orthogonal Matching Pursuit-Based Approach for Image Compression and Reconstruction .....	73
<i>Sai Sylesh Gupta Namburu, Nandu Vasudevan, Vudhya Muni Sai Karthik, M. Nimal Madhu, and V. Hareesh</i>	
Capabilities for Digital Transformation and Sustainability in an Emerging Economy .....	83
<i>Suzana Franco Juliano, Elizete da Costa Silva de Paula, and Selma Regina Martins Oliveira</i>	



Non-invasive Glucose Measurement with 940 nm Sensor Using Short Wave NIR Technique ..... 95  
*M. Naresh, Samineni Peddakrishna, and M. Thirupathi*

Managing the Purchase-Sale Process of Digital Currencies Under Fuzzy Conditions ..... 104  
*V. Malyukov, B. Bebashko, V. Lakhno, O. Smirnov, I. Malyukova, and H. Mohylnyi*

A Comparative Analysis of the Impacts of Traditional and Digital Billing Methods ..... 113  
*Tasnim Faruki, Rafa Tasnim, Malyha Bintha Mabud, Rashedul Amin Tuhin, Ahmed Wasif Reza, and Mohammad Shamsul Arefin*

Chest X-ray Image Classification Using Convolutional Neural Network to Identify Tuberculosis ..... 127  
*Fahmida Nusrat Promy, Tasnia Afrin Chowdhury, Omar Tawhid Imam, Farhana Alam, Ahmed Wasif Reza, and Mohammad Shamsul Arefin*

Digital Wireless Mini-transduce of Plant Thermoregulation ..... 140  
*A. Grishin, A. Grishin, N. Semenova, V. Grishin, and V. Panchenko*

**Convolution Neural Network, Deep Learning, and Machine Learning**

Effective Fault Prediction Techniques for the Green Cloud Computing Environment Applying Machine Learning to Enhance Network Management ..... 153  
*Hasnath Ahmed Tamim, Md. Sagor Hossain, Md. Asif Uzzaman Asif, Borhan Uddin, Ahmed Wasif Reza, and Mohammad Shamsul Arefin*

Transforming the Financial Industry Through Machine and Deep Learning Innovations ..... 167  
*Sweta Singh, Rahul Bhagat, S. H. Preeti, and G. P. Girish*

MRI-Based Brain Tumor Classification Using Various Deep Learning Convolutional Networks and CNN ..... 177  
*Md. Saiful, Sakib Haider, S. M. Arafat Rahman, Nahid Reza, Ahmed Wasif Reza, and Mohammad Shamsul Arefin*

Deciphering Handwritten Text: A Convolutional Neural Network Framework for Handwritten Character Recognition ..... 189  
*Md Jakir Hossain, Sarah Samiha Zaman, Fardin Rahman Akash, Farhana Alam, Ahmed Wasif Reza, and Mohammad Shamsul Arefin*

Applying Machine Learning Techniques to Forecast Demand in a South African Fast-Moving Consumer Goods Company .....	199
<i>Martin Chanza, Louise De Koker, Sasha Boucher, Elias Munapo, and Gugulethu Mabuza</i>	
A Review on Machine Learning Algorithms for Cost Estimation in Construction Projects .....	209
<i>Vijay Kumar, Sandeep Singla, and Aarti Bansal</i>	
A Computer Assisted Detection Framework of Kidney Diseases Based on CNN Model .....	217
<i>Tanjina Akter Ripa, Nafis Faiyaz, Mahmud Hassan, Rehnuma Naher Sumona, Mohammed Sharafullah Anem, Ahmed Wasif Reza, and Mohammad Shamsul Arefin</i>	
Rice Blast Disease Detection Using CNN Models and DCGAN .....	231
<i>Abdullah Al Munem, Lamyea Tasneem Maha, Rafid Mahmud Haque, Noor Fabi Shah Safa, Mozammel H. A. Khan, and Mohammad Ashik Iqbal Khan</i>	
Evaluation of Performance of Different Machine Learning Techniques for Structural Models .....	243
<i>Melda Yücel, Gebrail Bekdaş, and Sinan Melih Nigdeli</i>	
Age Estimation from Human Facial Expression Using Deep Neural Network .....	252
<i>Md. Ashiqur Rahman, Shuhena Salam Aonty, and Kaushik Deb</i>	
Recognition and Classification of Crop Images by Convolutional Neural Network of Hybrid Architecture .....	263
<i>K. Tokarev, N. Lebed, and I. Yudaev</i>	
A Comparative Study of Deep Learning Algorithms and SARIMA Models for Forecasting Monthly Solar Radiation and UV Index: Case Study for Mauritius .....	273
<i>Janvee Dabedoal, Ravindra Boojhawon, Oomesh Gukhool, and Deepanjali Shrestha</i>	
A Study on Fault Classification and Location Using Supervised Machine Learning .....	284
<i>Nanda Kumari and Channarong Banmongkol</i>	
Deep Learning-Based Time Series Forecasting for CO <sub>2</sub> Emission .....	294
<i>Abhishek Anilkumar, V. Yadukrishnan, M. Nimal Madhu, V. Hareesh, and B. Premjith</i>	

Manila City House Prices: A Machine Learning Analysis of the Current Market Value for Improvements ..... 304  
*Lejan Daniel I. Perdio, Marife A. Rosales, and Robert G. de Luna*

An Analysis of the Relationship Between Temperature Rise and GDP Using Machine Learning Techniques ..... 315  
*Iwan Halim Sahputra, Feren Nathan Wedianto, Novendra Imanuel, Daniel Jaya, and Andre*

Machine Learning Techniques for Predicting Remaining Useful Life (RUL) of Machinery for Sustainable Manufacturing Lines ..... 325  
*Lim Khai Sian and J. Joshua Thomas*

Machine Learning-Based Predictive Modelling of Spot-Welding Process Parameters ..... 337  
*Dinesh V. Burande, Kanak Kalita, and Jasgurpeet Singh Chohan*

**Author Index** ..... 347

## About the Editors



**Pandian Vasant** is Research Associate at Modeling Evolutionary Algorithms Simulation and Artificial Intelligence, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam, and Editor in Chief of *International Journal of Energy Optimization and Engineering* (IJEOE). He holds Ph.D. in Computational Intelligence (UNEM, Costa Rica), M.Sc. (University Malaysia Sabah, Malaysia, Engineering Mathematics) and B.Sc. (Hons, Second Class Upper) in Mathematics (University of Malaya, Malaysia). His research interests include soft computing, hybrid optimization, innovative computing and applications. He has co-authored research articles in journals, conference proceedings, presentations, special issues Guest Editor, chapters (312 publications indexed in Research-Gate) and General Chair of EAI International Conference on Computer Science and Engineering in Penang, Malaysia (2016) and Bangkok, Thailand (2018). In the years 2009 and 2015, he was awarded top reviewer and outstanding reviewer for the journal *Applied Soft Computing* (Elsevier). He has 35 years of working experience at the universities. Currently, Dr. Pandian Vasant is General Chair of the International Conference on Intelligent Computing and Optimization (<https://www.icico.info/>) and Research Associate at Modeling Evolutionary Algorithms Simulation and Artificial Intelligence, Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, HCMC, Vietnam.



**Professor Dr. Mohammad Shamsul Arefin** is in lien from Chittagong University of Engineering and Technology (CUET), Bangladesh and currently affiliated with the Department of Computer Science and Engineering (CSE), Daffodil International University (DIU), Dhaka, Bangladesh. Earlier he was the head of CSE Department, CUET. Prof. Arefin received his Doctor of Engineering Degree in Information Engineering from Hiroshima University, Japan with support of the scholarship of MEXT, Japan. As a part of his doctoral research, Dr. Arefin was with IBM Yamato Software Laboratory, Japan. His research includes data privacy and mining, big data management,

IoT, Cloud Computing, Natural Language processing, Image Information Processing, Social Networks Analysis and Recommendation Systems and IT for agriculture, education and environment. Prof. Arefin is the Editor in Chief of *Computer Science and Engineering Research Journal* (ISSN: 1990-4010) and was the Associate Editor of *BCS Journal of Computer and Information Technology* (ISSN: 2664-4592) and a reviewer as well as TPC member of many international journals and conferences. Dr. Arefin has more than 120 referred publications in international journals, book series and conference proceedings. He delivered more than 30 keynote speeches/invited talks. He also received a good number of research grants/funds from home and abroad. Dr. Arefin is a senior member of IEEE, Member of ACM, Fellow of IEB and BCS. Prof. Arefin involves/earlier involved in many professional activities such as Chairman of Bangladesh Computer Society (BCS) Chittagong Branch; Vice-President (Academic) of BCS National Committee; Executive Committee Member of IEB Computer Engineering Division; Advisor, Bangladesh Robotic Foundation. He was also a member of pre-feasibility study team of CUET IT Business Incubator, first campus based IT Business Incubator in Bangladesh. Prof. Arefin is an Principle Editor of the **Lecture Notes on Data Engineering and Communications Technologies book series** (LNDECT, Volume 95) published by Springer and an editor of the books on *Applied Informatics for Industry 4.0*, *Applied Intelligence for Industry 4.0* and *Computer Vision and Image Analysis for Industry 4.0* to be published Taylor and Francis. Prof. Arefin is the Vice-Chair (Technical) of IEEE CS BDC for the year 2022. He was the Vice-Chair (Activity) of IEEE CS BDC for the year 2021 and the Conference Co-Coordinator of IEEE CS BDC for two consecutive years, 2018 and 2019. He is acting as a TPC Chair of MIET 2022 and the TPC Chair of IEEE Summer Symposium 2022. He was the Organizing Chair of International Conference on Big Data, IoT and Machine Learning (BIM 2021) and National Workshop on Big Data and Machine Learning (BDML 2020). He served as the TPC Chair, International Conference on Electrical, Computer and Communication Engineering (ECCE 2017); Organizing Co-chair, ECCE 2019, Technical Co-chair, IEEE CS BDC Winter Symposium 2020 and Technical Secretary, IEEE CS BDC Winter Symposium 2021. Dr. Arefin helped different international conferences

in the form of track chair, TPC member, reviewer and/or secession chair etc. He is a reviewer of many reputed journals including *IEEE Access*, *Computing Informatics*, *ICT Express*, *Cognitive Computation* etc. Dr. Arefin visited Japan, Indonesia, Malaysia, Bhutan, Singapore, South Korea, Egypt, India, Saudi Arabia and China for different professional and social activities.



**Vladimir Panchenko** is an Associate Professor of the “Department of Theoretical and Applied Mechanics” of the “Russian University of Transport”, Senior Researcher of the “Laboratory of Non-traditional Energy Systems” of the “Federal Scientific Agroengineering Center VIM” and the Teacher of additional education. Graduated from the “Bauman Moscow State Technical University” in 2009 with the qualification of an engineer. Ph.D. thesis of the specialty “Power plants based on renewable energy” was defended in 2013. From 2014 to 2016 Chairman of the Council of Young Scientists and the Member of the Academic Council of the All-Russian Institute for Electrification of Agriculture, Member of the Council of Young Scientists of the Russian University of Transport, Member of the International Solar Energy Society, Individual supporter of Greenpeace and the World Wildlife Fund, Member of the Russian Geographical Society, Member of the Youth section of the Council “Science and Innovations of the Caspian Sea”, Member of the Committee on the use of renewable energy sources of the Russian Union of Scientific and Engineering Public Associations. Diplomas of the winner of the competition of works of young scientists of the All-Russian Scientific Youth School with international participation “Renewable Energy Sources”, Moscow State University M.V. Lomonosov in 2012, 2014, 2018 and 2020, Diploma with a bronze medal of the 15th Russian agro-industrial exhibition “Golden Autumn—2013”, Diploma with a gold medal of the 18th Russian agro-industrial exhibition “Golden Autumn—2016”, Diploma with a silver medal of the XIX Moscow International Salon of Inventions and Innovative technologies “Archimedes—2016”, Diploma for the winning the schoolchildren who have achieved high results in significant events of the Department of Education and Science of the City of Moscow (2020–2021, School No. 2045). Scientific adviser of schoolchildren-winners and prize-winners of the Project and Research Competition “Engineers of the Future” at NUST MISiS 2021 and RTU MIREA 2022.

Invited expert of the projects of the final stages of the “Engineers of the Future” (2021, 2022) and the projects of the “Transport of the Future” (2022, Russian University of Transport). Grant “Young teacher of MIIT” after competitive selection in accordance with the Regulations on grants for young teachers of MIIT (2016–2019). Scholarship of the President of the Russian Federation for 2018–2020 for young scientists and graduate students carrying out promising research and development in priority areas of modernization of the Russian economy. Grant of the Russian Science Foundation 2021 “Conducting fundamental scientific research and exploratory scientific research by international research teams”. Reviewer of articles, chapters and books IGI, Elsevier, Institute of Physics Publishing, *International Journal of Energy Optimization and Engineering*, *Advances in Intelligent Systems and Computing*, *Journal of the Operations Research Society of China*, *Applied Sciences*, *Energies*, *Sustainability*, *AgriEngineering*, *Ain Shams Engineering Journal*, *Concurrency and Computation: Practice and Experience*. Presenter of the sections of the Innovations in Agriculture conference, keynote speaker of the ICO 2019 conference session, key speaker of the special session of the ICO 2020 conference. Assistant Editor since 2019 of the “International Journal of Energy Optimization and Engineering”, Guest Editor since 2019 of the Special Issues of the journal MDPI (Switzerland) “Applied Sciences”, Editor of the book of the “IGI GLOBAL” (USA), as well as book of the “Nova Science Publisher” (USA). Participated in more than 100 exhibitions and conferences of various levels. Published more than 250 scientific papers, including 14 patents, 1 international patent, 6 educational publications, 4 monographs.



**J. Joshua Thomas** is an Associate Professor at UOW Malaysia KDU Penang University College, Malaysia since 2008. He obtained his Ph.D. (Intelligent Systems Techniques) in 2015 from University Sains Malaysia, Penang, and Master’s degree in 1999 from Madurai Kamaraj University, India. From July to September 2005, he worked as a research assistant at the Artificial Intelligence Lab in University Sains Malaysia. From March 2008 to March 2010, he worked as a research associate at the same University. Currently, he is working with Machine Learning, Big Data, Data Analytics, Deep Learning, specially targeting on Convolutional Neural Networks (CNN) and Bi-directional

Recurrent Neural Networks (RNN) for image tagging with embedded natural language processing, End to end steering learning systems and GAN. His work involves experimental research with software prototypes and mathematical modelling and design He is an editorial board member for the *Journal of Energy Optimization and Engineering (IJEQE)*, and invited guest editor for *Journal of Visual Languages Communication (JVLC-Elsevier)*. Recently with *Computer Methods and Programs in Biomedicine (Elsevier)*. He has published more than 40 papers in leading international conference proceedings and peer reviewed journals.



**Elias Munapo** has a Ph.D. obtained in 2010 from the National University of Science and Technology (Zimbabwe) and is a Professor of Operations Research at the North West University, Mafikeng Campus in South Africa. He is a Guest Editor of the *Applied Sciences Journal* and has co-published two books. The first book is titled *Some Innovations in OR Methodology: Linear Optimization* and was published by Lambert Academic publishers in 2018. The second book is titled *Linear Integer Programming: Theory, Applications, and Recent Developments* and was published by De Gruyter publishers in 2021. Professor Munapo has co-edited a number of books, is currently a reviewer of a number of journals, and has published over 100 journal articles and book chapters. In addition, Prof. Munapo is a recipient of the North West University Institutional Research Excellence award and is a member of the Operations Research Society of South Africa (ORSSA), EURO, and IFORS. He has presented at both local and international conferences and has supervised more than 10 doctoral students to completion. His research interests are in the broad area of operations research.



**Gerhard-Wilhelm Weber** is a Professor at Poznań University of Technology, Poznan, Poland, at Faculty of Engineering Management. His research is on mathematics, statistics, operational research, data science, machine learning, finance, economics, optimization, optimal control, management, neuro-, bio- and earth-sciences, medicine, logistics, development, cosmology and generalized space-time research. He is involved in the organization of scientific life internationally. He received Diploma and Doctorate in Mathematics, and Economics/Business Administration, at RWTH Aachen, and Habilitation at TU



Darmstadt (Germany). He replaced Professorships at University of Cologne, and TU Chemnitz, Germany. At Institute of Applied Mathematics, Middle East Technical University, Ankara, Turkey, he was a Professor in Financial Mathematics and Scientific Computing, and Assistant to the Director, and has been a member of five further graduate schools, institutes and departments of METU. G.-W. Weber has affiliations at Universities of Siegen (Germany), Federation University (Ballarat, Australia), University of Aveiro (Portugal), University of North Sumatra (Medan, Indonesia), Malaysia University of Technology, Chinese University of Hong Kong, KTO Karatay University (Konya, Turkey), Vidyasagar University (Midnapore, India), Mazandaran University of Science and Technology (Babol, Iran), Istinye University (Istanbul, Turkey), Georgian International Academy of Sciences, at EURO (Association of European OR Societies) where he is “Advisor to EURO Conferences” and IFORS (International Federation of OR Societies), where he is member in many national OR societies, honorary chair of some EURO working groups, subeditor of IFORS Newsletter, member of IFORS Developing Countries Committee, of Pacific Optimization Research Activity Group, etc. G.-W. Weber has supervised many M.Sc. and Ph.D. students, authored and edited numerous books and articles, and given many presentations from a diversity of areas, in theory, methods and practice. He has been a member of many international editorial, special issue and award boards; he participated at numerous research projects; he received various recognitions by students, universities, conferences and scientific organizations. G.-W. Weber is an IFORS Fellow.



**Roman Rodriguez-Aguilar** is a professor in the School of Economic and Business Sciences of the “Universidad Panamericana” in Mexico. His research is on large-scale mathematical optimization, evolutionary computation, data science, statistical modeling, health economics, energy, competition, and market regulation. He is particularly interested in topics related to artificial intelligence, digital transformation, and Industry 4.0. He received his Ph.D. at the School of Economics at the National Polytechnic Institute, Mexico. He also has a master’s degree in Engineering from the School of Engineering at the National University of Mexico (UNAM), a master’s degree in Administration and

Public Policy in the School of Government and Public Policy at Monterrey Institute of Technology and Higher Education, a postgraduate in applied statistics at the Research Institute in Applied Mathematics and Systems of the UNAM and his degree in Economics at the UNAM. Prior to joining Panamericana University, he has worked as a specialist in economics, statistics, simulation, finance, and optimization, occupying different management positions in various public entities such as the Ministry of Energy, Ministry of Finance, and Ministry of Health. At present, he has the second-highest country-wide distinction granted by the Mexican National System of Research Scientists for scientific merit (SNI Fellow, Level 2). He has co-authored research articles in science citation index journals, conference proceedings, presentations, and book chapters.

# **Digital Transformation, Image Analysis and Sensor Technology**



# Digitally-Enabled Dynamic Capabilities for Digital Transformation

Leonardo Pitança Formoso, Gabriel Souza Pinho,  
and Selma Regina Martins Oliveira<sup>(✉)</sup>

Fluminense Federal University, Volta Redonda, Street, Des. Ellis Hermydio Figueira, 783, Rio  
de Janeiro, RJ 27213-145, Brazil

{leonardopitanca, gabriel\_pinho, selmaregina}@id.uff.br

**Abstract.** The journey to digital transformation has become a strategic imperative on managers' agendas. To realize the benefits associated with a successful adoption of digital transformation, companies need dynamic digitally-enabled capabilities to respond to environmental changes and seize opportunities. While the potential benefits of digital transformation have been widely publicized, little is known about the current cutting-edge literature on how digitally-enabled dynamic capabilities support digital transformation. Primary research-based data was collected from experts at multinational companies to examine how dynamic digitally enabled capabilities contribute to digital transformation in an emerging economy, in this case Brazil. Multivariate analysis techniques were used to analyze the data. The results of our research suggest that dynamic digital capabilities contribute to digital transformation, indicating that reconfiguration and digital capabilities are the most relevant. Our findings fill a gap in the literature and make significant contributions: (i) we offer useful insights to understand the diffusion of dynamic digital capabilities for digital transformation in companies in an emerging economy; (ii) serve as a guide to managers on which capabilities to prioritize in their decision agendas; (iii) advances the body of knowledge on dynamic digitally-enabled capabilities and digital transformation. Furthermore, this study provides insights that can contribute to the current debate on the implementation of digitally-enabled dynamic capabilities in the context of digital transformation around the world.

**Keywords:** Dynamic digitally-enabled capabilities · Digital transformation · Decision-making · Accounting · Companies in an emerging economy

## 1 Introduction

While the potential benefits of dynamic capabilities for digital transformation have been widely reported by academics and practitioners, little is known about how dynamic digital capabilities support digital transformation in organizations. Building on the contemporary debate on the challenges and opportunities surrounding the adoption of dynamic capabilities at the enterprise level, this study tests an original research framework for the adoption of dynamic digital capabilities for digital transformation in an emerging

economy. Understanding how dynamic digital capabilities affect digital transformation in enterprises is a priority for several reasons:

- The prestigious literature [1] suggests that the digital transformation enhances the value proposition for stakeholders and influences the performance of companies [2].
- One of the main reasons for the lack of large-scale digital transformation can be related to the inability of companies to understand how dynamic digital capabilities contribute to digital transformation. Thus, the relationship between digital dynamic capabilities and digital transformation needs to be explored [3]. There is a lack of proposals to examine how dynamic digital capabilities support digital transformation at the firm level [3, 4].
- Dynamic IT-enabled capabilities drive enterprise performance [5] and help companies improve competitive performance.
- Existing studies address developed economies. We intend to shed light for companies in the context of Brazil, an emerging economy in South America, which faces challenges with digital transformation and can reap the benefits of digital transformation.

Existing studies highlight that digital transformation promotes substantial improvements for business [7]; for performance [2, 8] and for the value proposition to stakeholders [1], etc. Thus, digital transformation is on the agenda of managers and leaders as a relevant strategy for organizations [9–12]. For digital transformation to be successful, a combined and progressively aligned integration between digital technologies, resources, business models, structures and processes is required [1, 11, 13]. Thus, studies show that a successful digital transformation process depends on dynamic [1, 10, 11, 13] and digital [1]. We also highlight that previous studies focused mainly on capabilities internal to the organization with little attention to inter-organizational processes and the role of specific capabilities (digital capabilities) [3]. We highlight that, in fast-changing environments, the capabilities to detect and capitalize on changes enable companies to achieve sustainable competitive advantages [14]. Furthermore, we argue that business operations enabled by digital technologies can offer companies interesting opportunities. We believe that dynamic digital capabilities can accelerate digital transformation and respond quickly to environmental changes. Although digital transformation has become an imposing strategy on the agenda of companies [1, 2, 7, 9, 11, 12], little is known about empirical research that examines how digital transformation is supported by dynamic digital capabilities. Despite the strong appeal to the topic, there is still a limited understanding of the contributions of digital dynamic capabilities to digital transformation. In short, there is a growing body of literature that argues for the importance of adopting a dynamic digital capabilities approach to digital transformation [3, 4].

Based on the above assumptions, the aim of this proposal is to investigate how digitally enabled dynamic capabilities support the digital transition in the context of multinational companies from different sectors in Brazil. In this study, the idea of digitally enabled digital capabilities emphasizes the ability to mobilize and deploy capabilities based on digital technologies, combined and integrated with other resources and organizational capabilities aimed at renewing business strategies [6, 15]. In this work, the use of dynamic digital capabilities for digital transformation in multinational companies has the scope to release the full potential of companies to be successful in their digital

transformations. We believe that dynamic digital capabilities can offer opportunities for Brazilian companies that intend to move forward with the digital transformation agenda.

Thus, the question that motivates this article is: how do dynamic digital capabilities support digital transformation (performance) in Brazilian multinational companies? Thus, this article aims to broaden the understanding of Brazilian companies aimed at digital transformation. This empirical research is based on primary data from Brazilian companies from different sectors to answer the above questions. The selection of companies from different sectors for the research application is justified because it can show a broader view of digital dynamic capabilities to drive digital transformation. Studying companies located in Brazil can bring new insights to the debate on dynamic digital capabilities and drive digital transformation in Brazilian companies [16]. We thus offer useful insights to understand the diffusion of dynamic digital capabilities for digital transformation in companies in an emerging economy. Furthermore, this study provides insights that can contribute to the current debate on the implementation of dynamic digital capabilities in the context of digital transformation around the world. In summary, this article intends to fill a gap in the literature and contribute to a better understanding of the role of dynamic digital capabilities for digital transformation; serve as a guide to managers in decisions about priorities for dynamic digital capabilities to make digital transformation more effective; finally, advance in the body of knowledge emerging from the literature on dynamic digital capabilities and digital transformation. After this introduction, the remainder of the article is structured into five additional sections. Section 2 presents the theoretical concepts of the study. Section 3 describes the research methodology used in the study. Section 4 reports the results and discussion. Finally, Sect. 5 highlights the study's contributions and points out its limitations and some proposals for future research.

## 2 Dynamic Capabilities for Digital Transformation

The dominant literature [1, 7, 13] suggests several definitions for digital transformation. There is still no consensus on the best definition for digital transformation. We point out some definitions that support our study. Digital transformation can be defined as:

- The change facilitated by information technologies in organizations, through the digitization of products, services, main processes, customer touchpoints and business models [17].
- Big business improvements promoted by digital technologies (for example, improving the customer experience or creating new business models) [18].
- The use of new digital technologies (social media, mobile, analytics, or embedded) to enable major business improvements (such as enhancing the customer experience, streamlining operations, or creating new business models) [7].
- The use of new digital technologies, such as mobile devices, artificial intelligence, cloud, blockchain and Internet of Things (IoT) technologies, to enable major business improvements to enhance the customer experience, simplify operations or create new business models [13].

- A fundamental change process enabled by digital technologies that aims to bring radical improvements and innovations to an entity (for example, an organization, a business network, an industry or society) to create value for its stakeholders by strategically leveraging its main resources and capabilities [1].

Oliveira and Trento [19] indicate that successful digital transformation must consider dimensions such as digital technologies, people's behavior, leadership, organization, culture, stakeholders, institutions and regulations, and other relevant facilitators that allow improving the proposal and creation of value to customers. Matarazzo et al. [20] highlight organizational transformation, skills improvement and stakeholder involvement as essential dimensions to support digital transformation. In addition, Warner and Wäger [13] argue that digital transformation is a continuous process of using new digital technologies in organizations that allows for great improvements in business results. On the other hand, it is increasingly evident in the literature [11], the dependence of organizations on digital technologies to create new value propositions to satisfy customer needs. Vial [11] highlights that the nature (as a process) of digital transformation makes digital technologies create an impetus for organizations to implement responses and obtain or maintain their competitive advantage.

However, despite the relevance of digital technologies, the process of digital transformation depends on strategies [1, 13] for organizations to capitalize on the innovations of new and unexpected business models that optimize customer needs and experiences [21], explore opportunities and defend against threats arising from digital technologies [12]. Drawing on the dominant literature [1, 11, 13], we argue that the benefits of digital transformation can only be achieved through the lens of digital technologies, dynamic and digital resources and capabilities, processes, structures, etc. in an integrated way to adapt to environmental turmoil and take advantage of business opportunities. In this way, we argue that dynamic digital capabilities oriented towards innovation can be a facilitator of this process of digital transformation to respond quickly to environmental changes [3, 4, 14]. We believe that digitally enabled dynamic capabilities can facilitate companies to perceive the context of the environment and seize opportunities, manage threats, reconfigure the business model (and reallocate resources) in a context of digital transformation [6, 14, 22]. Digitization incorporates digital technologies with the aim of creating innovations of new business models, increasing revenues and creating value for customers [23].

That is, digitization transforms organizations, creates new value for customers and enhances value co-creation through advanced service offerings [24]. Thus, digital capabilities defined as “organizational capabilities that favor companies to combine digital assets and business resources, in a comprehensive way, make it possible to take advantage of digital networks aimed at innovation (products, processes and services), with an emphasis on organizational learning, customer-oriented value creation and innovation, with a focus on sustainable competitive advantage [3, 25]. Despite the relevance of digital capabilities for organizations to achieve future competitiveness, the full potential of digitization capabilities has not yet been fully exploited. Based on the above argument, we argue that one of the main reasons for the lack of large-scale digital transformation may be related to the inability of companies to identify, prioritize and implement dynamic digital capabilities to reap the benefits of this digital transition process. In this spectrum,

Warner and Wäger [13] highlighted in their study, digital detection capabilities (digital sense, e.g. technological trends, interpretation of digital scenarios, formulation of digital strategies, scenario planning, digital mindset, etc.), with an emphasis on digital reconnaissance and digital landscape planning as essential to quickly understand unexpected trends in rapidly changing environments. Strategic planning is an essential tool to better equip yourself for a digital age. Additionally, the authors highlight the digital mindset as another relevant digital detection capability.

In a digital context, Warner and Wäger [13] indicate that apprehension capacity (digital seizing, e.g. Strategic agility, balance of digital portfolios, etc.) is mainly supported by strategic agility, considered essential for innovation continuation of the business model. Finally, digital transformation/reconfiguration capability (digital transforming, e.g. redesigning structures, improving digital maturity, etc.) involves the continual strategic renewal of an organization's collaborative approach and, eventually, culture. Furthermore, digital maturity is a key dynamic capability for ongoing digital transformations [13]. In short, digital transformation to be successful requires a vibrant environment of specific dynamic capabilities (such as digital) of companies, technological aspects; etc. and essentially the reconfiguration of the existing organization, in terms of skills, business model and routines to adapt to technological changes for digital disruption [4, 5, 26, 27]. That is, to be successful, digital transformation depends on the development of dynamic digital capabilities as suggested in the editorial agenda proposed by [5]. Based on the premises above, our study seeks to fill this gap in the literature and answer how dynamic digital capabilities contribute to digital transformation (performance).

### 3 Methodology

To collect the survey data, a scalar-type questionnaire (Likert) (Appendix A) was prepared based on the literature review [1, 10, 13, 26, 28], in two parts: I—general information about the respondents; II—information to examine how dynamic digital capabilities affect digital transformation. First of all, content and structure validation was carried out with two respondents, one from the information technology area and the other from the management area. Few adjustments (such as shorter sentences) were made for final refinement. In addition, a pilot test was also carried out with a controller from a multinational company. Internal consistency was checked using Cronbach's alpha coefficient. The test value indicated 0.88, which is considered an almost perfect value [29]. The influence of digital dynamic capabilities on digital transformation performance was measured using a six-point Likert scale in relation to the level of influence exerted by each digital dynamic capability (1 = strongly disagree and 5 = strongly agree). The variables adopted to measure digital dynamic capabilities are based on studies by [1, 10, 13, 26, 28]. The performance of the digital transformation was based on the proposal by [10]. The research was directed to multinational companies from different sectors (Auditing, Financial, Construction and building materials, Manufacturing/electronics, machinery and equipment, Oil & Gas, and Others) in Brazil that are digitally enabled or that are in digital transition. The profile of respondents defined for this survey are professionals oriented to management, controlling, auditing, etc.: controllers, auditors, coordinators, managers, directors, etc. The professional social network LinkedIn was



used to map the respondents. 200 questionnaires were sent using the Google Forms platform. We received 20 completed questionnaires. Based on this sample, it is not possible to generalize the sample results to a broader population, even so our sample was representative and allowed us to validate our conclusions. Our results were analyzed using the multivariate techniques of descriptive statistics. Thus, Fig. 1a and b show the results of the respondents' profiles.

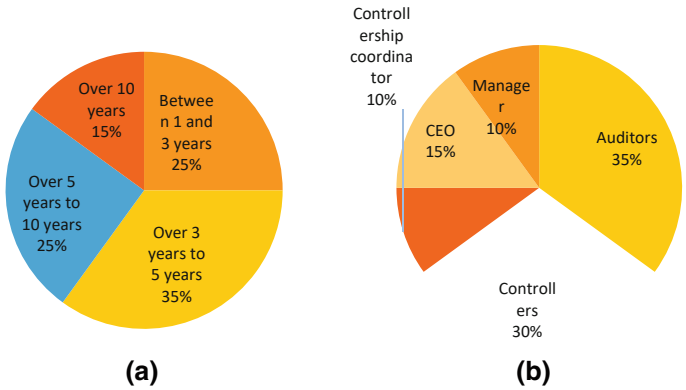


Fig. 1. a Experience in office. b occupations

The results shown in Fig. 1a and b suggest that respondents are auditors (35%), controllers (30%), controllership coordinator (10%), CEO (15%) and manager (10%); with more than 3 years of experience (75%).

### 4 Results and Analysis

Using descriptive statistics techniques, we present the results of the means (Mean) and standard deviation (SD) by categories of digital dynamic capabilities, with the integration of all sectors (Fig. 2 and Appendix A).

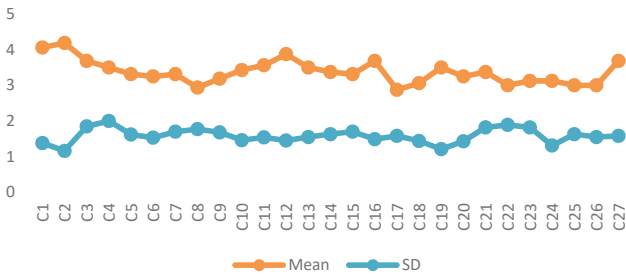


Fig. 2. Robustness test results—mean and standard deviation (SD)

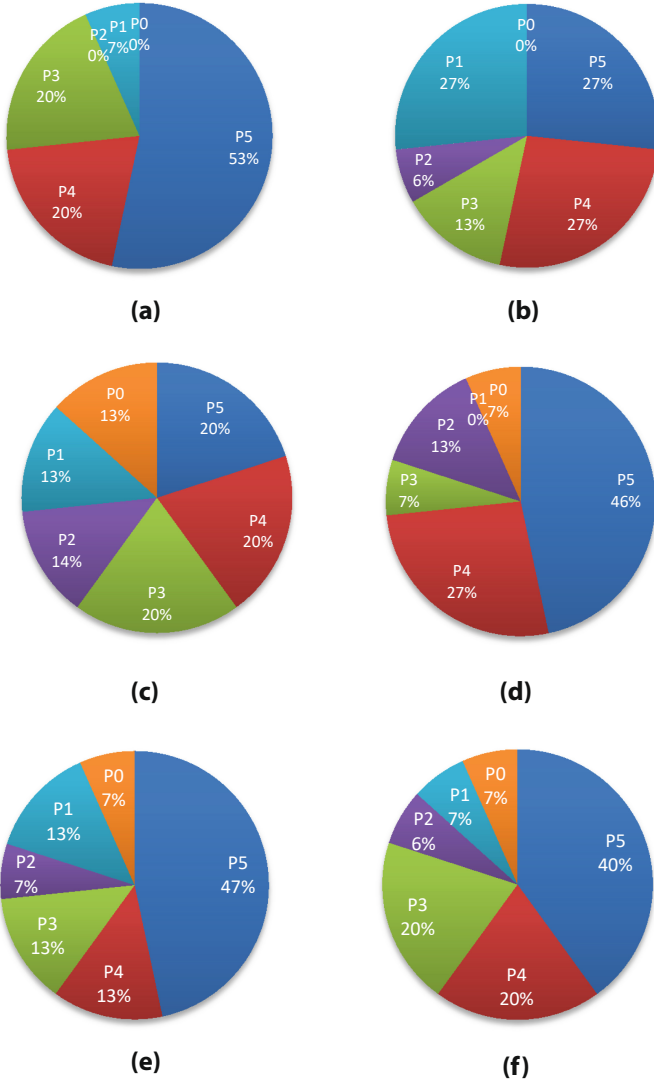
The results highlighted in Fig. 2 (and Appendix A) show that companies' digital transformation is moderately affected by digital transformation capabilities (M = 3.43).

On average ( $M = 3.86$ ), the ability to reconfigure has greater prominence for digital transformation, with emphasis on the ability to adapt to a data-driven culture (C1) and adapt to new digital technologies (big data, blockchain, IoT, etc.) (C2). Financial resource capabilities (C5), digital technologies (C6), and infrastructure for digital technologies (C7) are the most prominent investment capabilities ( $M = 3.29$ ) for business resilience. Customer-driven strategic transformation (C10) and customer-driven long-term digital vision (C11) capabilities are the digital/data-driven strategic planning capabilities ( $M = 3.28$ ) most relevant to business resilience. Interestingly, the ability ( $M = 3.38$ ) to formulate digital strategies (C8) was below average. The digital capabilities with the greatest relevance for digital transformation are trusted data-driven decisions (C12) and the ability to implement digital technologies (C16). What is intriguing is that customer-driven digital planning capability (C17) is below average. Also, human resources have moderate relevance for business resilience ( $M = 3.10$ ), with emphasis on the ability of technical teams to implement digital technology projects (C21). Finally, the ability to capture value (profit, return on investments, etc.) (C27) through new digital technologies is substantially strong ( $M = 3.68$ ). On the other hand, the values found for the standard deviation are greater than 1 for all capabilities, with emphasis on C3 (resource reallocation) ( $SD = 1.85$ ), C4 (expertise for successful digital transformation projects) (2.0) and C22 (work team with digital maturity) ( $SD = 1.89$ ), who presented SD close to 2 (Appendix A). We highlight that the means of influence of capabilities in relation to the manufacturing sector was less than 2 (see Appendix A and B). The results presented in Fig. 3a–d indicate the intensity of responses by levels (P0, P1, P2, P3, P4 and P5)—Likert scale.

The results indicate a greater intensity of responses at levels 4 and 5 (Fig. 3a–f), with greater prominence for reconfiguration (Fig. 3a) and digital (Fig. 3d) capacities (73%). Intriguing are the results indicated for the strategic planning capacity (3c), with intensity around 40% for levels 4 and 5. A joint analysis indicates that the analyzed sectors manage to capture value through digital technologies, with 60% of responses concentrated in levels 4 and 5. The intensity of responses by sector is presented in Appendix B.

## 5 Discussion and Conclusion

The unique contribution of this article lies in testing empirical evidence to relate dynamic digital capabilities and digital transformation, adding the results of this relationship in the context of an emerging economy. While the potential benefits of dynamic capabilities are widely publicized, little is known from the authoritative literature about how digital dynamic capabilities affect digital transformation in an emerging economy. A survey was directed at multinational companies from different sectors in Brazil. Previous studies have addressed the dynamic capabilities for digital transformation in developed markets. This study sheds light on an emerging economy in South America. We provide interesting insights that can contribute to the implementation of dynamic digital capabilities in companies and/or sectors that are in the process of digital transformation around the world. The conclusions of this study provide results and shed light on the field of knowledge about dynamic digital capabilities and digital transformation.



**Fig. 3.** **a** Reconfiguration. **b** Investment. **c** Strategic planning. **d** Digital. **e** Human resources. **f** Capturing value

Through an empirical study, we fill a gap in the literature on the current state of how dynamic digital capabilities affect digital transformation in multinational companies from different sectors in Brazil. In summary, our findings add to the findings of the existing literature by shedding light on linking dynamic digital capabilities and digital transformation in companies in emerging markets. Whereas digital transformation has become a strategic imperative for organizations to reconfigure their business models and create value for their customers and other stakeholders [1, 10, 13, 26], this study

provides insights for companies looking to reap the rewards of digital transformation. Cutting-edge literature argues that companies can be successful in their digital transformation projects when they are supported by specific technological capabilities [5, 27], organizational reconfiguration, resources human, financial, and digital knowledge and capabilities, to be able to adapt to technological changes aimed at digital disruption [1, 26]. Previous studies have focused on dynamic capabilities with little attention to digital dynamic capabilities [3, 4]. Thus, understanding the prominence of dynamic capabilities for digital transformation in organizations can be useful for companies in emerging economies in Latin America that intend to move forward with digital transformation. Based on these assumptions, the results of our empirical research suggest that dynamic digital capabilities affect digital transformation ( $M = 3.43$ ). This finding is in line with the prestigious literature [10, 13, 26].

Our findings indicate that reconfiguration capabilities to adapt to a data-driven culture and new digital technologies are the most relevant ( $M = 3.86$ ) for digital transformation in multinational companies. To summarize, although this research confirms the expected results that dynamic capabilities influence digital transformation in organizations, it also suggests unexpected and original insights: namely, that strategic planning capabilities have less relevance for digital transformation, which is unusual in the cutting-edge literature [13, 30], which suggests that organizations need to build increasingly digitized detection capabilities [13]. Thus, building digital landscape planning is essential to quickly understand unexpected trends in rapidly changing environments. Strategic planning is a protagonist in the digital age [31, 32]. Drawing an analogy with [28], we call on managers to focus their efforts on building strong dynamic digital capabilities to remain relevant in the context of an emerging digital economy [28]. The findings highlighted above have implications for both theory and practice.

## 5.1 Implications for Theory

The results of this study indicated that the digital dynamic capabilities (sensing, seizing and reconfiguration) influence the digital transformation in the analyzed sample companies. Thus, our findings advance the field of knowledge about dynamic digital capabilities and digital transformation in companies from different sectors in an emerging economy. Thus, our study extends the existing literature and suggests that digital dynamic capabilities have become prominent for digital transformation [3, 4]. We suggest that companies in the digital transition manage the development of strategic partnerships to leverage dynamic digital capabilities and reap the benefits of the digital transition journey. We highlighted at the beginning that some substantial studies suggest the positive effect of dynamic digital capabilities on outcomes [10, 13, 33]. Our findings advance the literature in the field in which dynamic digital capabilities drive organizational resilience, particularly in companies in the audit and finance sectors (Appendix B) of this sample.

## 5.2 Implications for Practice

The results of this study have implications for practice. Our results provide critical insights into how dynamic digital capabilities affect digital transformation. Thus, our study serves as a guide for managers to create business value from dynamic digital

capabilities. Therefore, we suggest that managers develop dynamic digital capabilities to make companies more successful in their digital transformations, in addition to significantly facilitating digital innovation in products and processes. In other words, for companies to reap good results from their digital transitions largely depends on digital reconfiguration capabilities; digital resources; digital strategic planning; and digital/technological capabilities. As companies develop and implement dynamic digital capabilities into their organizational processes and business routines, digital innovation can be achieved; as well as competitive advantage.

### 5.3 Limitations and Suggestions for Future Research

Although our study has substantial implications for theory and practice, it is not without limitations. This study is one of the first attempts to examine the effects of dynamic digital capabilities for digital transformation in an emerging South American economy—in this case, Brazil. Therefore, the findings of this research should be considered relevant to the sample of companies considered in this study and should not be extended beyond this limit. Furthermore, the methodology used in this study was based on a cross-sectional study, whose collected data are from a given moment. We also highlight that the sample is limited in terms of size. Future surveys should substantially increase the sample size and expand to other sectors and respondents. We analyze several sectors in an integrated way. Different results could emerge from the individual analysis of the surveyed sectors. Future studies could focus on one or another sector of this sample or simply expand to other segments and in other countries. Our study was directed to Brazilian companies. We suggest that future studies be directed to samples from other countries of emerging economies and mature economies for the purpose of comparing results, which could add value to the debate on the adoption of digital dynamic capabilities in digital transformation environments in companies from emerging economies.

**Acknowledgments.** Acknowledgment to the Fluminense Federal University.

## Appendix A: Robustness Test

See Table 1.

**Table 1.** Robustness test—digital dynamic capabilities

Digital dynamic capabilities LEGEND		Mean	SD
C1: Does our company have the ability to adapt to a data-driven culture?	Capacity for reconfiguration/transformation	4.06	1.38
C2: Does our company have the ability to adapt to new digital technologies (big data, blockchain, IoT, etc.)?		4.18	1.16
C3: Does our company have the ability to quickly redesign internal structures in the face of unexpected changes (rapid reallocation of resources)?		3.68	1.85
C4: Does our company have the digital expertise to be successful in digital transformation?		3.5	2.0
C5: Does our company invest financial resources in projects aimed at digital technologies?	Resources/investment capacity	3.31	1.62
C6: Does our company invest in human resources oriented towards digital/data technologies?		3.25	1.53
C7: Does our company invest in infrastructure for digital technologies?		3.31	1.7
C8: Does our company formulate digital strategies?	Capability for strategic planning	2.93	1.77
C9: Does our company have the capacity to interpret the scenario of the digital future?		3.18	1.68
C10: Does our company have the capacity for customer-oriented strategic transformation?		3.43	1.46
C11: Does our company have a long-term digital vision?		3.56	1.54
C12: Are our decisions based on reliable data?	Digital capabilities/technologies	3.87	1.45
C13: Does our company have the capacity to identify digital technological trends?		3.50	1.55
C14: Does our company have access to external sources of digital technologies?		3.37	1.63

*(continued)*

**Table 1.** (continued)

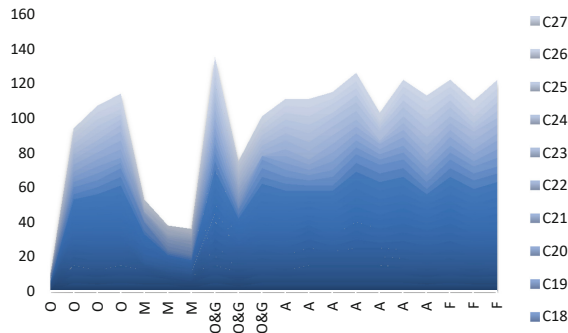
Digital dynamic capabilities LEGEND		Mean	SD
C15: Do we have the capacity to integrate multiple external data sources for analysis?		3.31	1.7
C16: Does our company have the capacity to implement digital technologies?		3.68	1.49
C17: Does our company prepare customer-oriented digital scenario planning?		2.87	1.58
C18: Does our company invest in customer-oriented digital innovation?		3.06	1.44
C19: Does our company have large-scale data-driven digital technologies?		3.50	1.21
C20: Does our company invest in customer-oriented digital partnerships/collaborations?		3.25	1.43
C21: Do we have teams with the technical skills needed to implement digital transformation/digital technology projects?	Human resource capacity	3.37	1.82
C22: Does our work team (task force) have digital maturity?		3.0	1.89
C23: We have a digital technology driven culture/ Does our company have a digital technology driven mindset?		3.12	1.82
C24: Do we have leadership that encourages the search for external knowledge and digital technologies?		3.12	1.31
C25: Does our company have the capacity to identify a digital task force?		3.0	1.63
C26: Does our company invest in leaders to improve digital strategies and capture potential business opportunities?		3.0	1.55
C27: Does our company have the capacity to capture value with new digital technologies (profit, return on investments, etc.?)	Cap. to capture value with digital tech	3.6	1.58

N = 20

Cronbach Alpha: 0.8863

## Appendix B

See Fig. 4.



**Fig. 4.** Intensity of responses by sector: auditing (a), financial (f), manufacturing (M), oil & gas (O&G), and others (O)

## References







1. Gong, C., Ribiere, V.: Developing a unified definition of digital transformation. *Technovation* **102**, 102217 (2021)
2. Oliveira, S.R.M., Trento, S.: How digital transformation capabilities influence firm performance? A methodology for supporting decision-making. In: *IEEE 15th International Conference on Application of Information and Communication Technologies*, pp. 1–4. AICT (2021)
3. Annarelli, A., Battistella, C., Nonino, F., Parida, V., Pessot, E.: Literature review on digitalization capabilities: co-citation analysis of antecedents, conceptualization and consequences. *Technol. Forecast. Social Change* **166**, 120635 (2021)
4. Mele, G., Passiante, G., Secundo, G., Capaldo, G., Corvello, V.: Digital Dynamic Capabilities for Digital Transformation of Organisations for Sustainability, pp. 7–9. IFKAD Conference (2023)
5. Battistella, C., Toni, A.F.D., Zan, G.D., Pessot, E.: Cultivating business model agility through focused capabilities: a multiple case study. *J. Bus. Res.* **73**, 65–82 (2017)
6. Mikalef, P., Pateli, A.: Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: findings from PLS-SEM and fsQCA. *J. Bus. Res.* **70**, 1–16 (2017)
7. Fitzgerald, M., Kruschwitz, N., Bonnet, D., Welch, M.: Embracing digital technology: a new strategic imperative. *MIT Sloan Manage. Rev.* **55**(2) (2014)
8. Ismail, M.H., Khater, M., Zaki, M.: Digital business transformation and strategy: what do we know so far. *Cambridge Serv. Alliance* **10**(1), 1–35 (2017)
9. Hess, T., Matt, C., Benlian, A., Wiesböck, F.: Options for formulating a digital transformation strategy. *MIS Q. Exec.* **15**(2), 123–139 (2016)
10. Souza-Zomer, T.T., Neely, A., Martinez, V.: Digital transforming capability and performance: a microfoundational perspective. *Int. J. Oper. Prod. Manage.* **40**(7/8), 1095–1128 (2020)



11. Vial, G.: Understanding digital transformation: a review and a research agenda. *J. Strateg. Inf. Syst.* **28**(2), 118–144 (2019)
12. Singh, A., Hess, T.: How chief digital officers promote the digital transformation of their companies. *MIS Q. Exec.* **16**(1), 1–17 (2017)
13. Warner, K.S.R., Wäger, M.: Building dynamic capabilities for digital transformation: an ongoing process of strategic renewal. *Long Range Plan.* **52**(3), 326–349 (2019)
14. Teece, D.J.: Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance. *Strat. Manage. J.* **28**(13), 1319–1350 (2007)
15. Bharadwaj, A.S.: A resource-based perspective on information technology capability and firm performance: an empirical investigation. *MIS Quart.* **24**(1), 169–196 (2000)
16. OECD.: *A Caminho da Era Digital no Brasil*. OECD Publishing (2020)
17. Hartl, E., Hess, T.: *The Role of Cultural Values for Digital Transformation: Insights from a Delphi Study*. Americas Conference on Information Systems, Boston (2017)
18. Piccinini, E., Hanelt, A., Gregory, R.W., Kolve, L.: *Transforming Industrial Business: The Impact of Digital Transformation on Automotive Organizations*. International Conference of Information Systems, Fort Worth (2015)
19. Oliveira, S.R.M., Trento, S.: Innovation ecosystem and digital transformation: linking two emerging agendas through knowledge. In: *IEEE 15th International Conference on Application of Information and Communication Technologies (AICT)*, pp. 1–6 (2021)
20. Matarazzo, M., Penco, L., Giorgia, P., Roberto, Q.: Digital transformation and customer value creation in made in Italy SMEs: a dynamic capabilities perspective. *J. Bus. Res.* **123**, 642–656 (2021)
21. Rogers, D.: *The Digital Transformation Playbook: Rethink Your Business for the Digital Age*. Columbia University Press, New York (2016)
22. Teece, D.J.: *Dynamic Capabilities and Strategic Management: Organizing for Innovation and Growth*, pp. 509–533. Oxford University Press, Oxford (2009)
23. Parida, V., Sjödin, D., Reim, W.: Leveraging digitalization for advanced service business models: reflections from a systematic literature review and special issue contributions. *Sustainability* **11**(2), 391 (2019)
24. Porter, M.E., Heppelmann, J.E.: How smart, connected products are transforming competition. *Harvard Bus. Rev.* **92**, 64–88 (2014)
25. Annarelli, A., Colabianchi, S., Nonino, F., Palombi, G.: *Digital Transformation and Organizational Resilience: Managing Knowledge to Nurture Capabilities*. IFKAD Conference (2023)
26. Ellström, D., Holtström, J., Berg, E., Josefsson, C.: Dynamic capabilities for digital transformation. *J. Strateg. Manag.* **15**(2), 272–286 (2022)
27. Eller, R., Alford, P., Kallmuenzer, A., Peters, M.: Antecedents consequences, and challenges of small and medium-sized enterprise digitalization. *J. Bus. Res.* **112**, 119–127 (2020)
28. Teece, D.J.: Business models and dynamic capabilities. *Long Range Plan.* **51**(1), 40–49 (2018)
29. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977)
30. Nambisan, S., Iyminen, K., Majchrzak, A., Song, M.: Digital innovation management: reinventing innovation management research in a digital world. *MIS Quart.* **41**(1) (2017)
31. Dong, A., Garbuio, M., Lovallo, D.: Generative sensing: a design perspective on the microfoundations of sensing capabilities. *Calif. Manage. Rev.* **58**(4), 97–117 (2016)
32. Matt, C., Hess, T., Benlian, A.: Digital transformation strategies. *Bus. Inf. Syst. Eng.* **57**(5), 339–343 (2015). <https://doi.org/10.1007/s12599-015-0401-5>
33. Shen, L., Zhang, X., Liu, H.: Digital technology adoption, digital dynamic capability, and digital transformation performance of textile industry: moderating role of digital innovation orientation. *Manag. Decis. Econ.* **43**(6), 2038–2054 (2021)



# Digitization of Feeding Processes in Pond Aquaculture Using a Cyber-Physical System for Analyzing Monitoring Data and Transmitting Information Using LoRaWAN Technology

V. Kostin<sup>1</sup> , N. Sokolova<sup>1</sup> , V. Kochetkov<sup>1</sup> , I. Yudaev<sup>2</sup>  , A. Silaev<sup>1</sup> ,  
and R. Borzin<sup>1</sup>

<sup>1</sup> Volzhsky Polytechnic Institute (Branch) of Volgograd State Technical University, 42a Engelsa St., Volzhskiy, Volgograd Region 404101, Russia

kochetkov@post.volpi.ru

<sup>2</sup> Kuban State Agrarian University, Kalinina St. 13, 350044 Krasnodar, Russia

zirochka2505@gmail.com, etsh1965@mail.ru

**Abstract.** For the effective development of fish in pond farms, it is necessary to observe comfortable conditions for the life and growth of aquaculture, which include such indicators of water quality as temperature, pH, concentration of dissolved oxygen. Accounting for all these parameters in manual mode is a very time-consuming process and carries the risks of the human factor. The purpose of the work is the use of systems for automatic monitoring. These systems allow online monitoring of water parameters and, in case of deviations, quickly take appropriate measures to normalize them. The objectives of the study are the development of a hardware-software complex for monitoring water parameters using LoRaWAN wireless technology; development of software for wireless transmission of parameters in pond farming. The main research method is a full-scale experiment, which consists in determining the water parameters and transmitting data from a transmitter installed on a watercraft to a receiver using LoRaWAN wireless data transmission technology. The result of the work of the system under study is an increase in economic efficiency by reducing the loss of fish and feed resources. During experiments on the basis of the Priboy fish farm, the effectiveness of the presented solution based on the LoRaWAN wireless data transmission technology was revealed. The introduction of a mobile complex designed for automated environmental monitoring of water bodies in the future will make it possible to receive accurate, reliable and complete information on the state of water bodies in a timely manner, to determine the occurrence of adverse environmental situations in time and be able to predict their consequences.

**Keywords:** Aquaculture · Pond farming · Automatic monitoring of water parameters · LoRaWAN wireless technology · Hardware and software complex

## 1 Introduction

In pond aquaculture, one of the most important tasks is to increase the efficiency of feeding through the rational use of feed, taking into account the maximum possible number of parameters in real time. To do this, it is necessary to monitor the quality of water in the ponds of the fish farm, quickly transfer and process the information received in order to form feeding standards. The main indicators of water quality that directly affect the formation of feeding norms are water temperature and the concentration of oxygen dissolved in it. Also, an important factor influencing the decision-making on the technological operations of water deoxidation is the pH value [1].

It is known that an increase in the water temperature in ponds to a certain value (depending on the composition of aquaculture) contributes to an increase in the intensity of metabolic processes, as well as an increase in appetite and, consequently, growth rates. An increase in temperature above the optimal value for a given aquaculture has a negative impact, namely, inhibition of the physiological state and a decrease in fish appetite, which is determined by the deterioration of the hydrochemical state of the ponds. The current methods for measuring water temperature in ponds recommend measuring at 12–13 o'clock at the outlet structure at a depth of 0.5–0.8 m. When the water temperature is close to or exceeds the optimal value for this aquaculture, measurements are taken daily [2].

In the main period of feeding, which is characterized by the accumulation of significant amounts of organic matter, the first distribution of feed is made no earlier than 2–3 h after sunrise. The main condition for this mode is the oxygen content not lower than 2 mg/l. If there is a steady decrease in oxygen concentration, and its readings in the morning are less than 2 mg/l, the first feeding is done at 10–11 am. Feeding is not recommended before sunset when the oxygen regime is tense and there is a risk of starvation. The level of oxygen saturation in fish ponds must be carried out in the morning at feeding places, which is technically practically impossible to implement with the existing organization of technological processes in fish farms.

For hydrobionts, an important characteristic of a reservoir is the acidity of the medium, which is characterized by the pH value (concentration of hydrogen ions). The pH is expressed in dimensionless units from 1 to 14. The best environment for the development and growth of aquaculture is a neutral or slightly alkaline water reaction (pH equal to 7 or slightly higher). The pH during the day can vary within 2–3 units [3]. During the mass development of algae and higher aquatic vegetation, plants consume all free carbon dioxide from the water during the day, and by evening its concentration drops to almost zero. With an insufficient amount of carbon dioxide in the water, the water becomes alkaline, during the night its concentration increases again due to the respiration of all aquatic organisms, including plants. The pH level of water in fish ponds must be measured twice a day—in the morning and in the evening.

For real-time monitoring of water parameters, transmission and processing of information necessary for calculating feeding norms, the development of a cyber-physical system for wireless transmission and analysis of data is an urgent task [4, 5]. Currently, there are many varieties of wireless data transmission technologies that are actively used to organize systems for collecting and processing data for Industry 4.0 [6, 7]. LoRaWAN

technology seems to be the most suitable wireless data transmission technology available. The advantages of this technology are high range (up to 10 km), ultra-low power consumption, and the ability to seamlessly develop own stack. The structure of this technology includes end device, gateways, network server and application server [8].

The gateway and end devices organize a star-type network topology. As a rule, this device includes multichannel transceivers for signal processing in several channels at the same time, or several signals in one channel [9].

## 2 Data and Methods

A prototype of a cyber-physical system for analyzing water monitoring data was developed on the Arduino Nano platform with SX1276 transceivers equipped with a LoRa Long Range modem [10, 11]. Three measurements of the studied parameters were carried out at three points of the pond.

The temperature sensor is implemented on the DS18B20 chip. It detects the ambient temperature in the range from  $-55$  to  $+125$  °C and transmits data as a digital signal with 12-bit resolution using the 1-Wire protocol, which allows you to connect a large number of such sensors using just one controller digital port. In this case, the so-called “parasitic power” is used, in which the sensor receives energy directly from the signal line. Each sensor has a unique factory-coded 64-bit code that can be used by the microcontroller to communicate with a specific sensor on a common bus. The code of each sensor is read by a separate command.

To determine dissolved oxygen, the DFRobot gravity analog sensor is used, compatible with Arduino, ESP8266, ESP32, STM32 microcontrollers and Raspberry Pi.

To measure the pH, a pH meter is used, which consists of two parts: a remote probe for measurements and a sensor in the form of a Troyka module, which is designed to filter and amplify the signal.

Readings from temperature, pH and dissolved oxygen sensors in mV are converted into °C, pH and mg/l values, respectively, for which the sensor calibration function was performed, with specified conversion factors.

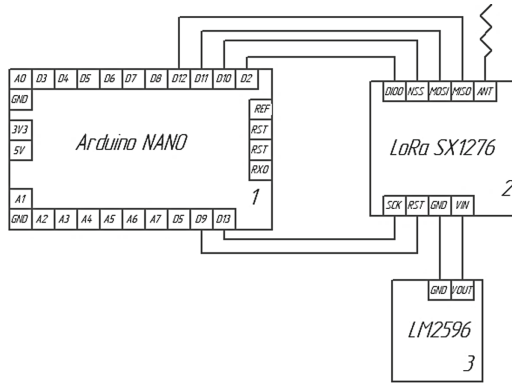
Figure 1 shows the circuit diagram for connecting the LoRa SX1276 module to the Arduino Nano microcontroller, which operates in receiver mode [10–12].

Figure 2 shows the circuit diagram of connecting the LoRa SX1276 module to the Arduino Nano controller, which operates in the transmitter mode and is used in the monitoring system of water parameters sensors [10–12].

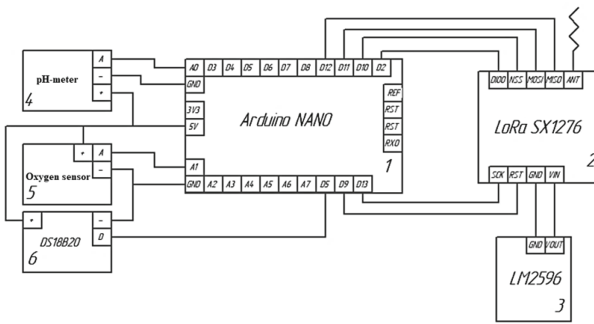
The block diagram of the water parameters monitoring system, shown in Fig. 3, consists of two devices.

The first device is installed directly on the pond with three sensors connected to it. The second device is installed directly at the fisheries management point.

In order to transmit data using LoRa technology using Arduino, you need to use the LoRa library. The library provides for the restoration of a connection in emergency situations, the creation of a pair between the sender and the recipient, automatic routing of the created network [4, 9].



**Fig. 1.** Receiver connection diagram: 1—Arduino Nano controller; 2—LoRa SX1276 transmitter receiving module; 3—voltage converter LM2596



**Fig. 2.** Transmitter connection diagram: 1—Arduino Nano controller; 2—LoRa SX1276 transmitter receiving module; 3—voltage converter LM2596; 4—pH-meter (Troyka-module); 5—DFRobot dissolved oxygen sensor; 6—temperature sensor DS18B20

Block diagrams of the developed algorithms for transmitting and receiving messages are shown in Figs. 4 and 5.

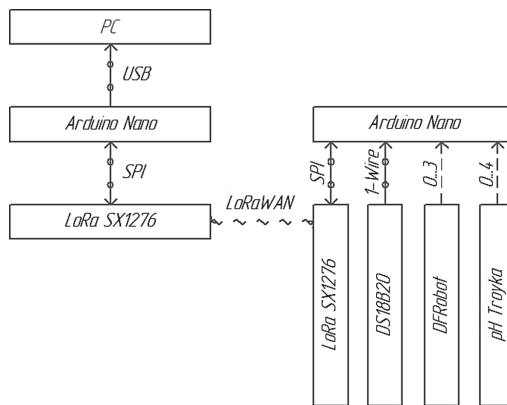
After readings are received by the receiving device, they are transferred to a PC for further processing by connecting to the COM port. Then a successful connection is checked. If the connection is successful, the message is split into parameters and displayed on the screen, after which recommendations are selected in a loop. The number of passes in a cycle directly depends on the number of ponds under supervision.

After the successful selection of recommendations and entering them into the array, they are displayed on the screen, and then saved to the database for the possibility of analyzing the state of water over a long period. The block diagram of this algorithm is shown in Fig. 6. To store and view statistics, a database was created that contains one table with the following fields:

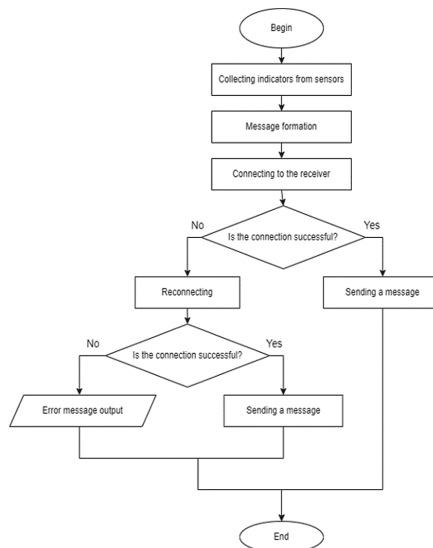
- ID;
- date and time of measurement;

- pond area;
- stocking density;
- average weight of fish;
- temperature;
- pH;
- oxygen content.

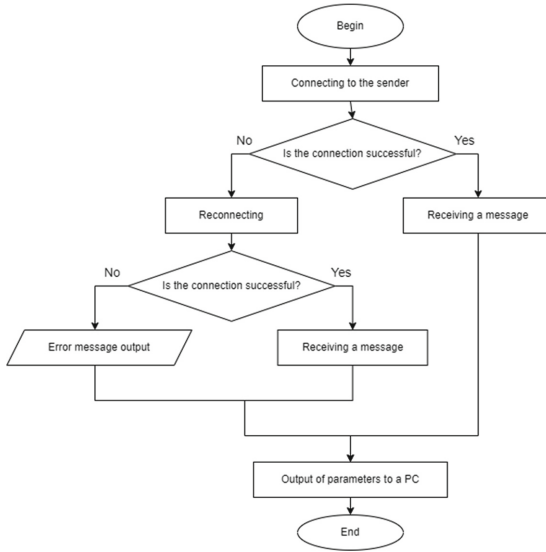
The program starts its execution after connecting the receiver to the PC and the sender to the power source. This algorithm allows you to transmit and receive data from sensors through a wireless communication device.



**Fig. 3.** Block diagram of the water parameters monitoring system



**Fig. 4.** Block diagram of the message-passing algorithm



**Fig. 5.** Block diagram of the algorithm for receiving a message

### 3 Results and Discussion

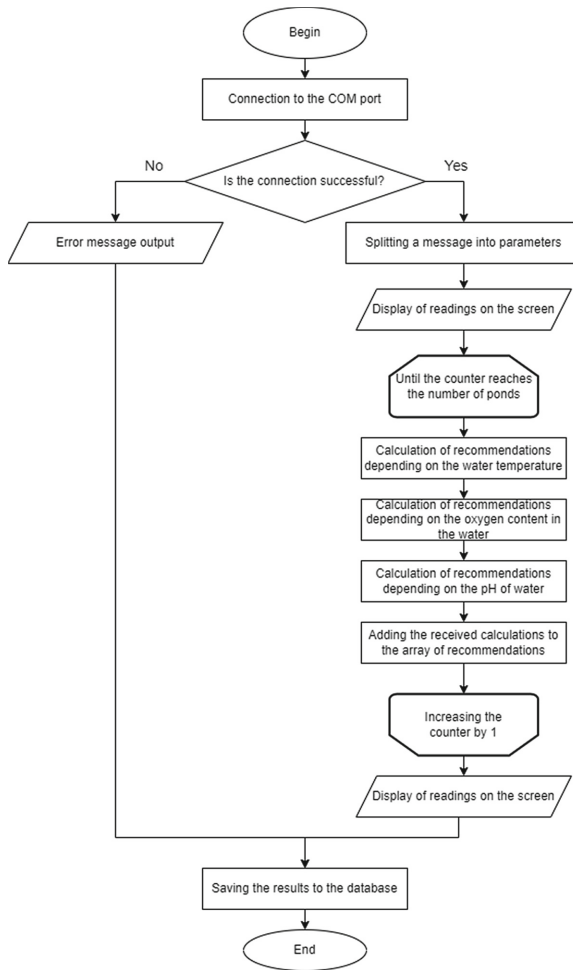
To test the data transfer rate, three measurements of the studied parameters were carried out at three points of the pond. According to the principle of the receiver and the transmitter, two devices were assembled, on which, in test mode, data packets of 17 bytes in size were sent at a distance of 1.5, 5, 10 km with fixing the time of sending and receiving messages.

A production experiment on the transfer of monitoring data of water parameters to a remote computer was carried out on June 1, 2022 on the basis of the Priboy fish farm located on the left bank of the Volgograd reservoir. During the experiment, the transmitting device was installed in one of the fish ponds, and the receiving device in the administrative building of the farm, the distance between the devices was 257 m (Fig. 7). According to the results of the experiment, the average waiting time for the transmission of readings and the average error in the readings of the sensors were established.

The AMTAST AMT08 thermooximeter was used to measure the reference readings of temperature and the amount of dissolved oxygen. The results of the experiment are shown in Table 1. Thus, the error of the sensors corresponds to the declared ones, and the data transmission is correct and uninterrupted.

For the effective functioning of the fish farm and the rational use of feed, prompt correction of feeding rates is necessary depending on the actual water temperature, oxygen concentration and the level of water acidity.

When implementing a cyber-physical system for analyzing data from monitoring water parameters and wirelessly transmitting data to a remote PC in a pond farm, it is necessary to develop an automated system for managing feeding rates based on specially developed software. This software analyzes the input values and received monitoring data to generate recommendations for fish feeding and the need for deacidification process



**Fig. 6.** Block diagram of the PC parameter output function

steps. Information for the user is displayed on the screen in the form of a table, which displays the main parameters and recommendations for each pond, the date and time of measurement, as well as the values of the controlled parameters themselves (Fig. 8).

Thus, with the help of the developed software, it becomes possible to monitor ponds in real time, as well as view the history of measurements and recommendations for feeding, depending on the current state of the water.

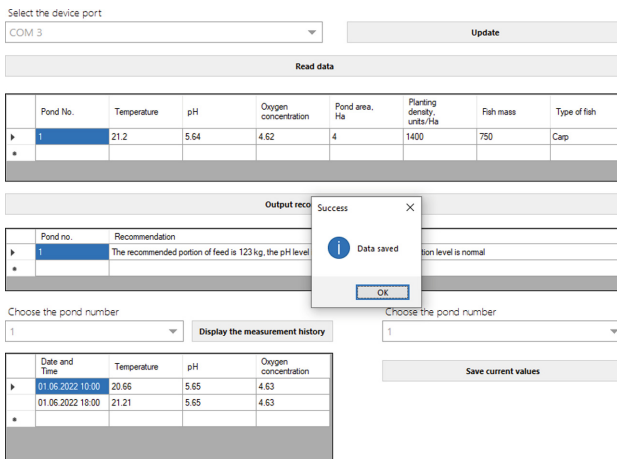




**Fig. 7.** Location of the receiver and transmitter on the map

**Table 1.** Results of the experiments

Parameter	Reference	1 sample	2 sample	3 sample	Accuracy%
Waiting time, sec		2.2	2	2.1	
Temperature level, °C	22.3	22.8	21.8	22.1	± 0.5
pH level	6.88	6.98	6.95	6.76	± 0.1
Dissolved oxygen level, mg/l	4.29	4.19	4.34	4.39	± 0.1



**Fig. 8.** The result of the program

## 4 Conclusion

A system for monitoring the parameters of pond water parameters based on the LoRaWAN wireless technology has been investigated and modeled. The LoRaWAN wireless transmission technology is considered, the principle of its operation is described. Algorithms have been developed to study the principles of data transmission over long distances. The developed system reduces the likelihood of the human factor to a minimum, reduces the loss of fish and feed resources, thus bringing economic benefits. The conducted experiments confirm the correct operation of the water monitoring system using LoRaWAN technology. Digitalization makes it possible to expand the opportunities of the agricultural sector, including agricultural production, to increase the efficiency of resource use. With its help, the efficiency of agricultural production is increased due to the optimal planning of the crops' structure, automation of plants' irrigation and extra-feed feeding, digital modeling of crop yields, feed ration optimization of farm animals [13–19].

## References

1. Averin, P.I., Danilov, M., Burdakova, N.V., Kashkovskaya, V.P., Sharaviev, P.V.: Technology of carp reproduction in pond fisheries. *Youth Sci.* **10**, 26 (2016)
2. Ivanov, A.P.: *Fish Farming in Natural Reservoirs*. Agropromizdat, Moscow, USSR (1988)
3. Petin, A.N., Lebedeva, M.G., Krymskaya, O.V.: *Analysis and Assessment of Surface Water Quality Textbook*. Publishing House of BelGU, Belgorod, Russian Federation (2006)
4. Pekshin, K.: *Communications in IoT*. <https://future2day.ru/>. Accessed 31 Jan 2023
5. Gorbachenko, I.M.: Modeling the process of packet data transmission. *Sci. Rev. Tech. Sci.* **1**, 121 (2014)
6. Haka, A.M., Aleksieva, V.P., Valchanov, H.G.: Deployment and analysis of Bluetooth low energy network. *IOP Conf. Ser. Mater. Sci. Eng.* **1032**, 012016 (2021)
7. Proshin, A.A., Goryachev, N.V., Yurkov, N.K., et al.: Wireless network WI-FI. *Innov. Inform. Commun. Technol.* **1**, 452–455 (2018)
8. Augustin, A., Yi, J., Clausen, T.H., et al.: A study of LoRa: Long range & low power networks for the internet of things. *Sensors* **16**, 1466 (2016)
9. Codeluppi, G., Cilfone, A., Davoli, L., Ferrari, G.: LORAFARM: a LoRaWAN-based smart farming modular IOT architecture. *Sensors* **20**(7), 2028 (2020)
10. Julie, M.: McGee SX1272/3/6/7/8: LoRamodem low energy consumption design AN1200.17. <https://www.semtech.com/products/wireless-rf/lora-core/sx1272#resources>. Accessed 31 Jan 2023
11. Marc, P.: SX1272/3/6/7/8: LoRa modem designer's guide AN1200.13 TCo. <https://www.semtech.com/products/wireless-rf/lora-core>. Accessed 31 Jan 2023
12. Sornin, N., Luis, M., Eirich, T., Kramp, T., Hersent, O.: LoRaWAN™ Specification, V1.0 (2015)
13. Petrukhin, V., Feklistov, A., Yudaev, I., et al.: Modeling of the device operating principle for electrical stimulation of grafting establishment of woody plants. *Lecture Notes Netw. Syst.* **569**, 667–673 (2023)
14. Yudaev, I.V.: Analysis of variation in circuit parameters for substitution of weed plant tissue under electric impulse action. *Surf. Eng. Appl. Electrochem.* **55**(2), 219–224 (2019)
15. Tokarev, K., Lebed, N., Prokofiev, P., et al.: Monitoring and intelligent management of agrophytocenosis productivity based on deep neural network algorithms. *Lecture Notes Netw. Syst.* **569**, 686–694 (2023)

16. Yudaev, I., Eviev, V., Sumyanova, E., et al.: Methodology and modeling of the application of electrophysical methods for locust pest control. *Lecture Notes Netw. Syst.* **569**, 781–788 (2023)
17. Baev, V.I., Yudaev, I.V., Petrukhin, V.A., et al.: Electrotechnology as one of the most advanced branches in the agricultural production development. In: *Handbook of Research on Renewable Energy and Electric Resources for Sustainable Rural Development*. IGI Global, Hershey, PA (2018)
18. Daus, Y.V., Kharchenko, V.V., Yudaev, I.V., et al.: Improving the efficiency of the power supply to agricultural facilities by means of roof-top photovoltaic installations. *Appl. Solar Energy* **56**(3), 207–211 (2020)
19. Yudaev, I.V., Daus, Y.V., Zharkov, A.V., Zharkov, V.Y.: Private solar power plants of ukraine of small capacity: features of exploitation and operating experience. *Appl. Solar Energy* **56**(1), 54–62 (2020)



# Energy Efficient Routing Approaches in Wireless Sensor Networks: A Review

Priyankaben K. Patel<sup>(✉)</sup> and Amrutbhai N. Patel

Ganpat University, Kherva, Mehsana, Gujarat, India  
{pkp02, amrut.patel}@ganpatuniversity.ac.in

**Abstract.** Due to their outstanding capabilities and wide range of potential applications, Wireless Sensor Networks have received more attention. Routing is very challenging task in wireless sensor networks, as packets are routed to the BS through a number of nodes. The lifetime of WSNs is extremely restricted. Wireless sensor network is a resource constrained network where reducing energy consumption is very crucial in order to enhance the network lifetime. It is important to share the network packet in an energy-efficient way. It also considers the residual power of battery to extend the life of the network. Energy conservation is very crucial for Wireless Sensor Networks. In this paper, we have surveyed various existing energy efficient routing approaches with their merits and demerits. We have also discussed some open research issues that will help researchers for further investigation in this field.

**Keywords:** Routing · Network lifetime · Residual · Energy

## 1 Introduction

Numerous tiny sensor nodes make up wireless sensor networks. The nodes are densely deployed. The sensor node's functions include sensing, monitoring, measuring and gathering data. After gathering the data, the sensor node transmits it to the sink node, where the user can access it remotely. The processor, sensor, battery, and transceiver are the main parts of the sensor node. Wireless sensor networks have applications in environmental detection, healthcare, tracking endangered species, military monitoring etc.

The nodes are low powered devices. They are operated by battery. They are widely dispersed in difficult or even unexplored environments. Once deployed, it is impractical to resupply them. When some proportion of the nodes becomes dead, network coverage is lost and the network lifetime will be decreased. It is an important issue to enhance the network lifetime. Therefore, reducing the energy consumption and improving the network lifetime has received attention recently [1–4].

The rest of the paper is organized as follows. The protocol stacks of wireless sensor networks and energy consumption characteristics at each layer are discussed in Sect. 2. Section 3 presents various existing energy efficient methods with their merits and demerits. Section 4 presents open research issues in this field. Section 5 draws our conclusion.

## 2 Energy Consumption and Protocol Stacks

Protocol stacks of wireless sensor networks consist of five layers: the physical layer, the data link layer, the network layer, the transport layer and the application layer. By reducing the amount of redundant data being transmitted and by reducing the data sampling frequency at the physical layer, energy consumption can be decreased. Through a properly constructed MAC mechanism, the energy cost resulting from over listening, idle listening and data collision at the data connection layer can be decreased. The flooding issues during the route establishment process must be taken into account at the network layer. Energy balance at the transport layer can be attained with the right multi-path multiplexing technology design. With the aid of lower layers and in accordance with the particular data generating pattern, energy efficiency can be achieved at the application layer.

## 3 Related Work

### 3.1 Leach [5]

Low Energy Adaptive Clustering Hierarchy has been proposed in [5] for periodical data gathering applications. Cluster heads are selected randomly. Data is received by the cluster head, which then aggregates it and transmits it via a single hop to the BS. Cluster head is rotated in every round. It does not consider residual energy while electing cluster head.

**Merits**—Energy conservation, load balancing to a certain point, avoiding unnecessary collisions.

**Demerits**—There is no multi-hop connection, there are energy holes, and communication is expensive.

### 3.2 Leach-C [6]

The algorithm presented in [6] is a centralized clustering. Each node broadcasts to the BS its location (as determined by a GPS receiver) and energy level during the initialization phase. The node with lesser energy cannot become cluster head. The cluster head for that round will be the node with the energy higher than the average as determined by the base station.

**Merits**—Cluster heads are well distributed over the network, good clustering.

**Demerits**—Do not scale for large area network.

### 3.3 Cluster Fuzzy-Based Algorithm [7]

In [7], authors have proposed a strategy for increasing network lifetime that combines an A-star method, a fuzzy approach, and a clustering algorithm. It prevents the low-energy node to become a cluster head. It also ensures that the numbers of cluster heads are

optimal and uniform distribution. It takes into account the factors including minimum hops, remaining power and node traffic density.

**Merits**—Improved network lifetime, best results are achieved in small-scale, static networks.

**Demerits**—Communication overhead, deterioration of performance in large-scale networks.

### 3.4 Leach-Mac [8]

In [8], authors have presented a method for reducing the randomness in the LEACH clustering algorithm. It stabilizes the cluster head count. It tries to keep the number of cluster head advertisements to a minimum.

**Merits**—Randomized cluster head count, load balancing and energy efficiency.

**Demerits**—Low energy consumption, absence of multihop intra cluster communication, randomized cluster head selection.

### 3.5 Energy-Aware Distributed Unequal Clustering [9]

Different competition radii are assigned to nodes depending on the base station's location and available energy. Number of nodes in the neighborhoods also considered while electing cluster heads. The size of the cluster is smaller which are nearer to the base station than the clusters that are far away. Compared to the existing methods, it provides better energy balancing.

**Merits**—Longer network lifespan, energy efficiency, and no equal clustering.

**Demerits**—Message overhead.

### 3.6 Unequal Clustering Size Model [10]

For balancing the CH's energy consumption, authors in [10] presented an unequal cluster size model for network structure. Cluster head nodes dissipate energy more uniformly and thereby extending the lifetime of the network.

**Merits**—Performs effectively in homogeneous networks and conserves energy.

**Demerits**—Number of nodes can change.

### 3.7 Energy-Aware Distributed Clustering [11]

In [11], authors have presented a cluster based routing protocol for wireless sensor networks with uneven node distribution. Using competition range, it creates even size clusters. In scarcely covered areas, it increases forwarding tasks by choosing nodes having higher energy.

**Merits**—Load balancing and energy-saving.

**Demerits**—Inefficient for randomized deployments, energy and communication overhead.

### 3.8 Fuzzy Based Balanced Cost CH Selection [12]

The authors of [12] have developed a balanced cost cluster head selection algorithm based on the concept of fuzzy. For the purpose of choosing the cluster head, it takes into account the remaining energy, distance from the sink, and density of the nodes nearby. Each node's eligibility index for the selection of the cluster head is calculated as an output of the fuzzy inference system. It ensures load balancing by choosing the best candidate for the role of cluster head.

**Merits**—Balance energy usage by taking into account the node's residual energy and the distances between nodes.

**Demerits**—Neglect the risk of CH failure and high energy consumption.

### 3.9 Distributed CH Scheduling [13]

In [13], authors have proposed a two-tier architecture based on sensor nodes' received base station signal strength indication. There are two layers in the network: primary and secondary. It guarantees that the cluster head is distributed evenly among the sensor nodes. It prevents the cluster head from being often chosen depending on signal strength and remaining energy. It can be successfully used for energy sensitive applications in wireless sensor networks.

**Merits**—Use of real data and dynamic cluster head selection based on the strength of received signals.

**Demerits**—The study system is small and the sensor status is constant.

### 3.10 K-means Based Clustering Algorithm [14]

In [14], authors have presented energy efficient K-means based clustering protocol. For improving the initial centroid selection procedure, midpoint algorithm is used. It takes into account Euclidean distance and residual energy for cluster head selection. It determines the ideal quantity of the intended clusters. The balanced cluster produced by the midpoint technique has uniformly distributed CHs and nearly equal numbers of sensor nodes in each cluster.

**Merits**—Adjusting the system to prolong the lifetime of nodes.

**Demerits**—Estimating the k-value is difficult and there is a significant time delay because of the heavy computational work.

### 3.11 Pareto Optimization Based Approach [15]

In [15], the goal is to identify the optimal routing tree that links the cluster heads to the base station. The authors have presented a centralized pareto optimization based

approach. It is suggested that a repair function be used to fix any flawed routing trees and direct the search toward the optimal routing tree. By increasing the throughput at the BS, the use of a dedicated routing tree improves the reliability of data delivery.

**Merits**—Using real data to enhance communication quality.

**Demerits**—Multiple interfaces are required to establish connection between nodes, and the requirement to look for the best power to convey data.

### 3.12 Energy Efficiency Semi Static Routing Algorithm [16]

In [16], residual energy is considered for clustering operation. The nodes would be arranged in a list at the time of cluster formation, and the cluster head would automatically rotate according to the order of the list. Every node functions initially as a single node cluster; after that, these smaller clusters start to merge with the larger ones. It is repeated until the scale of all clusters satisfies the threshold.

**Merits**—It addresses the issue of energy holes and has good scalability.

**Demerits**—Low robustness and lack of consideration of a node's distance from the cluster head.

### 3.13 Hybrid Energy Efficient Routing [17]

In [17], authors have proposed Hamilton energy-efficient routing protocol. During the network initialization phase, it creates clusters and connects each cluster's members on a Hamilton path for data transfer. It is not necessary to reconstitute the cluster; instead, the members of the path will alternate as cluster heads. It enables energy savings for network administration.

**Merits**—It extends the network's life and extends its stability period.

**Demerits**—Data flooding is caused by a lack of data aggregation at the sink node.

### 3.14 Improved ABC Algorithm [18]

In [18], the objective is to maximize system data acquisition quantity and to guarantee that the mobile sink moves in the shortest possible path while collecting data. The artificial bee colony algorithm that the authors have presented can effectively decrease data transmission, improve the efficiency and reliability of network data collection, minimize energy consumption, and enhancing the network lifetime. The number of mobile sinks can be increased to improve network connectivity.

**Merits**—The search space has been modified to improve convergence and balancing the energy of nodes in various clusters.

**Demerits**—Incorrect performance when there is noise and the dynamic state was not considered when changing the structure of clusters.



### 3.15 Hierarchical Energy Balancing Multipath [19]

A unique hierarchical strategy known as the Hierarchical Energy Balancing Multipath Routing Protocol for Wireless Sensor Networks is proposed in [19]. Cluster heads are chosen and dispersed throughout the region in an ideal manner that leads to effective load balancing utilization. To save energy, nodes' radios are turned off for a fixed amount of time as sleeping control rules.

**Merits**—While cluster heads are chosen, residual energy and distance from neighbor nodes are taken into account, and load balancing and shorter message delays are provided.

**Demerits**—It requires additional time to calculate cluster size and collect network data.

### 3.16 Novel Energy Aware Hierarchical Cluster Based Protocol [20]

New energy aware hierarchical cluster based protocol for extending network lifetime has been proposed by authors in [20]. Depending on the amount of energy left, cluster heads are chosen. Nodes with low energy alternate between sleeping and being active to balance energy usage.

**Merits**—Less frequently do the energy holes to occur.

**Demerits**—The network uses more energy when using a cluster head, and nodes in sleep mode may cause the risk of disconnecting.

### 3.17 Heuristic Algorithm for Clustering Hierarchy [21]

A new heuristic technique for clustering hierarchy is proposed in [21]. The inactive nodes and cluster leader nodes are chosen in each round. Selected nodes can be put into sleep mode using sleep scheduling without degrading network performance. The distribution of cluster head nodes can be improved by combining two separate solutions using the crossover operator. An objective function is used to evaluate the quality of the solution.

**Merits**—It functions well even at high degrees of heterogeneity. The scheduling of sleep time enhances the network lifetime.

**Demerits**—Long distances may need to be covered by a cluster head, which will use more energy.

### 3.18 Multi-level Route Aware Clustering [22]

In [22], authors have proposed distributed multi level route aware energy efficient clustering algorithm. Nodes can obtain the needed knowledge regarding potential pathways to the destination. It stops producing additional routing packets. Effective criteria are used to elect cluster heads.

**Merits**—It reduces transmission costs and prolongs network lifetime.

**Demerits**—It isn't flexible.

### 3.19 Double Phase Cluster Head Election Scheme [23]

Double phase cluster head election system is a new distributed energy-efficient clustering protocol proposed in [23]. There are two stages to the cluster head selection process: (1) The tentative cluster heads are chosen by probabilities based on the respective levels of initial and residual energy. (2) If any member of the tentative cluster heads' cluster has greater residual energy, that member will take their position to form the final set of cluster heads. Cluster heads are more likely to be selected from nodes with more energy.

**Merits**—There is no time wasted in waiting for the cluster head to be chosen because the network is continually active and network stability is assured.

**Demerits**—Up to two different cluster heads may be chosen in each round, increasing the network's energy usage.

## 4 Open Research Issues

The sensor nodes' collected data is redundant. The transmission of redundant data uses more energy. Therefore, new data aggregation techniques need to be developed. Whenever a sensor node becomes inactive, the remaining active nodes' communication cost is increased. Thus, maintaining network connectivity is essential for extending the network's lifespan. Most research papers focus on homogeneous sensor networks. Heterogeneous networks also need to be taken into account in real world circumstances. Mobile network routing is a challenging task. Therefore, creating an energy-efficient routing system for mobile wireless sensor networks is a difficult task.

Due to resource constraints, conventional cryptographic techniques cannot be employed to offer security in sensor networks. It is necessary to develop energy-efficient, lightweight protocol. The energy of the nodes is depleted by packet losses and retransmissions brought on by congestion. As a result, approaches for congestion avoidance and control are used [24]. Wireless sensor networks are dynamic by their nature. Using machine learning techniques to design energy efficient routing protocols is a good alternative to enable self learning from prior experience [25]. In order to support the development of energy efficient routing protocols in wireless sensor networks, the use of artificial intelligence techniques such as fuzzy logic, genetic algorithms, ant colony optimization, and particle swarm optimization are very promising [26, 27].

## 5 Conclusion

WSNs are resource limited devices. Therefore, developing energy efficient routing algorithm is very crucial task for increasing network lifetime. In this paper, we have discussed several energy efficient routing techniques with their merits and demerits. Finally, some open issues have been pointed out. This survey helps researchers to design appropriate routing protocol for their specific applications.

## References

1. Dogra, R., Kavita, S.R., Shafi, J., Kim, S., Ijaz, M.F.: ESEERP: enhanced smart energy efficient routing protocol for internet of things in wireless sensor nodes. *Sensors* **22**, 6109 (2022)
2. Dogra, R., Rani, S., Babbar, H., Krah, D.: Energy-efficient routing protocol for next-generation application in the internet of things and wireless sensor networks. *Wirel. Commun. Mobile Comput.* **2022**, 10 (Article ID 8006751) (2022)
3. Suresh Kumar, K., Vimala, P.: Energy efficient routing protocol using exponentially-ant lion whale optimization algorithm in wireless sensor networks. *Comput. Netw.* **197**, 108250 (2021)
4. Ahmad, I., Rahman, T., Zeb, A., Khan, I., Othman, M.T.B., Hamam, H.: Cooperative energy-efficient routing protocol for underwater wireless sensor networks. *Sensors* **22**(18), 6945 (2022)
5. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: Energy-efficient communication protocol for wireless micro sensor networks. In: *Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS '00)* (2000)
6. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans. Wirel. Commun.* **1**(4), 660–670 (2002)
7. Yuan, Y., Li, C., Yang, Y., Zhang, X., Li, L.: CAF: cluster algorithm and a-star with fuzzy approach for lifetime enhancement in wireless sensor networks. *Hindawi Publ. Corp.* **2014**, 17 (Article ID 936376) (2014)
8. Batra, P.K., Kant, K.: LEACH-MAC: a new cluster head selection algorithm for wireless sensor networks. *Wirel. Netw.* **22**, 49–60 (2016)
9. Gupta, V., Pandey, R.: An improved energy aware distributed unequal clustering protocol for heterogeneous wireless sensor networks. *Int. J. Eng. Sci. Technol.* **19**, 1050–1058 (2016)
10. Soro, S., Heinzelman, W.B.: Prolonging the lifetime of wireless sensor networks via unequal clustering. In: *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium*, p. 8. Denver, CO (2005)
11. JiguoYu, Y.Q., Wang, G., Xin, G.: A cluster-based routing protocol for wireless sensor networks with non uniform node distribution. *AEU-Int. J. Electron. C.* **66**(1), 54–61 (2012)
12. Mehra, P.S., Doja, M.N., Alam, B.: Fuzzy based enhanced cluster head selection (FBECS) for WSN. *J. King Saud Univ. Sci.* **32**(1), 390–401 (2020)
13. Kannan, G., Raja, T.S.R.: Energy efficient distributed cluster head scheduling scheme for two tiered wireless sensor network. *Egypt. Inform. J.* **16**, 167–174 (2015)
14. Ray, A., De, D.: Energy efficient clustering protocol based on K-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network. *IET Wirel. Sensor Syst.* **6**, 181–191 (2016)
15. Elhabyan, R., Shi, W., St-Hilaire, M.: A Pareto optimization-based approach to clustering and routing in wireless sensor networks. *J. Netw. Comput. Appl.* **114**, 57–69 (2018)
16. Du, T., Qu, S., Liu, F., Wang, Q.: An energy efficiency semi-static routing algorithm for WSNs based on HAC clustering method. *Inform. Fusion* **21**, 18–29 (2015)
17. Yi, D., Yang, H.: HEER- a delay-aware and energy-efficient routing protocol for wireless sensor networks. *Comput. Netw.* **104**, 155–173 (2016)
18. Yue, Y., Li, J., Fan, H., Qin, Q.: Optimization-based artificial bee colony algorithm for data collection in large-scale mobile wireless sensor networks. *Hindawi Publ. Corp. J. Sensors* **2016**, 12 (Article ID 7057490) (2016)
19. Gherbi, C., Aliouat, Z., Benmohammed, M.: An adaptive clustering approach to dynamic load balancing and energy efficiency in wireless sensor networks. *Energy* **114**, 647–662 (2016)

20. Ke, W., Yangrui, O., Hong, J., Heli, Z., Xi, L.: Energy aware hierarchical cluster based routing protocol for WSNs. *J. China Univ. Posts Telecommun.* **23**(4), 46–52 (2016)
21. Oladimeji, M.O., Turkey, M., Dudley, S.: HACH: heuristic algorithm for clustering hierarchy protocol in wireless sensor networks. *Appl. Soft Comput.* **55**, 452–461 (2017)
22. Sabet, M., Naji, H.: An energy efficient multi-level route-aware clustering algorithm for wireless sensor networks: a self-organized approach. *Comput. Elect. Eng.* **56**, 399–417 (2016)
23. Han, R., Yang, W., Wang, Y., You, K.: DCE: a distributed energy-efficient clustering protocol for wireless sensor network based on double-phase cluster-head election. *Sensors* **17**, 998 (2017)
24. Bohloulzadeh, A., Rajaei, M.: A survey on congestion control protocols in wireless sensor networks. *Int. J. Wirel. Inf. Netw.* **27**(3), 365–384 (2020). <https://doi.org/10.1007/s10776-020-00479-3>
25. Praveen Kumar, D., Amgoth, T., Annavarapu, C.S.R.: Machine learning algorithms for wireless sensor networks: a survey. *Inform. Fusion*, **49**, 1–25 (2019)
26. Thangaramya, K., Kulothungan, K., Logambigai, R., Selvi, M., Ganapathy, S., Kannan, A.: Energy aware cluster and neuro-fuzzy based routing algorithm for wireless sensor networks in IoT. *Comput. Netw.* **151**, 211–223 (2019)
27. Sabor, N., Abo-Zahhad, M.: A comprehensive survey of intelligent-based hierarchical routing protocols for wireless sensor networks. In: *Nature Inspired Computing for Wireless Sensor Networks*, pp. 197–257. Springer, Singapore (2020)



# Robust Vehicle Speed Estimation Based on Vision Sensor Using YOLOv5 and DeepSORT

Dea Angelia Kamil, Wahyono, and Agus Harjoko<sup>(✉)</sup>

Department of Computer Science and Electronics, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia

dea.angelia.kamil@mail.ugm.ac.id, {wahyo,aharjoko}@ugm.ac.id

**Abstract.** For driving safety purposes, the policymakers make regulations on the speed limit on the highway because vehicles high speed may cause an accident. Vision-based vehicle speed estimation requires limited human resources and minimizes human error. Therefore, the authors of this research propose a vehicle speed estimation pipeline using You Only Look Once version 5 (YOLOv5) for vehicle detection, tracking vehicles using Simple Online and Realtime Tracking with a Deep Association Metric (DeepSORT), and speed measurement process. The dataset built for this work is taken in Gadjah Mada University using a camera in bird view orientation. The detection precision and recall of the YOLOv5 pre-trained model used are both about 100% and has fps 10.762. The speed value reaches a Root Mean Square Error (RMSE) of 3.926 and a Mean Absolute Error (MAE) of 3.155. The RMSE obtained is compared with previous research and has the most significant value.

**Keywords:** Vehicle speed estimation · YOLOv5 · DeepSORT · Computer vision · Intelligent transportation system

## 1 Introduction

Driving safety scheme is improved with the popularity and development of the autonomous vehicle and part of Intelligent Transportation Systems: traffic density classification [1], vanishing point detection [2], illegally parked vehicles detection [3], traffic sign recognition [4], and vehicle classification [5]. One part of this scheme is vehicle speed detection. Not only do autonomous vehicles have to control the level of speed, but conventional vehicles also need to be managed. It is because vehicles moving at high speed may cause an accident. Therefore, the law enforcer makes the policy about managing the speed limit. This regulation may be applied on the highway. In fact, to this regulation, the speed detector is usually installed on the side road to detect the vehicle over the speed limit. A speed gun or radar gun is a frequently used tool for detecting vehicle speed. This method is based on non-intrusive technology [6]. There is much research using speed guns for obtaining ground truth information or to be the primary approach for detecting speed itself [7–10]. However, speed gun has disadvantages, and it depends on human power, which may cause human error. Radar guns have a simple

method to implement. People need to shoot a speed gun at the vehicle they want to detect the speed. However, human hands and eyes speed cannot handle it. The other technology is the intrusive-based method using an inductive loop detector. Luvizon et al. use this tool to generate the ground truth of their dataset [6]. This method avoids human error but has complicated installation. An inductive loop detector is installed under the asphalt and is sensitive to water. Moreover, the two leading technologies above are expensive.

Because of the problem mentioned, the authors proposed vehicle speed estimation based on computer vision technology. The joint vision-based speed estimation is divided into three main processes. Step one is detecting and classifying vehicles. The vehicle shown in the video is detected using a detector algorithm. The learning-based method is commonly used for detecting vehicles in speed estimation schemes such as SSD [11], YOLO family; YOLOv3 [12], YOLOv4 [12], Faster RCNN [13, 14], and Mask RCNN [15]. In this study, YOLOv5 is utilized for detecting vehicles. Step two is vehicle tracking. One of the famous tracking algorithms is Kalman Filter. Single Online and Realtime Tracking (SORT) [15, 16] is a tracker based on Kalman Filter and the Hungarian method for predicting motion. In this research, DeepSORT is used [15, 17]. After localization of the car, tracking is needed to track the bounding boxes in the whole appearance of the video. Furthermore, the distance and time travel ratio of vehicles passing the RoI is used for computing the vehicle velocity.

The following sections will cover the whole process of this research. Section 2 discusses the previous research on vehicle speed estimation using three-based processes, including detection, tracking, speed measurement, and learning-based approaches. Section 3 describes the proposed method's experimental, starting with how the dataset was collected, the experimental on object detection by YOLOv5, tracking using DeepSORT, and the method implemented for measuring vehicle speed. Section 4 presents the conclusion of the result, the weakness of the research, and the scheme of the future study.

## 2 Related Works

### 2.1 Learning-Based

The learning-based approach is rarely used for detecting vehicle speed because of the lack of data that has vehicle speed information in the video dataset. Generally, the learning-based method [18–21] discusses how to map the sequence of the image into regression information, in this case, speed. To realize the concept mentioned, Martinez et al. [20] adapted two different deep learning models: 3D CNN and CNN-RNN. The author generated dataset from synthetics driving simulator. Adapted two deep learning models: 3D CNN and CNN-RNN. The author generated a dataset from a synthetic driving simulator. However, due to the small number of data, the model needs to be more robustly explored whether environmental conditions influence the model. In the following research [21], the author increased the amount of data, and the experiment result indicates that if the model the author used, 3D CNN, was trained in a large dataset, it could be adequate to detect the speed of traveling vehicles with more outstanding results.

## 2.2 Three-Based Process

At least the process of getting vehicle speed value consists of the vehicle detecting, multi-object tracking, and estimating speed. In the part of vehicle detection influenced by the popularity of deep learning, the amount of research implementing deep learning for detecting and classifying vehicles is shown in the video data for the first step of the vehicle speed estimation scheme [12–17, 22]. The work of contribution to this step can be seen in the research of Lin et al. [12]. They compared YOLOv3 and YOLOv4 to detect and classify vehicles with an accuracy of about 98.02% and 99.50%, respectively. Furthermore, in the pipeline of estimating speed Inter-frame difference method is applied [20, 21]. This method consists of binary logical information and morphological information. Not only utilizing the inter-frame difference method but Trivedi et al. [23] also employ blob analysis to detect objects like in [6, 14]. However, in blob analysis, the different colors of the vehicle need to be sufficiently recognized. Still, in vehicle detection, Luvizon et al. [6] extracted features from vehicle movement using MHI (Motion Histogram Image) and matched the motion feature by SIFT (Scale Invariant Feature Transform).

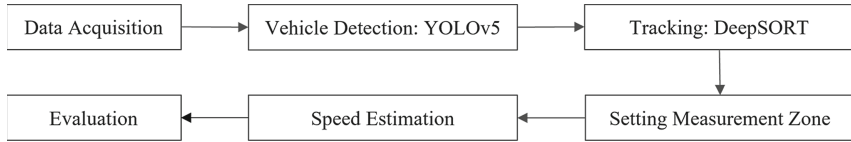
The following process of detecting vehicle speed is tracking. The amount of research employed Kalman Filter for the base of the tracking object [14–17]. Sochor et al. [14] applied Kalman Filter to track a feature group of blobs representing one car. Moreover, SORT [15, 16] and DeepSORT [15, 17] are the famous trackers used in vehicle speed measurement studies. SORT and DeepSORT employ Kalman Filter to predict the bounding box for comparison with the detected one. If the Kalman Filter is used for predicting the localization of an object, Kanade-Lucas, and Tomasi usually use the corner feature. In [6], after the extracted feature, the author applies Kanade Lucas Tomasi to track the feature to produce the motion vector. The last process is measuring vehicle speed. In general, we divide the technique of measuring speed into calibrated [6, 14–16] or uncalibrated [7, 12, 13, 17, 24]. Detailed information on vehicle speed measurement will be described in Sect. 2.3.

## 2.3 Speed Estimation

Mostly, the vehicle speed estimation technique without calibrating the camera inputs the distance information obtained from the laser meter or meter wheel in the between of line or region of interest. Then, the distance information operates with the pixel information. Hence, the velocity values get from the distance and time travel ratio. In [13], the author uses the warping technique, so the image seen in the eagle view and the road is straight and flat. This technique makes the measurement process more robust and does not need camera distortion. On the other hand, the calibration-based technique uses camera parameters to obtain image orientation for accurate distances. In the research by Kumar et al. [15], the authors used an affine rectification technique to calibrate the camera. This method uses using Homography matrix to get intrinsic and extrinsic camera parameters. After calculating the homography operation, the distance information can be computed. In [25], the authors used the frame rate information and starting and endpoint of the region of interest (RoI), in this research they utilized white dashed detection to localize the RoI using color thresholding, for estimating vehicle speed.

### 3 Proposed Method

The method proposed in this study is to estimate vehicle speed using a traffic video dataset. The three-based process is adopted to detect the vehicle speed: vehicle detection, vehicle tracking, and speed estimation. The whole process is included in Fig. 1.



**Fig. 1.** Speed estimation process

#### 3.1 Data Acquisition

The author collects the data built for this research from the veterinary faculty overpass of Gadjah Mada University, Indonesia. The video was captured using Canon 1200D. The testing video has 12 min long with 30 fps and the resolution is  $1920 \times 1088$ . The field of view of the camera setting is bird view. A frame of the video dataset is illustrated in Fig. 2.



**Fig. 2.** Image of traffic video dataset (left) Bushnell Speed Gun (right)

For the evaluation process, the ground truth value of vehicle speed is obtained by speed gun, Bushnell Speed Gun/Radar Gun Velocity, from above the overpass. The image of the radar gun is shown in Fig. 3. Because of the limitation of human power to catch the vehicle speed, we just captured one class of vehicle that is the car for this research, and not all the cars shown in the video have the ground truth value. The total number of cars evaluated in the system and the actual speed value is 38.

#### 3.2 Detection: YOLOv5

The limitation of this research is using only some of the vehicles shown in the video; the only vehicle detected is the car. Pre-trained YOLOv5s, YOLOv5m, YOLOv5l, and



YOLOv5x, are used for detecting cars. The pre-trained models were trained in the COCO dataset. Among 80 classes of the COCO dataset, the class used only cars. This pre-trained model is utilized due to the small number of datasets. Our dataset contains 449 images for training, 56 for validating, and 57 for testing. The labeling dataset is illustrated in Fig. 3.



Fig. 3. Dataset labelling (left) detected car in the testing dataset (right).

In this research, the authors use the architecture of YOLOv5. YOLOv5 is divided into three main parts: Backbone, Neck, and Head. In the backbone, CSPDarknet53 is used. Four pre-trained models were trained in a custom dataset with 16 batches and 100 epochs. For the research experiment, we train four different models of YOLOv5: YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. To evaluate their detection performance, we calculate the precision and recall value. Then, to evaluate the time assuming or computational complexity, we estimate the average frame rate of the processing video. The precision and recall result are written in Fig. 4.

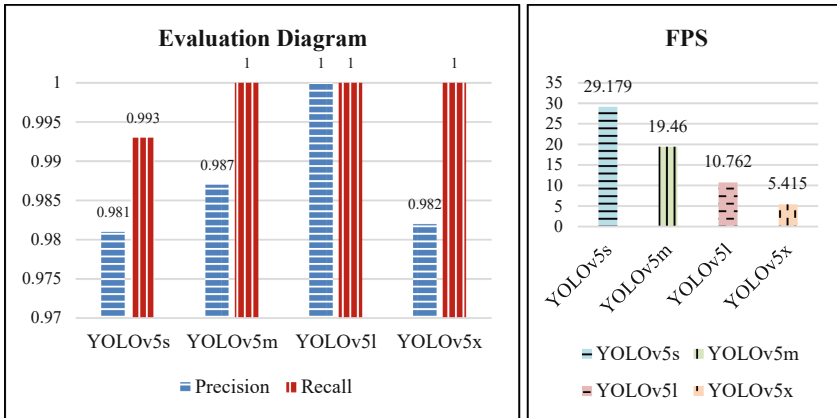


Fig. 4. Evaluation diagram (left) FPS diagram (right).

From the experiment, the highest precision and recall is reached by YOLOv5l of about 100% in both precision and recall. YOLOv5s have the lowest computation complexity, with a frame rate value reaching 29.179 fps. However, YOLOv5s have the lowest

precision and recall, 98.1% and 99.3%, respectively. In this research, we tend to use the YOLOv5l model since have the best performance, 100% of precision and recall, and the frame rate, 10.762, is better than YOLOv5x with a fps of about 5.415.

### 3.3 Tracking: DeepSORT

DeepSORT or Simple Online and Realtime Tracking with a Deep Association Metric is used for tracking the vehicle after detected. In DeepSORT, there are at least three steps: tentative, confirmed, and deleted. The whole process is included in Fig. 5 The cars detected using YOLOv5 would be tracked using Kalman Filter to predict the tracks of the bounding box from the past frame to the next frame. The tracker will be deleted if the track is lost when less than five frames are reached. The scenario of the tracking is included in dimensional state space:

$$(u, v, \gamma, h, \dot{x}, \dot{y}, \dot{\gamma}, \dot{h}), \quad (1)$$

where  $(u, v)$  is the bounding box position,  $\gamma$  is the aspect ratio,  $h$  is height, and  $(\dot{x}, \dot{y}, \dot{\gamma}, \dot{h})$  are the respective velocities in image coordinates. The detection is based on their two metrics using verified tracks and matching cascade. The first metric is Mahalanobis distance, and the second metric is feature metrics generated using a pre-trained CNN model for bounding box appearance descriptor; in this research, mobilenet architecture is applied. Unmatched tracks and detections will be passed to the next filter using the Hungarian algorithm with the IoU cost matrix.

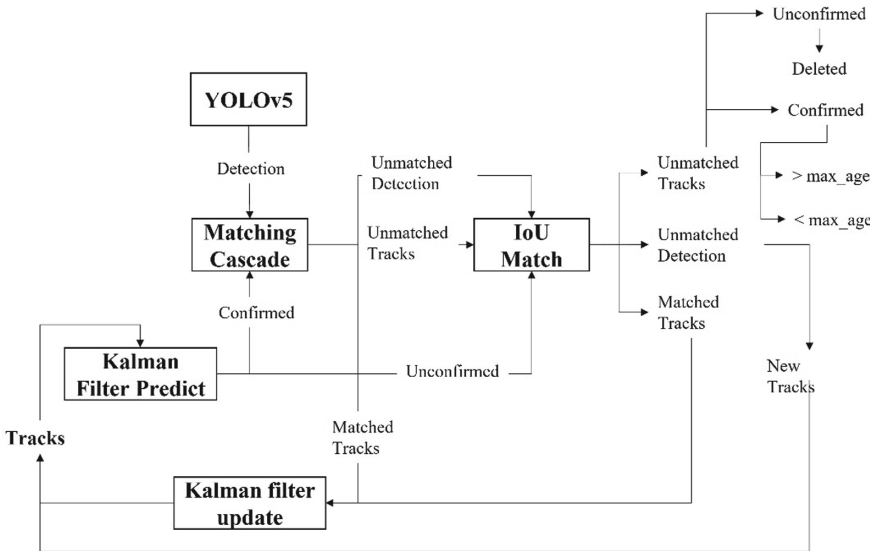


Fig. 5. DeepSORT workflow diagram

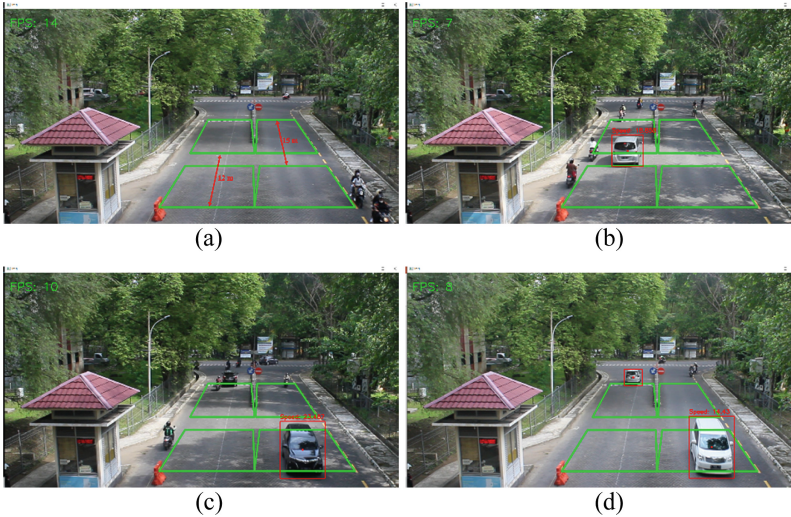
### 3.4 Speed Estimation

For estimating vehicle speed, pre-trained model YOLOv5s since they have the lowest time complexity. The following equation computes the vehicle speed measurements:

$$t = \frac{N}{fps}, \quad (2)$$

$$Speed = \frac{x \times 3.6}{t}, \quad (3)$$

where  $N$  is the number of frames, and  $fps$  is the frame rate. For obtaining the distance of the vehicle, four Region of Interest is localized. In this experiment, two vehicles' directions are computed both toward and away from the camera. The distance between two regions  $x$  in each direction was measured by a meter laser, which is 12 m and 10 m. For the detail, look at Fig. 6. Furthermore, the elapsed time is calculated from the vehicles entering the first ROI and out of the second one for each direction based on generated centroid from the middle of the bounding box. The experiment result is also illustrated in Fig. 6. Including toward the camera, away from the camera, and both. Since the road segment has a roadblock in the center, the ROI can work adequately in each direction. Moreover, even if vehicles are in either direction, the system for both vehicles can operate.



**Fig. 6.** Vehicle speed estimation: **a** Region of Interest (RoI) setting, **b** toward the camera, **c** away from the camera, and **d** both.

The number of ground truth data obtained from radar guns is 38 data. Table 1 contains vehicle speed results by the proposed system, speed ground truth by speed gun, and the difference between them. The car's speed is at most 34 km/h since the road is between the crossroads and the entrance gate of the college. The most considerable difference

between the system and ground truth is 11.593 km/h. The metrics utilized for evaluation are Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{r}_i - r_i)^2}, \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{r}_i - r_i|, \quad (5)$$

where,  $\hat{r}_i$  is the prediction rating and  $r_i$  is the actual rating in data set,  $n$  is the number of samples, and  $e$  is the model error composed of as  $e_i$  ( $i = 1, 2, \dots, n$ ).

**Table 1.** Vehicle speed evaluation

Vehicle	Speed Estimation (km/h)			Vehicle	Speed Estimation (km/h)		
	System detection	Speed-gun detection	Difference		System detection	Speed-gun detection	Difference
1	32.995	29.000	- 3.995	20	35.564	34.000	- 1.564
2	24.379	26.000	1.621	21	22.412	25.000	2.588
3	24.23	19.000	- 5.23	22	17.31	24.000	6.69
4	22.251	20.000	- 2.251	23	20.788	18.000	- 2.788
5	22.883	21.000	- 1.883	24	23.175	18.000	- 5.175
6	16.335	21.000	4.665	25	18.006	23.000	4.994
7	24.209	20.000	- 4.209	26	20.005	23.000	2.995
8	21.595	22.000	0.405	27	24.691	23.000	- 1.691
9	18.335	17.000	- 1.335	28	23.829	25.000	1.171
10	23.968	20.000	- 3.968	29	21.59	27.000	5.41
11	21.595	22.000	0.405	30	29.024	32.000	2.976
12	23.968	20.000	- 3.968	31	21.415	25.000	3.585
13	17.241	17.000	- 0.241	32	20.088	23.000	2.912
14	30.706	29.000	- 1.706	33	24.182	28.000	3.818
15	31.351	30.000	- 1.351	34	25.738	21.000	- 4.738
16	24.948	24.000	- 0.948	35	30.588	31.000	0.412
17	30.531	24.000	- 6.531	36	19.407	21.000	1.593
18	28.279	24.000	- 4.279	37	20.503	19.000	- 1.503
19	32.714	26.000	- 6.714	38	19.407	31.000	11.593

The result reaches an RMSE of 3.926 and an MAE of 3.155. The result is excellent but still needs to be improved. The proposed study and the previous research are shown in

Table 2. It is shown that the proposed method's result of this study has the minimum error. However, the dataset used differs from the previous study, so the RMSE comparison is not totally equal.

**Table 2.** Vehicle speed comparison result

References	Method	RMSE
Tang et al. [22]	YOLOv2, 3D SCT, Vehicle speed estimation	4.096
Kumar et al. [15]	Mask RCNN, SORT, Image rectification, Vehicle speed estimation	9.540
Kumar et al. [15]	Mask RCNN, DeepSORT, Image rectification, Vehicle speed estimation	10.100
Huang [13]	Faster RCNN, Histogram-based Tracking, Speed conversion	8.609
Proposed method	YOLOv5, DeepSORT, Vehicle speed estimation	3.926

## 4 Conclusion

We proposed the vehicle speed estimation method based on three steps: vehicle detection using YOLOv5, multi-object tracking using DeepSORT, and vehicle speed measurement. The authors build a primer traffic video dataset with ground truth value obtained from speed gun. Four YOLOv5 pre-trained models are trained using the custom dataset that our build contains 449 images for training, 56 for validating, and 57 for testing and comparing the precision, recall, and fps value. The experiment's highest precision, 100%, and recall, 100%, are reached by YOLOv5l with 10.762 fps. For the vehicle speed estimation result, we reach RMSE 3.926 and MAE 3.155 through 38 data. The results are excellent. This research helps the law enforcer manage the speed limit in traffic. However, the dataset developed only contains slow vehicles speed, at most 34 km/h. In future research, we suggest including an automatic calibration technique to minimize error and append high vehicle speed.

**Acknowledgments.** This research was financially supported by the PMDSU Scholarship (Master Program of Education Leading to Doctoral Degree for Excellent Graduates), funded by the Ministry of Education, Culture, Research, and Technology.

## References

1. Kholik, A., Harjoko, A., Wahyono, W.: Classification of traffic vehicle density using deep learning. *Indones. J. Comput. Cybern. Syst.* **14**, 69 (2020)
2. Kamil, D.A., Wahyono, Harjoko, A.: Vanishing Point Detection Using Angle-Based Though Transform and RANSAC, pp. 1–5 (2023)

3. Wahyono, Jo, K.H.: Cumulative dual foreground differences for illegally parked vehicles detection. *IEEE Trans. Ind. Inform.* **13**, 2464–2473 (2017)
4. Gadri, S., Adouane, N.E.: Efficient traffic signs recognition based on CNN model for self-driving cars BT—intelligent computing & optimization. Presented at the (2022)
5. Saha, R., Debi, T., Arefin, M.S.: Developing a framework for vehicle detection, tracking and classification in traffic video surveillance BT—Intelligent Computing and Optimization. Presented at the (2021)
6. Luvizon, D.C., Nassu, B.T., Minetto, R.: A video-based system for vehicle speed measurement in urban roadways. *IEEE Trans. Intell. Transp. Syst.* **18**, 1393–1404 (2017)
7. Sundoro, H.S., Harjoko, A.: Vehicle counting and vehicle speed measurement based on video processing. *J. Theor. Appl. Inf. Technol.* **84**, 233–241 (2016)
8. Rodríguez-Rangel, H., Morales-Rosales, L.A., Imperial-Rojo, R., Roman-Garay, M.A., Peralta-Peñuñuri, G.E., Lobato-Báez, M.: Analysis of statistical and artificial intelligence algorithms for real-time speed estimation based on vehicle detection with YOLO. *Appl. Sci.* **12**, 1–20 (2022)
9. Famouri, M., Azimifar, Z., Wong, A.: A novel motion plane-based approach to vehicle speed estimation. *IEEE Trans. Intell. Transp. Syst.* **20**, 1237–1246 (2019)
10. Sun, R., Zhuang, X., Wu, C., Zhao, G., Zhang, K.: The estimation of vehicle speed and stopping distance by pedestrians crossing streets in a naturalistic traffic environment. *Transp. Res. Part F Traffic Psychol. Behav.* **30**, 97–106 (2015)
11. Liu, C., Huynh, D.Q., Sun, Y., Reynolds, M., Atkinson, S.: A vision-based pipeline for vehicle counting, speed estimation, and classification. *IEEE Trans. Intell. Transp. Syst.* **22**, 7547–7560 (2020)
12. Lin, C.J., Jeng, S.Y., Lioa, H.W.: A real-time vehicle counting, speed estimation, and classification system based on virtual detection zone and YOLO. *Math. Probl. Eng.* **2021** (2021)
13. Huang, T.: Traffic speed estimation from surveillance video data Tingting Huang institute for transportation, Iowa State University. *IEEE Conf. Comput. Vis. Pattern Recognit.* 161–165 (2018)
14. Sochor, J., Juránek, R., Herout, A.: Traffic surveillance camera calibration by 3D model bounding box alignment for accurate vehicle speed measurement. *Comput. Vis. Image Underst.* **161**, 87–98 (2017)
15. Kumar, A., Khorramshahi, P., Lin, W.A., Dhar, P., Chen, J.C., Chellappa, R.: A semi-automatic 2D solution for vehicle speed estimation from monocular videos. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* **2018**(June), 137–144 (2018)
16. Revaud, J., Humenberger, M.: Robust automatic monocular vehicle speed estimation for traffic surveillance. *Proc. IEEE Int. Conf. Comput. Vis.* 4531–4541 (2021)
17. Huu, P.N., Duy, M.B.: An algorithm using YOLOv4 and DeepSORT for tracking vehicle speed on the highway. *Indones. J. Electr. Eng. Inform.* **10**, 90–101 (2022)
18. Lee, J., Roh, S., Shin, J., Sohn, K.: Image-based learning to measure the space mean speed on a stretch of road without the need to tag images with labels. *Sensors* **19**, 1–19 (2019)
19. Dong, H., Wen, M., Yang, Z.: Vehicle speed estimation based on 3D ConvNets and non-local blocks. *Futur. Internet.* **11**, 123 (2019)
20. Martínez, A.H., Díaz, J.L., Daza, I.G., Llorca, D.F.: Data-driven vehicle speed detection from synthetic driving simulator images. *IEEE Conf. Intell. Transp. Syst. Proc. ITSC.* **2021**(September), 2617–2622 (2021)
21. Martínez, A.H., Llorca, D.F., Daza, I.G.: Towards view-invariant vehicle speed detection from driving simulator images (2022)
22. Tang, Z., Wang, G., Xiao, H., Zheng, A., Hwang, J.N.: Single-camera and inter-camera vehicle tracking and 3d speed estimation based on fusion of visual and semantic features. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn. Work.* **2018**(June), 108–115 (2018)

23. Trivedi, J.D., Mandalapu, S.D., Dave, D.H.: Vision-based real-time vehicle detection and vehicle speed measurement using morphology and binary logical operation. *J. Ind. Inf. Integr.* **27**, 100280 (2021)
24. Grents, A., Varkentin, V., Goryaev, N.: Determining vehicle speed based on video using convolutional neural network. *Transp. Res. Proc.* **50**, 192–200 (2020)
25. Ramasamy, S., Joshua Thomas, J.: AutoMove: an end-to-end deep learning system for self-driving vehicles BT—intelligent computing and optimization. Presented at the (2021)



# Automatic Alignment of Aerospace Images Based on the Search for Characteristic Points

A. Kolesenkov<sup>1</sup>, B. Kostrov<sup>1</sup>, V. Panchenko<sup>2,3</sup> , and Yu Daus<sup>4</sup>  

<sup>1</sup> Ryazan State Radio Engineering University Named After V.F. Utkin, Gagarin St., 59/1, 390005 Ryazan, Russia

kolesenkov.a.n@kt.rsreu.ru

<sup>2</sup> Russian University of Transport, Obraztsova St. 9, 127994 Moscow, Russia  
pancheska@mail.ru

<sup>3</sup> Federal Scientific Agroengineering Center VIM, 1st Institutsky Passage 5, 109428 Moscow, Russia

<sup>4</sup> Kuban State Agrarian University, Kalinina St. 13, 350044 Krasnodar, Russia  
zirochka2505@gmail.com

**Abstract.** The purpose of the research is to design a method for automatic alignment of aerospace images (ASI) based on the use of sequence analysis and allowing achieving high speed with a low error probability, taking into account distortions and the influence of external factors. The article analyzes existing solutions, explores their shortcomings and possibilities for their elimination. Method for ASI alignment based on the use of the “autocorrelation” function to search for characteristic alignment points in the current image is proposed. Step-by-step description of the algorithm is presented; the corresponding mathematical apparatus is given. We used an automatic image alignment technique using Principal Component Analysis (PCA), which is based on assignment rules derived from eigenvectors given by PCA. Testing was performed on real pairs of ASI from spacecraft. The dependence of the calculation of the cross-correlation function on the rotation angle of images relatively to each other is demonstrated. The implementation of the proposed method in the form of the algorithm is carried out and comparison with other known algorithms is performed. Approbation of the algorithm confirmed its effectiveness in solving problems of automatic image alignment. It is conducted that the reliability of the proposed alignment method increases with the elimination of redundancy in ASI.

**Keywords:** Digital image processing · Correlation method · Image combination · Aerospace images · Remote sensing · Autocorrelation function · Redundancy · Sequential analysis · Walsh transform

## 1 Introduction

Image alignment is an important part of data processing information systems, which consists in comparing the pixels of two images corresponding to the same points on the earth’s surface [1, 2]. Alignment requires the establishment of the correspondence



between the elements of two images, which boils down to the selection of the so-called characteristic (reference or check) points, according to which the images are coordinated [3]. One of the sections of image processing is the processing of aerospace images (ASI), which have their own characteristics and require an individual approach. For example, correlation methods work better for them [4, 5].

The following tasks are solved by combining ASI:

- georeferencing of aerospace images [6];
- update of aerospace data (replacement of old images with new ones);
- “stitching” of frames of remote shooting of the earth’s surface [7];
- autonomous navigation of unmanned aerial vehicles;
- construction of virtual 3D panoramas.

In the process of combining ASI, a number of difficulties can be identified [8, 9]:

- the existing algorithms for combining aerospace images work effectively on certain types of underlying surface, containing, for example, a large number of contrast transitions, while on other types of surface, the correctness of identifying a pair of characteristic points is significantly reduced;
- superimposed ASI, as a rule, have a temporal mismatch and, accordingly, differ greatly in the spatial plane, which worsens the identification of characteristic points;
- in practice, ASI of the same area obtained from different sources may differ significantly from each other due to cloudiness or the influence of other external factors, which complicates the alignment process.
- disposition of characteristic points in current systems is performed in semi-automatic or manual modes, which increases the likelihood of errors;
- classical correlation matching, which is reduced to finding the maximum of the two-dimensional correlation function of images, is characterized by high computational costs.

It is also worth noting that it is problematic to fully automate the process of combining two ASI, since they are often rotated relatively to each other by several degrees, and their parts can be partially or completely covered by clouds or precipitation [10].

In this regard, it is an urgent task to design new effective methods for combining ASI that can work with any Earth remote sensing data and are distinguished by the automatic search for characteristic points and their displays.

## 2 Materials and Methods

The aerospace image is a picture that is obtained as a result of remote sensing of the earth’s surface using satellite vehicles, multicopters or unmanned aerial vehicles.

Any ASI can be represented as a discrete signal [11], given at finite intervals of coordinate readings in the corresponding plane. If the numbers  $f(i, j)$  are the brightness values of the bitmap at each point (pixel), then in this case the image itself is given by a

matrix of the form:

$$\mathbf{F} = \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1N} \\ f_{21} & f_{22} & \cdots & f_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ f_{M1} & f_{M2} & \cdots & f_{MN} \end{pmatrix}. \quad (1)$$

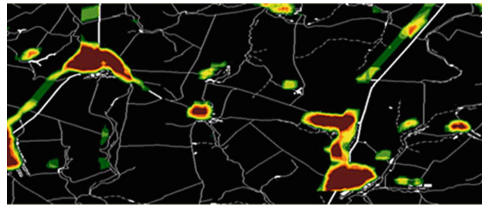
The paper proposes a method for automatic alignment of ASI, which consists of the following steps:

- search for favorable areas (having a large high-frequency component) (Fig. 1) and their corresponding characteristic points based on the calculation of the “autocorrelation” function  $R_a(p, g)$  with the size  $K \times K$  at each point of the current ASI; these areas may contain elements of coastlines, rivers, road network, residential and non-residential buildings.
- identification of reference fragments on ASI of the given size  $K \times K$  with the maximum value of the “autocorrelation” function in each favorable area (Fig. 2).
- search for coincidence points on the combined ASI using the correlation-extremal approach (Fig. 3) and calculation of the cross-correlation function (CCF)  $R$  at the point with coordinates  $(p, g)$ :

$$R(p, g) = \frac{1}{K} \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} [f_{RF}(i, j)] * [f_{CF}(i + p, j + g)] \quad (2)$$

where  $[f_{RF}(i, j)]$  and  $[f_{CF}(i + p, j + g)]$ —matrixes of pixels of the reference fragment (RF) and the fragment (CF) of the current image (CI) at the point with coordinates  $(p, g)$ ;  $N \times N$ —CI size;  $K \times K$ —RF size;  $p = 0, N - K - 1, g = 0, N - K - 1$ —RF shift relative to the zero point of CI; \*—element wise multiplication symbol;  $\Sigma$  sign denotes summation of elements across rows and columns.

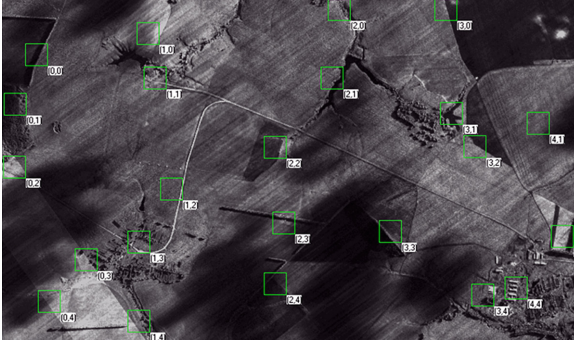
Superposition of overlapping elements of the combined ASI on the current one by characteristic points. Removing seams in the resulting image that appeared as a result of the merging operation.



**Fig. 1.** Search for favorable sites for combination

There are many image alignment methods that use various measures of image similarity, which are applied as the quality objective function in the synthesis of algorithms, among which are [12, 13]:

- correlation algorithms (cross-correlation coefficients, Tanimoto, Spearman and Kendall rank correlation);



**Fig. 2.** Reference fragments on ASI

- informational algorithms (entropy of the joint brightness probability distribution density, Shannon, Renyi, Tsallis mutual information, F-information measures).

As a result of the research, it was detected that the use of the correlation function in the basis of piecewise constant functions shows the best results when solving the problem of combining ASI among all the above approaches, because they are better suited for working with digital images that have a high-frequency component of the brightness difference [13]. They are capable of minimizing alignment errors arising from terrain inaccuracies, features of the survey equipment, errors in geometric correction and georeferencing of images.

The methods of classical analysis when combining ASI use calculations in the spatial domain, have low speed and are distinguished by the detection of numerous local extrema in the calculation of the CCF [1]. More modern methods of ASI processing are based on the use of the Fourier transform, but have a greater computational complexity [15].

In the process of conducting research on the spectral processing of ASI, it was revealed that the transition from trigonometric functions to piecewise constants can significantly simplify mathematical calculations and thereby reduce the requirements for computing resources [16].

The most commonly used is the generalized Walsh–Hadamard transform based on the Walsh functions (with two values  $+1$  and  $-1$ ). The technique for using this transformation is based on theorems that differ from the theorems of classical spectral analysis, the consequences of which lead to the development of new methods for analyzing ASI. In this case, the digital image spectrum matrix is calculated as:

$$\mathbf{S} = \frac{1}{K^2} \mathbf{W} \mathbf{F} (\mathbf{W})^T, \quad (3)$$

where  $\mathbf{F}$ —image pixel matrix;  $\mathbf{W}$ —Walsh matrix;  $K \times K$ —image matrix size.

The application of non-trigonometric basis functions is advisable when working with discrete images.

In the design process, there was used the well-known fact that superimposed ASIs have a certain amount of data redundancy, the elimination of which will lead to a significant reduction in computational costs with a slight decrease in the superimposition accuracy [17].

To reduce the complexity and reduce the dependence on distortions, noise or changes in the ASI caused by external factors or time, it is proposed to use a real-dyad convolution of two images of the following form as a CCF:

$$R(p, g) = \frac{1}{K} \sum_{i=0}^{K-1} \sum_{j=0}^{K-1} [s_{RF}(i, j)] * [s_{CF}(i + p, j + g)] \quad (4)$$

where  $\sum_{j=0}^{K-1} [s_{RF}(i, j)]$  and  $[s_{CF}(i + p, j + g)]$ —RF and SF spectra in the Walsh basis;  $N \times M$ —CI size;  $K \times K$ —RF size;  $p = \overline{0, N - K - 1}$ ;  $g = \overline{0, M - K - 1}$ —RF shift relatively to the zero point of CI.

An important advantage of using the Walsh spectra is that in this case, the normalization of the CCF is also fully or partially achieved.

To reduce the initial information content of the ASI, it is proposed to apply the **P** filter, zeroing out some of the rows and columns in the Walsh matrix [18].

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix}, \quad (5)$$

In this case, there is a modified Walsh matrix of the following form:

$$\mathbf{W}_m = \mathbf{W} * \mathbf{P}. \quad (6)$$

In this case, the CCF will be calculated according to formula (4) with the replacement of the Walsh matrix by the modified matrix (6).

Applying image filtering at an early stage improves the performance of the algorithm.

“Zeroing” some functions in the Walsh matrix when searching for characteristic points on the current ASI leads to a more explicit selection of image elements [19], which increases the reliability of the search for matching points and can be used to automate this process using the “autocorrelation” function obtained using a double substitution spectrum of the CF  $[s_{CF}(i + p, j + g)]$  in the Walsh basis into formula (2). When processing ASI, the Walsh–Hadamard fast transformation algorithm is used.

Evaluation of performance and reliability. To reject characteristic points when combining ASI, the  $r$  evaluation criterion is introduced that characterizes the ratio of local and global extrema:

$$r = 1 - \frac{R_{02}}{R(i_{\max}, j_{\max})} \geq r^*, \quad (7)$$

where  $R_{02}$ —value of the second extremum of the correlation function,  $R(i_{\max}, j_{\max})$ —CCF value at its maximum.

The larger the  $r$  value is, the less chance of false image alignment is. The result will be considered false if the inequality is true  $r < r^*$ , where  $r^*$ , is the value of the

assessment of the combination quality  $r$ , at which the result of the calculation of the CCF is reliable.

Performance evaluation is based on comparison of the proposed method with the classical one.  $T_1$ —the complexity of the basic method of image alignment;  $T_i$ —the complexity of the studied method of combining images.

**Table 1.** Estimation of  $r$  for different numbers of “nullable” functions

Noat	$T_i/T_1$	$r$ at $a = 0/7$ — number of “nullable” functions							
		0	1	2	3	4	5	6	7
1	1.00	0.23	0.32	0.31	0.22	0.12	0.15	0.13	0.15
2	0.47	0.24	0.34	0.47	0.51	0.55	0.49	0.48	0.54

Comparing the experimental results (Table 1), there was observed an increase in the reliability of the algorithm with the noticeable decrease in computational costs compared to the classical approach based on the calculation of the CCF in the brightness space.

The results of the experiments also show that the application of the proposed method makes it possible to perform automatically ASI alignment.

The implementation of the proposed method in the form of the algorithm made it possible to identify its features, advantages and disadvantages:

(1) Features are:

- a significant reduction in computational costs due to the work in the spectral region in a non-trigonometric basis;
- the ability to work even in the presence of interference and geometric distortions;

(2) advantages:

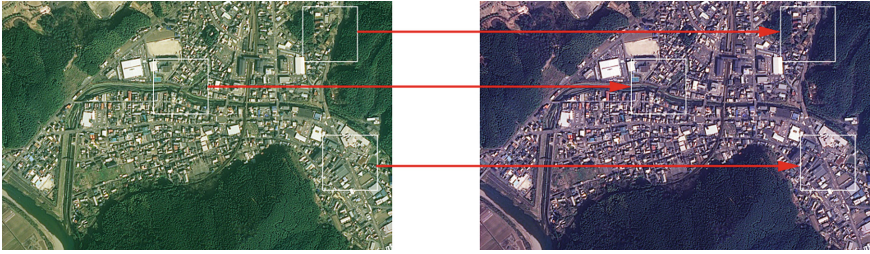
- improving the quality of the CCF by increasing the sharpness of the global extremum;
- good results on multi-temporal ASI;
- high resistance to mutual rotation of combined images up to 5 degrees;
- the possibility of additional filtering and “removal” of redundant information or interference;

(3) disadvantages:

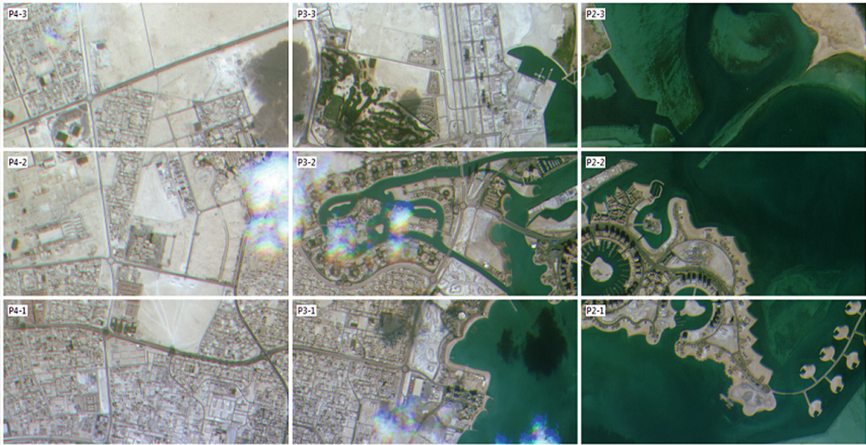
- an increase in the number of “false positives” in the event of cloudiness on the ASI;
- the impossibility of combining different-scale images;
- the method has been tested on ASI, for its application in other areas, additional studies and selection of parameters are required.

The application of the proposed method was tested on the technology of “stitching” [19, 20] adjacent ASI frames (Fig. 4) into a single image (Fig. 5) with further smoothing of the seams (Fig. 6). The search for overlapping points of adjacent frames of aerospace survey is presented at Fig. 3.

Experiments were carried out to compare the proposed method with the most popular image alignment methods: SIFT [21], SURF [22], and ORB. The size of the test collection included 100 ASI.



**Fig. 3.** Search for displaying reference fragments



**Fig. 4.** Scheme of joining adjacent ASI frames



**Fig. 5.** ASI combination

According to the results of experimental studies, the proposed method gives a significant gain in speed (up to 74.2%) with comparable or better accuracy and the number



**Fig.6.** The result of ASI “crosslinking”

of “false comparisons” (up to 57.3%). First of all, these results relate to ASI, since the proposed method has the following features:

- resistance to noise and illumination, since different-time ASIs are usually made at different times of the day;
- resistance to distortion and rotation;
- resistance to the appearance of additional objects, such as clouds.

The proposed method also showed high accuracy being used in automatic data update technology in information and analytical systems.

### 3 Conclusion

The application of the proposed method for solving the problem of automatic alignment of ASI with pixel accuracy shows the viability of the approach based on real-dyadic convolution and elimination of redundancy due to filtering in the frequency domain [11]. The combination of images is possible even with significant seasonal variability of the terrain, as well as the presence of mutual geometric distortions. It is possible to introduce restrictions on the allowable number of “resettable” functions when calculating the cross-correlation function [23]. The proposed image alignment method does not use the specifics of any satellite vehicles and is also applicable for processing images from multicopters and unmanned aerial vehicles.

### References

1. Aksenov, O.Y.: Combining images. *Digital Signal Process.* **3**, 51–55 (2005)
2. Domov, M.V., Zotin, A.G.: Constructing a scene image by combining consecutive frames. *Vestnik Siberian State Aerospace Univ M. F. Reshetnev*, **5**(31), 212–216 (2010)
3. Gulyaev, P.V.: Application of reference marks for coordinate reference to the surface in scanning probe microscopy. *Comput. Opt.* **3**(44), 420–426 (2020)
4. Kozhin, A.I., Pochivalov, S.G., Topchy, P.N.: Method of correlation recognition of objects observed from multispectral data. *Telecommun. Transp.* **6** (7), 38–41 (2013)
5. Zlobin, V.K., Eremeev, V.V.: *Processing of Aerospace Images*. FIZMATLIT, Moscow (2006)

6. Egozhkin, N.A., Ereemeev, V.V., Kozlov, E.P., et al.: Geodesic binding of images from geostationary satellites along the contour of the Earth's disk and electronic maps. *Modern Probl. Remote Sensing Earth Space* **1**(6), 132–138 (2009)
7. Chekhov, D.V., Tsudikov, M.B., Balyasny, S.V.: Investigation of the process of stitching several frames into a single image. *Izvestiya TULSU. Tech. Sci.* **1**, 306–313 (2013)
8. Aleksanin, A.I., Morozov, M.A., Fomin, E.V.: Problems of combining images with pixel accuracy. *Modern Probl. Remote Sensing Earth Space* **1**(16), 9–16 (2019)
9. Frolov, V.N., Tupikov, V.A., Pavlova, V.A., Alexandrov, V.A.: Methods of information combination of images in multichannel optoelectronic systems. *Izvestiya Tula State University. Tech. Sci.* **11–3**, 95–104 (2016)
10. Biktimirov, L.S., Tashlinsky, A.G.: Features of image combinations in conditions of intense interference. *REDS Telecommun. Dev. Syst.* **4**(3), 321–324 (2013)
11. Kostrov, B.V., Kolesenkov, A.N., Ruchkina, E.V.: Algorithms of aerospace photo processing in a quasi-two-dimensional non-trigonometric spectral space. *ICMSC Int. Conf. Mech. Syst. Control Eng.* 305–309 (2017)
12. Barabin, G.V., Gusev, V.Y., Zaitsev, V.E., Yurov, N.N.: Methods of constructing a single image in satellite imagery with a partitioned image sensor. *Bull. Comput. Inform. Technol.* **4**(118), 15–20 (2014)
13. Kostrov, B.V., Svirina, A.G., Zlobin, V.K.: *Spectral analysis of images in finite bases: monograph.* KURS Publishing House, Moscow, Russia (2016)
14. Tolmacheva, A.V.: Review and comparative analysis of image matching methods. *Methods and means of information processing and storage. Interuniv. Collection Sci. Papers* 40–42 (2018)
15. Chochia, P.A.: Fast correlation combination of quasi-regular images. *Inform. Process.* **3**(9), 117–120 (2009)
16. Babaev, S.I., Baranchikov, A.I., Grinchenko, N.N., et al.: The directions for collaborate usage of flight apparatus technical vision system information and electronic cartography. In: *MECO: The 5th Mediterranean Conference on Embedded Computing*, pp. 153–157 (2016)
17. Goshin, E.V., Kotov, A.P., Fursov, V.A.: Two-stage formation of a spatial transformation for combining image. *Comput. Opt.* **4**(38), 886–891 (2014)
18. Golubov, B.I., Skvortsov, V.A., Efimov, A.V.: *Walsh Series and Transformations: Theory and Applications.* URSS, Moscow (2007)
19. Petrov, E.P., Albakhtin, K.Y., Kharina, N.L.: The method of crosslinking satellite images. *DSPA Questions Appl. Digital Signal Process.* **3**(7), 128–131 (2017)
20. Silantyev, R.V.: Crosslinking of digital overlapping images in aerial photography problems. *Nauchny Obozrevatel* **12**, 89–92 (2013)
21. Wu, J., Cui, Z., Sheng, V.S., Zhao, P., Su, D., Gong, S.: A comparative study of SIFT and its variants. *Measurement Sci. Rev.* **3**(13), 22–29 (2013)
22. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: SURF: speeded up robust features. *Comput. Vis. Image Understand.* **3**(110), 122–131(2008)
23. Kolesenkov, A.N., Kostrov, B.V.: Method of thinning of basic functions in correlation-extreme algorithms of image matching. *Questions Ra-dioelectron.* **1**(1), 176–183 (2010)





# Method for Plant Leaves Square Area Estimation Based on Digital Image Analysis

Y. Proshkin<sup>1</sup>(✉) , A. Smirnov<sup>1</sup> , D. Burynin<sup>1</sup> , and V. Panchenko<sup>1,2</sup> 

<sup>1</sup> Federal Scientific Agroengineering Center VIM, 1st Institutsky Passage 5, 109428 Moscow, Russia

yproshkin@mail.com

<sup>2</sup> Russian University of Transport, Obraztsova St. 9, 127994 Moscow, Russia

**Abstract.** When examining plant reactions to external climatic and other factors, it is necessary to record and analyze a large number of plants morphological features. The research is aimed at the optimization of the existing applied methods for values obtainment of plants morphological characteristics based on digital image processing, at the obtained solutions adaption for use in non-destructive monitoring systems, and at the calculation of the average relative error of the developed method for the leaf area estimation. For the plant leaf morphological parameters finding out, the method was developed allowing to determine the area of the plant leaf based on the analysis of the digital images obtained by optical scanning and photographing with a digital camera. Scanning of the plants was carried out by the optical scanner with the resolution of 150, 300 and 600 dpi. Photographing was carried out with the digital camera with the matrix resolution of 18.2 million pixels. Based on the results of the work, the table was formed indicating the area of the analyzed objects. There were calculated the total area of the leaves of the studied plant groups and the relative deviation from the values obtained with aid of the planimeter. The results of the measurements obtained by scanning have a small average deviation relative to the planimeter readings. (The minimum average deviation makes 0.37% for the resolution of 300 dpi and the maximum 1.16% for the resolution of 150 dpi.) The relative average deviation in the case of the leaf area estimation with aid of the digital camera (0.67%) is not much different from the results obtained with the scanner.

**Keywords:** Digital image analysis · Leaves square area estimation · Plant monitoring · Phenotyping

## 1 Instruction

When examining plant reactions to external climatic and other factors, it is necessary to record and analyze a large number of plants morphological features. Traditionally, one of the key indicators is the square area of plant leaves [1–6]. In absence of specialized equipment (planimeters), the measurement of this parameter can cause significant difficulties, especially in case of large subsets, when a number of images reaches several hundreds.

The manual measurement techniques and methods for the leaf area estimation include as follows: the leaf area calculation by the largest length and width, the leaf shape projection onto millimeter paper, and the leaf area calculation by the ratio of its specific index and mass [7–9].

The above mentioned techniques and methods of the leaf area estimation have two main disadvantages: the large complexity in the measurement process and the inadequate accuracy of the results obtained, which is either low or highly dependent on the human factor. At present time, the most attractive methods are those based on the optical or laser scanning with the subsequent software processing of the obtained digital images [10–15]. There has been developed a number of software products and research reports [16–19] that allow to estimate the values of the leaf areas and other morphological characteristics of a plant by the digital images software-based analysis. Those technical and software solutions allow to automate the procedures of the plants morphological parameters obtainment with high accuracy. On the other hand, it should be noted that the offered programs are provided on a paid basis, and, in addition, not always, the methods described in the articles are applicable; actually, they have a number of conventions and restrictions. Thus, as the plants morphological parameters determination is concerned, timely relevant is the task of both optimization of the existing estimation techniques and methods and the creation of better universal, adaptive and generally available methods.

The aim of the study is to optimize the existing applied methods for obtaining the values of the morphological characteristics of plants based on digital image processing, adapt the obtained solutions for use in non-destructive testing systems and calculate the average relative error.

In the course of the study, a technique for estimating the area of leaves, an algorithm and recommendations for its application were developed. Comparative tests of the developed method for estimating the leaf area with other two optical measurement methods were also carried out, the results of the work are given in the Sect. 3.

The developed method can be used in platforms for plant phenotyping, and in non-destructive testing systems, as well as for applied research.

## 2 Methods and Materials

For the plant leaf morphological parameters finding out, the method was developed allowing to determine the area of the plant leaf based on the analysis of the digital images obtained by optical scanning and photographing with a digital camera.

Scanning of the plants was carried out by the optical scanner with the resolution of 150, 300 and 600 dpi. Photographing was carried out with the digital camera with the matrix resolution of 18.2 million pixels.

The studied objects were leaves of the basil plant (*Ócimum*) divided into 2 subset groups depending on the variety of the plant, so that the group No.1 included the sweet basil and the group No. 2 the red basil.

The measurement of the leaves area was carried out for two subsets of the plant leaves harvested on the 60th day from the cultivation beginning. The process consisted of the following main steps:

1. Sampling (samples collection);
2. Samples preparation for digital imaging;
3. Getting digital images;
4. Digital image analysis;
5. Report preparation.

The selection and separation of the samples groups for the research was carried out as part of the effect studying of the ultraviolet radiation on content and concentration of the essential oils in the leaves of both the sweet basil and the red basil plants.

All selected samples were digitized using the digital camera and the scanner.

In order to optimize the well-known technique of the plant leaves area estimation and to make recommendations, two sources of the digital images obtainment were used for different measurement conditions: in the field (where the digital camera was used) and in the laboratory (the scanner).

When measuring in the laboratory, the separated plant leaves were laid down onto the scanner glass at the certain distance from each other equal to about 1 cm. Optical scanning was performed on the scanner CANON Canoscan LIDE400 at the resolution values 150, 300 and 600 dpi.

When in the field, for the leaf samples preparation before the digital images obtainment, the separated plant leaves were laid down onto a prepared base, which was a white paper sheet A4 with a red line 5 mm thick running along its edges. In order to facilitate the subsequent analysis of the digital images, the leaves were laid out at the certain distance from each other (about 1 cm). This enabled simplifying the processing algorithm and counting the total number of samples on the stage of the report preparation. Also onto the white base, a round black stencil of a known square area was laid out.—This was done in order to determine the scale of the samples easily. The prepared samples were photographed with the digital camera SONY Cyber-shot DSC-WX350 (with the matrix resolution of 18.2 million pixels) at a distance of about 1 cm.

In frame of the well-known leaves area estimation techniques, a number of operations are performed immediately before the image analysis, which significantly complicates the process. One of the needed operations is pre-processing in the graphic editor Adobe Photoshop for the purpose of noise removal and background segments selection in order to create a background reference. In the proposed method, those operations are omitted, which significantly reduces the time spent on the leaves area estimation.

The obtained digital images were processed by a software script in the environment MatLab. This script splits the digital image into separate spectral channels: red, green, and blue. On the next step with aid of the function `im2bw`, this script converts the original grayscale image to a two-level binary image (BW) as follows: it replaces all pixels, whose brightness corresponds to the specified intensity range, with the white color (1), while all the other pixels are colored black (0). Then the binary images of an individual channel (red, green, or blue) are summed into one. Then for the convenience of white (1) pixels counting, with aid of the function `imcomplement`, the images inversion is performed. Now, with aid of the functions `imfill` and `strel`, gaps are filled in, and morphological

structural elements are created according to the specified parameters. Now the functions `bwlabel` and `regionprops` find and read the parameters of the found elements; and this allows to count the number of objects and to calculate their square area in pixels. Then, knowing the scale of the image and the area of the target objects in pixels, the script translates these values into square centimeters. Based on the results of the operations performed, a report is generated in the form of a processed digital image plus a table indicating the number of the objects found, their area and the area of all objects in general. This software script was developed based on the MatLab standard functions and application packages.

The above described algorithm was developed in order to analyze images obtained by optical scanning. As for the case of the images obtained with the digital photo-camera, at first, it is necessary to carry out a number of preparatory transformations. For beginning, the zone is determined, in which the samples will be analyzed. The selection of the boundaries of the desired zone is based on a principle similar to the one described above for the software script. Upon that, the most significant reference point is considered to be the red line 5 mm thick running along the contour of the white base. In the determined zone, the stencil is searched judging on objects' contour and color. After detecting an object with the specified parameters, the software determines its square area in pixels. As the area of the stencil is the pre-known value, it is possible to calculate the scale of the image. On the next stage, the image analysis is performed according to the algorithm given above. When getting the final result, the area of the stencil is not taken into account.

### 3 Results and Discussion

With the purpose of the simplification of the method of leaf area finding and the developed method adaptation for use in the field, the following three main tasks were set: (a) analysis of known methods [10–19], (b) reduction of a number of operations during the work due to the reviewing of the well-known algorithms in general and of the scanning results processing unit in particular, and (c) adaptation of this method for use in the field.

During the analysis, it was found that the leaf preparation and its image pre-processing in the graphic editor are the most time-consuming operations. In the majority of the programs and methods under consideration, during the scan results processing, the pixels of the background of the presented image are counted; upon that, the leaf area is determined by the difference between the total area of the image and the pixel area corresponding to the search criteria. Nevertheless, the obtained results are not suited well for analysis and making generalized conclusions as they will contain only the total leaves area and the average leaf area (if there is information on the number of samples taken).

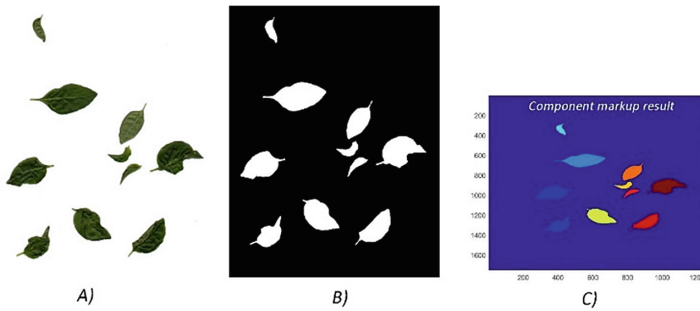
When processing images obtained with portable cameras, the images preparation time in the graphical editor will significantly increase: due to the heterogeneity of the illumination, there will be a lot of noise and background color distortion; always, under the leaves, there are areas with shadow and half-shadow. In most cases, in order to solve the problem of the image scale determination, special stencils of a known square area are used.

With due consideration of the above conclusions, in the developed method for the plant leaves area estimating, it is proposed firstly to perform recognizing of the leaf

samples themselves and secondly to fix the area for each of the two types. For this purpose, for processing of the scan results, the new algorithm was developed. The basic concept takes into account the fact that the background color will differ significantly from that of the plant leaves. As the background color is close to white, it is not difficult to impose boundary conditions for converting of an original halftone image into a two-level binary image (BW). It will be done by white-coloring (replacing) of all pixels, whose brightness corresponds to the specified intensity range, and by black-coloring of the rest of the pixels.

As preliminary this image was not processed in the graphic editor, it can contain some noise caused by the heterogeneity of the plant leaves. In order to eliminate these undulations, the following functions are used: `imfill` and `strel`. The function `imfill` makes filling-in of zones that meet the specified parameters; the function `strel` creates morphological structures of elements in order to smooth out the irregularities. For the function `imfill`, the property 'holes' is set, which allows to remove gaps in the image. For the function `strel`, morphological structures are defined as round ones by use of the property 'disk'; upon that, the disk radius is selected depending on the image expansion. After this is done, the individual elements are recognized by the specified parameters; and with aid of the functions `bwlabel` and `regionprops`, the characteristics of interest are read.

Figure 1 shows an example of the step-by-step conversion of a digital image.



**Fig. 1.** Example of step-by-step conversion of digital image: **a** initial image; **b** binary image; **c** layout of components and leaves area counting.

In its simplified form, the developed software script on the platform for programming and numerical calculations “MATLAB” will look like as follows:

```

I1 = imread('image.jpg');
% Reading the image
rmat1 = I1(:, :, 1); gmat1 = I1(:, :, 2); bmat1 = I1(:, :, 3);
levelr1 = 0.5; levelg1 = 0.5; levelb1 = 0.5;
% Setting the pixel brightness level criteria separately for the red, blue, and green
channels
i1 = im2bw(rmat1, levelr1); i2 = im2bw(gmat1, levelg1); i3 = im2bw(bmat1, levelb1);
I1sum = (i1&i2&i3);
% Obtaining a binary image

```

```

I1comp = imcomplement(I1sum);
% Inverting the binary image
Ifilled1 = imfill(I1comp,'holes'); se = strel('disk',2); Iopened1 = imopen(Ifilled1,se);
% Removing gaps and undulations
n = 8;
[L1,num] = bwlabel(Iopened1,n);
figure(3), imagesc(L1), colormap(jet), title ('Result of marking up of connected
components');
STATS1 = regionprops(L1,'Area'); for i = 1:num
STATS2(i,1) = i; % object number
STATS2(i,2) = STATS1(i,1).Area; SUM2(i,1) = STATS1(i,1).Area;
End SUM1 = sum(SUM2);
SUM3 = SUM1/scale
% Determining the square area of each individual leaf and their total sum

```

The scanner was calibrated using the calibration plate 9931-021 PERF CALIBR PLATE PACK (Fig. 2).



**Fig. 2.** Calibration plate 9931-021 PERF CALIBR PLATE PACK

The scanner was calibrated for the following resolution values: 150, 300 and 600 dpi. The received correction factors (shown in Table 1) were taken into account when calculating the average relative error. As control values, there were taken the results obtained on the planimeter LI-3100C Area Meter.

The software script adaptation for the leaves area estimation in the field conditions consisted in development of an additional unit for pre-processing of available digital images. This additional unit includes two stages: determination of the zone, in which the samples will be analyzed, and calculation of the image scale. The reference point for the boundaries recognition of the desired zone is the red line 5 mm thick running along the contour of the white base. In a case of absence of a suitable base, the preliminary preparation of an image can be carried out in the graphic editor (for example, Adobe Photoshop, Pixlr X, Photopea). The image scale is calculated based on size of a stencil placed with the selected samples. A search for a relevant stencil is fulfilled by its contour and color. As a result, in the final values, the area of the stencil is not taken into account.

The results obtained in this research are presented in Table 1.

In order to check the adequacy of the presented method, the comparative analysis was carried out with the computer program PETIOLE (<https://petioleapp.com/>) as this program is similar to our methods from point of view of its functions, is installed on

**Table 1.** Different measurement measuring mixed state estimation results.

Device	Calibration factor	Leaf area (cm <sup>2</sup> ) (1 group)	Relative deviation (%) (1 group)	Leaf area (cm <sup>2</sup> ) (2 group)	Relative deviation (%) (2 group)	Average Relative deviation (%)
Control (LI-3100C)	—	43.47	—	207.11	—	—
Scanner (150 dpi)	1.021	42.91	1.32	205.09	1.00	1.16
Scanner (300 dpi)	1.013	43.31	0.38	206.39	0.37	0.37
Scanner (600 dpi)	1.02	43.17	0.70	204.97	1.06	0.88
Digital camera (DSC-WX350)	—	43.73	0.60	207.11	0.74	0.67

mobile devices and is used for similar purposes [20, 21]. The objects of the study were the following 2 groups of plant leaf samples: dandelion leaves (*Taraxacum*) as the 1st group, and the parsley leaves (*Petroselinum*) as the 2nd group (3). The plants were scanned using the optical scanner CANON Canoscan LIDE400 at the resolution 300 dpi. Photographing was done with the digital camera with the matrix resolution 18.2 million pixels. The control was carried out on the planimeter LI-3100C Area Meter. The results obtained during the comparative tests are presented in Table 2.

The contraposition of the proposed method and the similar product enabled by these comparative tests showed the adequacy of the obtained values and allowed identification of the main positive and negative features of the proposed method. Its main advantages include as follows (Fig. 3):

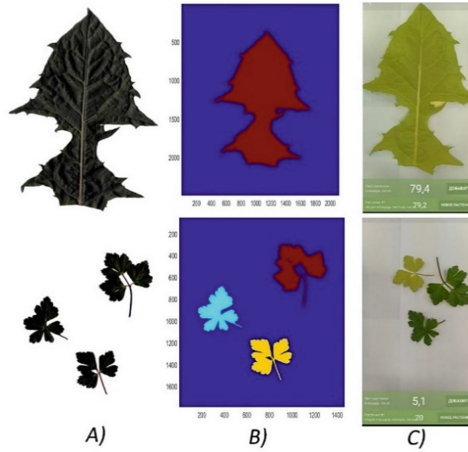
- high accuracy;
- universal functionality regardless of digital image sources (scanner, photo);
- possibility to accumulate and present the obtained data in tables;
- affordability of the proposed method with no need in expensive measuring devices.

The method disadvantages include:

- needed basic knowledge for work in the software environment MatLab or for programming languages such as Python, C++, etc.;
- strong dependence of measurement accuracy on leaves shapes and background shadows (though this disadvantage is typical for most part of the similar techniques).

## 4 Conclusions

Based on the results of the work, the table was formed indicating the area of the analyzed objects. There were calculated the total area of the leaves of the studied plant groups and the relative deviation from the values obtained with aid of the planimeter. The results of



**Fig. 3.** Comparative analysis of methods of digital image processing and leaf area estimation: **a** initial image; **b** leaf area estimation by the software-based script in the environment MatLab; **c** leaf area estimation in the program PETIOLE.

**Table 2.** Results of comparative tests at contraposition with analogues.

Device	Leaf area (cm <sup>2</sup> ) (1 group)	Relative deviation (%) (1 group)	Leaf area (cm <sup>2</sup> ) (2 group)	Relative deviation (%) (2 group)
Control (LI-3100C)	76.45	—	20.21	—
Scanner (300 dpi)	76.29	0.21	20.07	0.70
Digital camera (DSC-WX350)	77.76	1.69	20.37	0.79
PETIOLE	79.2	3.47	20.0	1.05

the measurements obtained by scanning have a small average deviation relative to the planimeter readings. (The minimum average deviation makes 0.37% for the resolution of 300 dpi and the maximum 1.16% for the resolution of 150 dpi.) This indicates the high accuracy and relevance of the proposed method.

The relative average deviation in the case of the leaf area estimation with aid of the digital camera (0.67%) is not much different from the results obtained with the scanner. From this fact, it is possible to conclude that the inaccuracy caused by the geometric distortions of the optical elements of the camera lens affects the results to such a small extent that it can be considered negligible.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



## References

- Zhang, Z., Huang, M., Zhao, X., Wu, L.: Adjustments of leaf traits and whole plant leaf area for balancing water supply and demand in *Robinia pseudoacacia* under different precipitation conditions on the Loess Plateau. *Agric. Forest Meteorol.* **279**, 107733 (2019). <https://doi.org/10.1016/j.agrformet.2019.107733>
- Teobaldelli, M., Basile, B., Giuffrida, F., Romano, D., Toscano, S., Leonardi, C., Rivera, C.M., Colla, G., Roupshael, Y.: Analysis of cultivar-specific variability in size-related leaf traits and modeling of single leaf area in three medicinal and aromatic plants: *Ocimum basilicum* L., *Mentha* Spp. and *Salvia* Spp. *Plants-Basel*, **9**(1), 13 (2019). <https://doi.org/10.3390/plants9010013>
- Mencuccini, M., Rosa, T., Rowland, L., Choat, B., Cornelissen, H., Jansen, S., Kramer, K., Lapenis, A., Manzoni, S., Niinemets, U., Reich, P., Schrod, F., Soudzilovskaia, N., Wright, I.J., Martinez-Vilalta, J.: Leaf economics and plant hydraulics drive leaf: wood area ratios. *New Phytol.* **224**, 1544–1556 (2019). <https://doi.org/10.1111/nph.15998>
- Gong, H., Gao, J.: Soil and climatic drivers of plant SLA (specific leaf area). *Global Ecol. Conserv.* **20**, e00696 (2019). <https://doi.org/10.1016/j.gecco.2019.e00696>
- Tatsumi, K., Kuwabara, Y., Motorayashi, T.: Monthly variability in the photosynthetic capacities, leaf mass per area and leaf nitrogen contents of rice (*Oryza sativa* L.) plants and their correlations. *J. Agric. Meteorol.* **75**(2), 111–119 (2019). <https://doi.org/10.2480/agrmet.D-18-00043>
- Dorokhov, A.S., Smirnov, A.A., Semenova, N.A., Akimova, S.V., Kachan, S.A., Chilingaryan, N.O., Glinushkin, A.P., Podkovyrov, I.Y.: The effect of far-red light on the productivity and photosynthetic activity of tomato. *IOP Conf. Ser. Earth Environ. Sci.* **663**, 012044 (2021). <https://doi.org/10.1088/1755-1315/663/1/012044>
- Hinnah, F.D., Heldwein, A.B., Heldwein, I.C., Loose, L.H., Lucas, D.D.P., Bortoluzzi, M.P.: Estimation of eggplant leaf area from leaf dimensions. *Bragantia* **73**(3), 213–218 (2014). <https://doi.org/10.1590/1678-4499.0083>
- Huang, W., Ratkowsky, D.A., Hui, C., Wang, P., Su, J., Shi, P.: Leaf fresh weight versus dry weight: which is better for describing the scaling relationship between leaf biomass and leaf area for broad-leaved plants? *Forests* **10**(3), 256 (2019). <https://doi.org/10.3390/f10030256>
- Donato, L.T.F., Donato, S.L.R., Brito, C.F.B., Fonseca, V.A., Gomes, C.N., Rodrigues, V.A.: Estimating leaf area of prata-type banana plants with lanceolate type leaves. *Revista Brasileira de Fruticultura*, **42**(4), (2020). <https://doi.org/10.1590/0100-29452020417>
- Itakura, K., Hosoi, F.: Voxel-based leaf area estimation from three-dimensional plant images. *J. Agric. Meteorol.* **75**(4), 211–216 (2019). <https://doi.org/10.2480/agrmet.D-19-00013>
- Valle, B., Simonneau, T., Bouldard, R., Sourd, F., Frisson, T., Ryckewaert, M., Hamard, P., Bricchet, N., Dauzat, M., Christophe, A.: PYM: a new, affordable, image-based method using a Raspberry Pi to phenotype plant leaf area in a wide diversity of environments. *Plant Methods* **13**, 98 (2017). <https://doi.org/10.1186/s13007-017-0248-5>
- An, N., Palmer, C.M., Baker, R.L., Markelz, R.J.C., Ta, J., Covington, M.F., Maloof, J.N., Welch, S.M., Weinig, C.: Plant high-throughput phenotyping using photogrammetry and imaging techniques to measure leaf length and rosette area. *Comput. Electron. Agric.* **127**, 376–394 (2016). <https://doi.org/10.1016/j.compag.2016.04.002>
- Neinavaz, E., Skidmore, A.K., Darvishzadeh, R., Groen, T.A.: Retrieval of leaf area index in different plant species using thermal hyperspectral data. *ISPRS J. Photogramm. Remote Sens.* **119**, 390–401 (2016). <https://doi.org/10.1016/j.isprsjprs.2016.07.001>
- Cargnelutti, A., Toebe, M., Burin, C., Fick, A.L., Neu, I.M., Facco, G.: Leaf area estimation of velvet bean through non destructive method. *Ciência Rural* **42**(2), 238–242 (2012). <https://doi.org/10.1590/S0103-84782012000200009>

15. Smirnov, A., Proshkin, Y., Sokolov, A., Dorokhov, A.: Portable spectral device for monitoring plant stress conditions. *E3S Web Conf.* **210**, 05016 (2020). <https://doi.org/10.1051/e3sconf/202021005016>
16. Yang, Z., Han, Y.: A low-cost 3d phenotype measurement method of leafy vegetables using video recordings from smartphones. *Sensors* **20**(21), 6068 (2020). <https://doi.org/10.3390/s20216068>
17. Xu, S., Wang, J., Tian, H., Wang, B.: Automatic measuring approach and device for mature rapeseed's plant type parameters. *J. Elect. Comput. Eng.* **2019**, 1–10 (2019). <https://doi.org/10.1155/2019/6834290>
18. Hang, T., Lu, N., Takagaki, M., Mao, H.: Leaf area model based on thermal effectiveness and photosynthetically active radiation in lettuce grown in mini-plant factories under different light cycles. *Sci. Hortic.* **252**, 113–120 (2019). <https://doi.org/10.1016/j.scienta.2019.03.057>
19. Cortazar, B., Koydemir, H., Tseng, D., Tseng, S., Ozcan, A.: Quantification of plant chlorophyll content using Google Glass. *Lab Chip* **15**(7), 1708–1716 (2015). <https://doi.org/10.1039/c4lc01279h>
20. Hrytsak, L.R., Herts, A.I., Nuzhyna, N.V., Cryk, M.M., Shevchenko, V.V., Drobyk, N.M.: The influence of light regime on the growth data and pigment composition of the plant *Gentiana lutea* cultured in vitro. *Regul. Mech. Biosyst.* **9**(2), 258–266 (2018). <https://doi.org/10.15421/021838>
21. Polunina, O.V., Maiboroda, V.P., Seleznov, A.Y.: Evaluation methods of estimation of young apple trees leaf area. *Bull. Uman Natl. Univ. Hortic.* **2**, 80–83 (2018). <https://doi.org/10.31395/2310-0478-2018-21-80-82>



# Digital Revolution Through Computational Intelligence: Innovative Applications and Trends

Ramsagar Yadav<sup>1</sup>(✉), Mukhdeep Singh Manshahia<sup>1</sup>, and M. P. Chaudhary<sup>2</sup>

<sup>1</sup> Department of Mathematics, Punjabi University Patiala, Punjab, India  
ramsagar.yadav@lsraheja.org

<sup>2</sup> International Scientific Research and Welfare Organization, New Delhi, India

**Abstract.** Computational Intelligence is sometimes also referred to as soft computing, which is a specific field of study where the task is to make computers learn some real-life or complex problems from the experimental data or observations. In computational intelligence, there is a set of approaches or methodologies used to address real-life or complex problems. Generally, it is impossible to solve real-life problems using traditional computing methods because of complexity, uncertainty, or problems that don't have a proper definition. Considering cognitive computing is an indispensable technology to develop these smart systems, this paper proposes human-centered computing assisted by cognitive computing and cloud computing. Computational Intelligence has significantly extended the possibility of computing, encompassing it from traditional computing on data to progressively diverse computing paradigms such as cognitive intelligence and human-computer fusion intelligence. Intelligence and computing have undergone paths of different evolution and development for a long time but have become increasingly intertwined in recent years. Intelligent computing is not only intelligence-oriented but also intelligence-driven. Such cross-fertilization has prompted the emergence and rapid advancement of intelligent computing.

**Keywords:** Data intelligence · Computing architectures and paradigms · Computational intelligence · Neural networks · Deep learning · Meta-heuristics · Real-world applications

## 1 Introduction

Computational intelligence is a new and modern tool for explaining difficult problems which are hard to be solved by the conventional techniques. Heuristic optimization techniques are broad purpose methods that are very flexible and can be applied to many types of objective functions and restrictions. In recent times, these new heuristic tools have been joined among themselves and new methods have emerged that combine elements of nature-based techniques or which have their foundation in stochastic and simulation methods [1–4].

Computational intelligence techniques comprise the use of computers to facilitate machines to mimic human performance. The protuberant paradigms used include AI

systems, fuzzy logic, artificial neural networks, evolutionary computing techniques, artificial life, computer vision, adaptive intelligence, and chaos engineering. These knowledge-based computational intelligence techniques have generated terrific interest among scientists and application engineers due to a number of benefits such as generalization, adaptation, fault tolerance and self-repair, self-organization and evolution. Successful demonstration of the applications of knowledge-based systems theories will aid scientists and engineers in finding sophisticated and low cost solutions to difficult problems [5–8].

## 2 The Five Main Principles of CI and Its Applications

The main applications of Computational Intelligence include computer science, engineering, data analysis and bio-medicine.

### 2.1 Fuzzy Logic

As explained before, fuzzy logic, one of CI's main principles, consists in measurements and process modeling made for real life's complex processes. It can face incompleteness, and most importantly ignorance of data in a process model, contrarily to Artificial Intelligence, which requires exact knowledge.

This technique tends to apply to a wide range of domains such as control, image processing and decision making. But it is also well introduced in the field of household appliances with washing machines, microwave ovens, etc. We can face it too when using a video camera, where it helps stabilizing the image while holding the camera unsteadily. Other areas such as medical diagnostics, foreign exchange trading and business strategy selection are apart from this principle's numbers of applications.

Fuzzy logic is mainly useful for approximate reasoning, and doesn't have learning abilities, a qualification much needed that human beings have. It enables them to improve themselves by learning from their previous mistakes [9].

### 2.2 Neural Networks

This is why CI experts work on the development of artificial neural networks based on the biological ones, which can be defined by 3 main components: the cell-body which processes the information, the axon, which is a device enabling the signal conducting, and the synapse, which controls signals. Therefore, artificial neural networks are doted of distributed information processing systems, enabling the process and the learning from experiential data. Working like human beings, fault tolerance is also one of the main assets of this principle.

Concerning its applications, neural networks can be classified into five groups: data analysis and classification, associative memory, clustering generation of patterns and control. Generally, this method aims to analyze and classify medical data, proceed to face and fraud detection and most importantly deal with nonlinearities of a system in order to control it. Furthermore, neural networks techniques share with the fuzzy logic ones the advantage of enabling data clustering [10].

### **2.3 Evolutionary Computation**

Based on the process of natural selection firstly introduced by Charles Robert Darwin, the evolutionary computation consists in capitalizing on the strength of natural evolution to bring up new artificial evolutionary methodologies. It also includes other areas such as evolution strategy, and evolutionary algorithms which are seen as problem solvers... This principle's main applications cover areas such as optimization and multi-objective optimization, to which traditional mathematical one techniques aren't enough anymore to apply to a wide range of problems such as DNA Analysis, scheduling problems [11].

### **2.4 Learning Theory**

Still looking for a way of "reasoning" close to the humans' one, learning theory is one of the main approaches of CI. In psychology, learning is the process of bringing together cognitive, emotional and environmental effects and experiences to acquire, enhance or change knowledge, skills, values and world views. Learning theories then helps understanding how these effects and experiences are processed, and then helps making predictions based on previous experience [12].

### **2.5 Probabilistic Methods**

Being one of the main elements of fuzzy logic, probabilistic methods firstly introduced by Paul Erdos and Joel Spencer, aim to evaluate the outcomes of a Computation Intelligent system, mostly defined by randomness. Therefore, probabilistic methods bring out the possible solutions to a problem, based on prior knowledge [13].

### **2.6 Issues of Traditional Computing**

The conventional computing functions logically with a set of rules and calculations. Conventional computing is often unable to manage the variability of data obtained in the real world. In conventional computing, programmer tells the system exactly how to solve the problem. Conventional computing can solve only one problem at a time in a given domain [14].

## **3 Digital Revolution and Artificial Intelligence**

Digital disruption means the rise of new technology and the effect that has on existing models and products. Digital disruption has an impact on businesses and economies, but AI takes it to next level with machine learning and big data which gives us insights for decision making with deep learning which haven't been experienced in other disruptions [15].

## 4 Big Data

Computational intelligence (CI) enables choice makers to develop and analyze the apprehended facts computationally and consequently to identify and illuminate the underlying patterns of the data, as well as to proficiently learn the precise tasks. CI insures a broad range of nature-inspired, multidisciplinary and computational approaches, such as fuzzy logic, artificial neural networks, evolutionary computing, learning theory, probabilistic methods, and so on. CI technologies are estimated to deliver effective and great tools that scale well with data volume for big data analytics and method, while addressing the trials brought by the massive amount of data [16–18].

## 5 Artificial Intelligence (AI) and Computational Intelligence (CI)

Artificial Intelligence (AI) is the study of intelligent conduct established by machines as divergent to the natural intelligence in human beings. It is an area of computer science that is apprehensive with the development of a technology that allows a machine or computer to reason, act, or perform in a more humane way. Computational Intelligence (CI), on the other hand, is more like a sub-branch of AI that accentuates on the design, application and advancement of linguistically inspired computational models. It is the study of adaptive contrivances to assist or expedite intelligent behavior in multifaceted and changing situations [19–21].

The expansion of CI methods follows a different pathway than that of the AI. AI aims to construct intelligent machines which can demonstrate intelligent behavior and which can contemplate and absorb like human beings. CI, on the other hand, is a subclass of AI which highlights on computational paradigms that create intelligent behavior conceivable in natural or artificial systems in difficult and fluctuating settings. Although, both AI and CI seek almost similar objectives, they are quite dissimilar from each other [22, 23].

## 6 Internet of Things (IoT) and Computational Intelligence (CI)

Computational intelligence conclusions, precision, and delicacy at different layers in IoT systems are monitored by computational intelligence techniques, which are accountable for data gathering, the interconnection amongst devices and the internet, data processing, and decision-making, without human interface. Therefore, super-fast computing with a desired thoroughgoing accuracy level has always been critical in the swift development of IoT applications [24, 25].

Development and realization of various machine learning and optimization methods, preferably with maximum accuracy and precision ratio, are the backbone of computational intelligence and smart IoT-based healthcare systems. Therefore, the integration of new learning and assessment methods is an emerging trend of great significance in next-generation of computational intelligence and IoT-based applications. These advancements in computational intelligence approaches can significantly enhance sustainability without compromising the quality of service. However, the vast use of these devices for next-generation IoT applications generates massive amounts of data, which can significantly deteriorate the computing efficiency [26–28].

## 7 Computational Intelligence as a New Paradigm

In the past era, Artificial Intelligence had an outstanding paradigm shift from its domain of representative to non-symbolic and numeric computation. Prior to the mid-eighties, symbolic logic was used as the exceptional tool in the advancement of algorithms for the classical AI problems like reasoning, planning, and machine learning. The incompleteness of the customary AI was shortly realized, but inappropriately no handy solutions were readily available at the time. In the nineties the enormous developments in fuzzy logic, artificial neural nets, genetic algorithms and probabilistic reasoning models motivated the researchers around the world to discover the possibilities of building more humanlike machines using these new tools [29–33].

## 8 Innovative Applications

In the present era of investigation and knowledge, several emergent concepts like Wireless Sensor Networks, Body Wireless Sensor Networks, Fog, Edge, and Big Data Analytics can assist the design and development of intelligent systems in miscellaneous domains, e.g., transportation, education, enterprise, and industry, etc. Wireless sensor network technologies are one of the key research areas in computer science. Recent Intelligent Computing Techniques have provided unexpected solutions for wireless network applications. Wireless network applications, such as real-time traffic data, sensor data from driverless cars, or entertainment streaming recommendations, generate enormous quantities of data for real-time collection and processing [34–36].

Computational Intelligence Techniques can be used as follows [37]:

- Intelligent Computing Techniques for mobile network design
- Intelligent Computing Techniques for wireless network applications
- Innovative intelligent computing architecture/algorithms for wireless networks
- Intelligent Computing Techniques in industrial-level systems.

## 9 Conclusion

The field of computational intelligence has developed enormously over that past five years, thanks to evolving soft computing and artificial intelligent methodologies, tools and techniques for visualizing the quintessence of intelligence entrenched in real life explanations. Therefore, scientists have been able to expound and comprehend real life processes and practices which earlier often remain unexplored by virtue of their underlying inaccuracy, uncertainties and dismissals, and the unapproachability of appropriate methods for describing the incompleteness and indistinctness of information represented. With the arrival of the field of computational intelligence, researchers are now able to explore and unearth the intelligence, otherwise overwhelming, embedded in the systems under consideration. Computational Intelligence is now not limited to only precise computational fields, it has prepared.

**Acknowledgements.** Authors are grateful to Punjabi University, Patiala for providing adequate library and internet facility.

## References

1. Smith, J.D., Johnson, A.B.: Computational intelligence: a review of recent advances in heuristic optimization techniques. *J. Artif. Intell. Res.* **52**, 789–813 (2018)
2. Lee, C.H., Chen, Y.H.: Nature-inspired computational intelligence: a comprehensive survey. *IEEE Trans. Evol. Comput.* **23**(2), 256–277 (2019)
3. Wang, G., Deb, S., Coello Coello, C.A.: Exploring the evolutionary search space: a survey on benchmark problems and evaluation measures for real-parameter optimization. *IEEE Comput. Intell. Mag.* **12**(3), 57–76 (2017)
4. Verma, A., Singh, D.: A review on the applications of computational intelligence techniques in computer vision. *Pattern Recogn. Lett.* **130**, 204–215 (2020)
5. Hsu, L.C., Lee, E.S.: Adaptive intelligence for fault tolerance and self-repair in complex systems: a survey. *IEEE Trans. Syst. Man Cybern. Syst.* **48**(1), 54–69 (2018)
6. Sareen, S., Choudhury, T.: Computational intelligence for self-organizing systems: a comprehensive survey. *IEEE Trans. Cogn. Dev. Syst.* **11**(3), 384–398 (2019)
7. Chen, S., Xu, X.: Chaos engineering: approaches, techniques, and applications. *ACM Comput. Surv.* **49**(2), 33:1–33:32 (2016)
8. Kumar, S., Sharma, P.: Recent trends and applications of artificial neural networks in computational intelligence: a survey. *Appl. Soft Comput.* **57**, 591–607 (2017)
9. Smith, J.D., Johnson, R.W.: Fuzzy logic applications in control theory and image processing. *IEEE Trans. Fuzzy Syst.* **16**(2), 418–435 (2008)
10. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *Science* **313**(5786), 504–507 (2006)
11. Eiben, A.E., Smith, J.E.: *Introduction to Evolutionary Computing*. Springer (2015)
12. Ormrod, J.E.: *Human Learning*, 4th edn. Prentice Hall (1999)
13. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer (2006)
14. Harel, D.: On visual formalisms. *Commun. ACM* **28**(7), 620–631 (1985)
15. Smith, A., Johnson, R.: The impact of artificial intelligence in the digital revolution. *J. Digital Disrup.* **12**(3), 45–62 (2020)
16. Chen, M., Mao, S., Liu, Y.: Big data: a survey. *Mobile Netw. Appl.* **19**(2), 171–209 (2014)
17. Jain, A.K., Chandrasekaran, M.: Big data analytics: a survey. *ACM Comput. Surv.* **47**(2), 1–36 (2016)
18. Zhang, L., Zhang, J., Zhang, G.: Big data analytics with computational intelligence. *IEEE Access* **4**, 2924–2940 (2016)
19. Russell, S.J., Norvig, P.: *Artificial Intelligence: A Modern Approach*. Pearson (2016)
20. Pedrycz, W., Gomide, F.: *Fuzzy Systems Engineering: Toward Human-Centric Computing*. Wiley (2017)
21. Engelbrecht, A.P.: *Computational Intelligence: An Introduction*. Wiley (2007)
22. Samanta, D.: *Artificial Intelligence and Machine Learning*. CRC Press (2017)
23. Yegnanarayana, B.: *Artificial Neural Networks*. PHI Learning Pvt. Ltd. (2009)
24. Zhang, Y., Liu, Q., Wang, J.: Computational intelligence in internet of things: methods, algorithms, and applications. *Futur. Gener. Comput. Syst.* **86**, 936–945 (2018)
25. Rahmani, A.M., Thanigaivelan, N.K., Gia, T.N., Granados, J.: Computational intelligence techniques for IoT and big data. *IEEE Internet Things J.* **6**(2), 1616–1627 (2019)
26. Wang, Z., Xu, Y., Wang, J., Zhang, Y., Gu, T.: Computational intelligence for internet of things: a survey. *ACM Trans. Internet Things* **1**(1), 1–26 (2020)
27. Oussous, A., Benhlilima, L., Ait Lahcen, A., Belfkih, S.: Computational intelligence techniques for Internet of Things security: a comprehensive review. *J. Ambient. Intell. Humaniz. Comput.* **11**(8), 3247–3264 (2020)



28. Mahmood, A.N., Yaqoob, I., Anpalagan, A., Huh, E.N., Ahmed, S.H.: Big data analytics for IoT-cloud supported real-time computational intelligence: a survey. *IEEE Access* **6**, 66336–66353 (2018)
29. Kurzweil, R.: *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*. Penguin (2000)
30. Sotelo, M.A., Fernández, A., Pelayo, F., Prieto, A., Karray, F.: Artificial intelligence in the twenty-first century. *ACM Comput. Surv.* **50**(5), 1–48 (2017)
31. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley (1989)
32. Luger, G.F.: *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*. Pearson Education (2012)
33. Mitchell, M.: *Machine Learning*. McGraw-Hill (1997)
34. Chen, C.M., Zhang, C.: Intelligent computing techniques for mobile network design. *IEEE Trans. Mob. Comput.* **18**(9), 2057–2072 (2019)
35. Karim, M.R., Shawkat, S.M., Akram, M.T., Pathan, A.S.K.: Intelligent computing techniques for wireless network applications: a survey. *IEEE Commun. Surv. Tutorials* **20**(2), 1226–1253 (2018)
36. Wang, C., Zhang, H., Song, J., Li, Y.: Innovative intelligent computing architecture for wireless networks. *IEEE Trans. Indus. Inf.* **16**(4), 2581–2593 (2020)
37. Dey, N., Ashour, A. S., Thampi, S. M. (Eds.): *Innovative computing and communication: second international conference, ICICC 2017, Kochi, India, August 17–18, 2017, Proceedings (Vol. 639)*. Springer (2017)



# Compressive Sensing and Orthogonal Matching Pursuit-Based Approach for Image Compression and Reconstruction

Sai Sylesh Gupta Namburu, Nandu Vasudevan, Vudhya Muni Sai Karthik<sup>(✉)</sup>,  
M. Nimal Madhu, and V. Hareesh

Center for Computational Engineering and Networking (CEN), Amrita Vishwa Vidyapeetham,  
Coimbatore, India

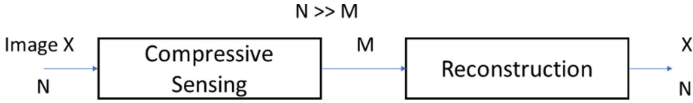
karthikvms95@gmail.com

**Abstract.** The focus of the paper is to reconstruct the images with a lower number of measurements by utilizing compressive sensing as the core architecture. The existing technologies utilize multiple frameworks to generate a compressed image using fewer measurements. However, a combination of compressive sensing, essential matrix operations, and an orthogonal matching pursuit (OMP) can generate a compressed image with a lower number of measurements and faster processing time. Key concepts include Compressive Sensing, Sparse Representation, Measurement Matrix, and Sensing Matrix.

**Keywords:** Compressive sensing (CS) · Random measurement process · Sparse representation · Measurement matrix · Orthogonal matching pursuit (OMP)

## 1 Introduction

Compressive sensing (CS) is a mathematical framework that enables the acquisition of signals that are sparse or compressible in some basis or dictionary, using fewer measurements than traditional sampling techniques. In compressive sensing, instead of directly measuring all the values of a signal, only a small number of linear projections are taken, and an algorithm is used to recover the original signal from these projections. This can result in a significant reduction in the number of measurements needed, making it useful for a variety of applications, especially in fields like imaging, audio processing, and communication systems. Figure 1 [1] depicts a traditional compressive sensing framework, where an image ( $X$ ) with  $N$  measurements is passed to a compressive sensing block that generates compressed data with  $M$  measurements, and the original image is reconstructed by utilizing reconstruction algorithms. Further, it is always ensured that the size of  $M$  should always be less than the size of  $N$ .



**Fig. 1.** Compressive sensing framework.

Compressive Sensing, as demonstrated in various studies [2], is a promising technique for data acquisition and compression. CS streamlines this conventional framework by merging the processes of sampling and compression.

In compressive sensing, the sparse representation of a signal is a key concept. It assumes that the signal can be represented as a linear combination of a few, or sometimes only one, basis elements from a set of basis functions known as a dictionary.

Mathematically, this can be expressed as [3]:

$$X = \Psi\alpha \quad (1)$$

where  $X$  is the original signal,  $\Psi$  is the dictionary matrix and  $\alpha$  is the sparse representation of  $X$  in the form of a coefficient vector with relatively few non-zero entries. Compressive sensing aims to reconstruct  $X$  from a small number of linear projections of  $X$ , which are represented by a measurement vector/measurement matrix ( $\Phi$ ) [4].

$$Y = \Phi\Psi\alpha \quad (2)$$

The sparsest vector  $\alpha$  that fulfills this measurement equation may be obtained. After resolving the sparsest Fourier coefficient vector, you may use the inverse Fourier transform to reconstruct the picture. The sparse representation  $\alpha$  can be recovered from  $Y$  and  $\Phi$  using optimization techniques.

A compressive sensing method for image reconstruction is presented in this paper. The focus is on accurately reconstructing the original image with a limited number of compressed measurements, which is determined by the design of the measurement matrix and the basis used for sparse representation. The orthogonal matching pursuit algorithm is utilized for image reconstruction.

## 2 Literature Review

Donoho [2], disclose a theory of compressive sensing, which demonstrated that sparse signals can be obtained using a reduced number of random measurements and explains how to use a non-linear approach to rebuild an unknown vector that represents the digital image or a signal. The number of measurements needed to reconstruct may be significantly lower than the image's actual dimensions only if the vector can be compressed via transform coding. Cands et al. [5] disclose that in compressive sensing, the sparse representation of a signal is a key concept. It assumes that the signal can be represented as a linear combination of a few, or sometimes only one, basis elements from a set of basis functions known as a dictionary. Yang et al. [6] proposed an improved Compressed Sensing (CS) algorithm that reduces the amount of raw data needed for high-resolution Synthetic Aperture Radar (SAR) imaging. Aharon et al. [7] disclose

that sparse representation of signals using overcomplete dictionaries which have gained significant attention in recent years. The focus has mainly been on pursuit algorithms that decompose signals concerning a given dictionary. This paper proposes a novel K-SVD algorithm for adapting dictionaries to achieve the best sparse signal representation.

Tropp and Gilbert [8] paper discloses that Orthogonal Matching Pursuit (OMP) is a reliable algorithm for signal recovery, requiring only  $O(m \log d)$  random linear measurements to recover a signal with  $m$  nonzero entries in dimension  $d$ . Jayaraman et al. [3] disclose that a signal  $X$  in  $\mathbb{R}^N$  can be represented by a sparse representation matrix  $\Psi$  ( $N \times N$ ), where  $X$  can be expressed as the sum of sparse coefficients ( $\alpha$ ) multiplied by the elements ( $\Psi_i$ ) of  $\Psi$ . This representation is commonly used in image data, where a fixed basis such as a wavelet basis is employed. The equation for this representation  $X = \Psi\alpha$ .

Li and Guoan [9] present an improved compressive sensing algorithm for image reconstruction based on wavelet transform. The algorithm improves the PSNR of reconstructed images by using the Hadamard matrix. Vahid [10] proposes a methodology, to improve the measurement matrices for compressive sampling of signals. Low mutual coherence between the measurement matrix and the representation matrix is required for a successful CS. This paper offers a gradient descent method for improving the measurement matrix. The method decreases mutual coherence by reducing the absolute off-diagonal members of the linked Gram matrix. In tests, the given method beats older methods under unoptimized conditions for sparse signals and random Gaussian matrices. In an underdetermined system of linear equations,  $y = C\Psi\alpha$  as a sparse solution, according to Donoho et al. [4]. The study shows that the sparsest solution of the linear system may be found using Stagewise OMP. A certain number of common linear algebraic operations are used to compress the signal using STOMP into a minimal residual, allowing different coefficients to enter the model at different stages.

Stagewise Orthogonal Matching Pursuit (STOMP) offers a solution for systems with typical/random distributions. Tropp [11] proposes a methodology for finding a signal's approximate value with  $T$  elementary signals is the task of simple sparse approximation. By simultaneously approximating many input signals with the same  $T$  elementary signals, simultaneous sparse approximation goes beyond this. Our approach combines compressive sensing with OMP for reconstructing the image with an optimal number of samples.

### 3 Proposed Approach

We use our dataset which contains color images with different kinds of lighting, textures, quality, and sharpness. To obtain a sparse representation of an image, we use a combination of fixed and adaptive bases, such as the Discrete Fourier basis. The image is transformed into a domain represented by  $\Psi$ , resulting in its sparse representation. The sparse representation is obtained by performing a matrix operation between the DCT basis, the Input Image, and the transpose of the DCT basis [1] (Fig. 2).

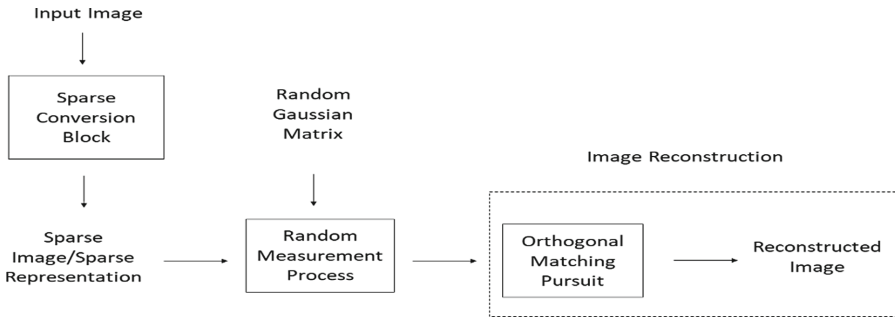


Fig. 2. Compressive sensing framework

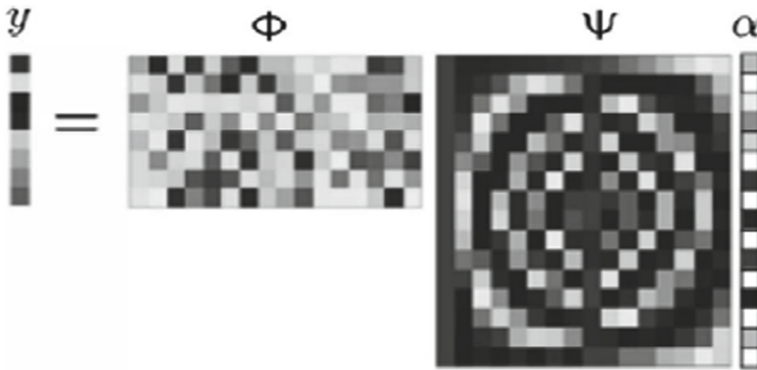


Fig. 3. Random measurement process [1]

Figure 3 explains the process of compressive sensing by creating a random measurement process, denoted as  $\Phi$ , which has a size of  $M \times N$ , where  $M$  is much smaller than  $N$  [1]. A random matrix is initially created using random Gaussian values. The random measurement is generated by performing matrix multiplication with the generated sparse representation and random Gaussian values. An optimization process is discussed in a subsequent section. During compression, the sparse data is measured using the random measurement process, resulting in a  $Y$  vector of size  $M \times 1$ , wherein the  $M$  is the required regeneration sample. Finally, the original image  $X$  is reconstructed from the compressed measurements  $Y$  using a reconstruction algorithm such as Orthogonal Matching [4, 11].

$$\min \|\alpha\|_1 \text{ such that } \|y - \Psi\alpha\|_2 \tag{3}$$

Orthogonal Matching Pursuit (OMP) was introduced by Tropp and Gilbert [11], an algorithm for recovery of a sparse vector from random linear measurements. It is a greedy algorithm that iteratively selects the most significant coefficients from the measurement vector and adds them to the sparse representation until a desired level of reconstruction accuracy is achieved.

At each iteration of OMP, the algorithm selects the basis function from the dictionary that is most correlated with the residual error between the measurement vector and the current estimate of the sparse representation. This basis function is then added to the current sparse representation and the process repeats until a stopping criterion is met. OMP has several advantages, including low computational complexity, and ease of implementation. However, OMP can be sensitive to the choice of the dictionary and may not converge to the optimal solution in some cases. Nevertheless, it is a widely used algorithm in compressive sensing and has been shown to provide good reconstruction results in many applications.

Here is the algorithm for the Orthogonal Matching Pursuit (OMP) [3, 11] method used in this paper in a step-by-step format:

1. Initialize:
  - i. Set the size of the original signal  $m$ .
  - ii. Set the number of iterations as  $\text{floor}(\text{length}(y)/4)$ .
  - iii. Initialize an empty solution  $\hat{x}$  with all elements equal to 0.
  - iv. Initialize an empty set of highly correlated columns  $\text{Aug}_t$ .
  - v. Set the residual  $r_n$  equal to  $y$ .
2. For each iteration from 1 to  $s$ :
  - i. Compute the product of the transpose of  $T\_Mat$  and  $r_n$  and store it in the product.
  - ii. Find the maximum value in the product and its corresponding column position  $pos$ .
  - iii. Augment the set of highly correlated columns  $\text{Aug}_t$  with the column  $T\_Mat(:, pos)$ .
  - iv. Set the column  $T\_Mat(:, pos)$  to zeros
  - v. Solve for  $aug\_x$  using the formula:  $aug\_x = (\text{Aug}_t' * \text{Aug}_t)^{-1} * \text{Aug}_t' * y$ .
  - vi. Compute the new residual  $u r_n = y - \text{Aug}_t * aug\_x$ .
  - vii. Store the position  $pos$  in the array  $pos\_array$ .
3. Set the values of  $\hat{x}$  at positions  $pos\_array$  equal to  $aug\_x$  and the rest of the entries equal to 0.
4. Return  $\hat{x}$  as the recovered signal.

## 4 Results

In this section, we describe the experiments conducted to evaluate the optimization algorithm. We have done experiments with three images and compared the peak signal-to-noise ratio (PSNR) values of those images having different input samples. The experiment focuses on the Compressed Sensing (CS) recovery of images. In the first experiment, the basis matrix  $\Psi$  is a Discrete Fourier basis. The Fourier coefficients were used and measured using a random Gaussian matrix. The reconstruction was performed using the Orthogonal Matching Pursuit (OMP) algorithm. Finally, the reconstruction performance of the optimized measurement matrix and random Gaussian matrix was compared for different values of  $M$ .

The results of the input image recovery with 3500 and 4250 samples are presented in the Figs. 4 and 5 respectively. From Figs. 4 and 5, we can understand that the quality of the reconstructed image relies upon the number of samples/measurements taken for the reconstruction. If we choose more number samples for the reconstruction, not only does the PSNR value or the signal-to-noise ratio increase linearly but also the elapsed time to reconstruct the image increases exponentially. This can be inferred from Fig. 6.

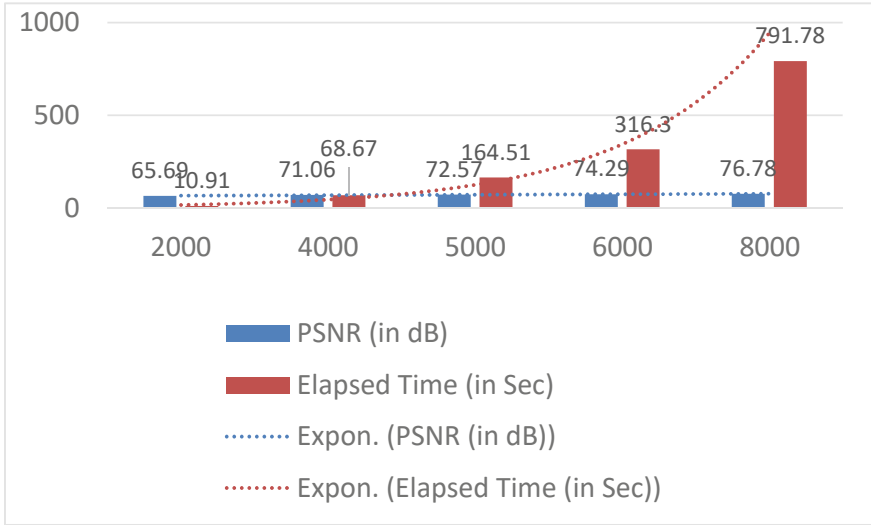


**Fig. 4.** Reconstructed using 3500 samples.



**Fig. 5.** Reconstructed using 4250 samples

In the second experiment, we took two more images with different lighting conditions and different characteristics. The images we took for the experiment are shown in Fig. 7.



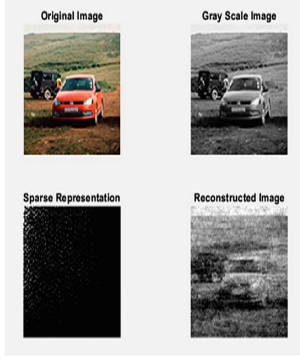
**Fig. 6.** PSNR values and elapsed time of the reconstructed image with 2000, 4000, 5000, 6000 and 8000 samples.

Table 1, shows that in different images having the same number of measurements, the quality and the elapsed time for the reconstructed image vary. The reconstruction also depends upon the characteristics of the image, such as high-texture images, are not getting properly reconstructed. The optimal number of measurements varies from image to image. From the reconstructed images, it is evident that the current approach is capable of reconstructing different texture images with a maximum number of details.

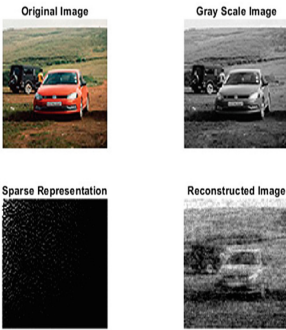




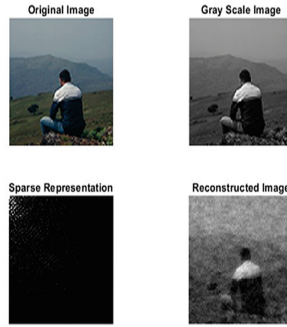
a. Reconstructed Cars Image using 4250 Samples.



b. Reconstructed Cars Image using 4500 Samples.



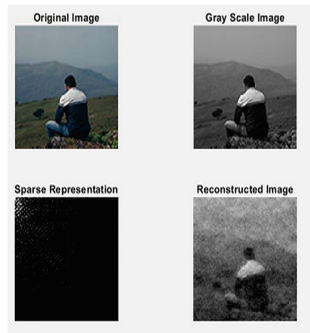
c. Reconstructed Cars Image using 4750 Samples.



d. Reconstructed Human Image using 4250 Samples.



e. Reconstructed Human Image using 4500 Samples



f. Reconstructed Human Image using 4750 Samples

**Fig. 7.** Image reconstruction of two images using three different sampling rates.

**Table 1.** The PSNR values, time taken to reconstruct, and the number of samples taken

Images	PSNR (in dB)	Elapsed time	Regeneration samples
Cars	67.98	100.54	4250
Cars	68.53	104.82	4500
Cars	68.80	127.90	4750
Human	72.80	94.30	4250
Human	73.44	107.90	4500
Human	72.40	126.05	4750

## 5 Conclusion

The process of reconstructing an image using compressive sensing is formulated and tested which involves taking a random number of samples for measurement and then using those measurements to reconstruct the image. We've utilized our images with different characteristics such as lighting conditions, textures, quality, and sharpness. The quality of the reconstructed images is evaluated using PSNR values with the elapsed time, and a comparison of these values is performed. Table 1 shows that we obtained a 73.4% signal-to-noise ratio when reconstructing the human image with an elapsed time of 107 s. From our experiment, this is the highest PSNR value obtained while reconstructing images using OMP. The study can be improved by finding the optimum number of samples for measurement. Also, additional techniques for finding the optimum sampling rate of an image that does not require a high elapsed time to reconstruct the image with a high PSNR value can be explored.

## References

1. Salan, S., Muralidharan, K.B.: Image reconstruction based on compressive sensing using optimized sensing matrix. In: 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kerala, India, 2017, pp. 252–256. <https://doi.org/10.1109/ICICICT1.2017.8342569>
2. Donoho, D.L.: Compressed sensing. *IEEE Trans. Inform. Theory.* **52**(4), 1289–1306 (2006)
3. Jayaraman, S., Esakkirajan, S., Veerakumar, T.: *Digital Image Processing*. Tata McGraw–Hill Education Pvt. Ltd, New Delhi, 2010
4. Donoho, D., Tsai, Y., Drori, I., Starck, J.-l.: Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. Technical report (2006)
5. Cands, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* **52**(2), 489–509 (2006)
6. Yang, Y., Wei, C., Tao, X.: An improved Compressed Sensing algorithm and its application in SAR imaging. In: 2015 IEEE 16th International Conference on Communication Technology (ICCT). IEEE (2015)
7. Aharon, M., Elad, M., Bruckstein, A.: K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006)

8. Tropp, J.A., Gilbert, A.C.: Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **53**(12), 4655–4666 (2007)
9. Li, X., Guoan, B.: Image reconstruction based on the improved compressive sensing algorithm. In: 2015 IEEE International Conference on Digital Signal Processing (DSP). IEEE (2015)
10. Vahid, A., Saideh, F., Bahador, M., Saeid, S.: On optimization of the measurement matrix for compressive sensing. In: 2010 18th European Signal Processing Conference. IEEE (2010)
11. Tropp, J.A., Gilbert, A.C., Strauss, M.J.: Simultaneous sparse approximation via greedy pursuit. In: Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, vol 5. IEEE (2005)
12. S. V. V.V., G. E.A., S. V., S. K.P.: A complex network approach for plant growth analysis using images. In: 2019 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, pp. 0249–0253 (2019). <https://doi.org/10.1109/ICCSP.2019.8698021>
13. Harichandana, M., Sowmya, V., Sajithvariya, V.V., Sivanpillai, R.: Comparison of image enhancement techniques for rapid processing of post flood images. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **XLIV-M-2–2020**, 45–50 (2020). <https://doi.org/10.5194/isprs-archives-XLIV-M-2-2020-45-2020>
14. Suchithra, K.S., Gopalakrishnan, E.A., Jürgen, K., Surovyatkina, E.: Emergency rate-driven control for rotor angle instability in power systems. *Chaos* **32**, 061102 (2022). <https://doi.org/10.1063/5.0093450>
15. Selvin, S., Vinayakumar, R., Gopalakrishnan, E.A., Menon, V.K., Soman, K.P.: Stock price prediction using LSTM, RNN, and CNN-sliding window model. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udipi, India, pp. 1643–1647 (2017). <https://doi.org/10.1109/ICACCI.2017.8126078>
16. Intelligent Computing & Optimization, Conference Proceedings ICO 2018, Springer, Cham. ISBN 978-3-030-00978-6



# Capabilities for Digital Transformation and Sustainability in an Emerging Economy

Suzana Franco Juliano, Elizete da Costa Silva de Paula,  
and Selma Regina Martins Oliveira<sup>✉</sup>

Fluminense Federal University, Volta Redonda, Street. Des. Ellis Hermydio Figueira, 783, Rio de Janeiro, RJ 27213-145, Brasil

{suzanafj, elizetesilva, selmaregina}@id.uff.br

**Abstract.** While the benefits of digital transformation have been widely publicized, little is known about digital transformation and sustainability in an emerging economy. In this article, we intend to identify the prominence of capabilities for digital transformation towards sustainability. Primary survey-based data were collected from multinational companies in Brazil. Descriptive statistics was used for data analysis. The results of our research indicate the relevance of capacities for digital transformation aimed at the sustainability of companies. Dynamic and digital capabilities are more substantive. We offer a contribution at the forefront of one of the most dramatic transformations in the global business scenario for the coming seasons. We intend to shed new light on how to govern for digitally-enabled sustainability.

**Keywords:** Digital transformation capability · Digitally-enabled sustainability · Emerging economies

## 1 Introduction

While the benefits of digital transformation are widely recognized in the cutting-edge literature [1, 2], little is known about digital transformation capabilities and sustainability. This article shows the prominence of capabilities for digital transformation aimed at the sustainability of companies in an emerging economy in South America, in this case Brazil. Understanding this relationship is important for several reasons. Are they:

- Contemporary companies are under more pressure to go beyond profit and operate more sustainably [3–7]. Sustainability is considered an essential component for socially responsible organizations [8]. At the same time, companies have been challenged to engage with digital technologies and reconfigure their business models to remain competitive [1].
- Emerging studies [5, 6, 9, 10] argue that industry 4.0 technologies have the potential to remodel organizations' operations and contribute to achieving sustainability goals in organizations, reducing costs, waste, etc.

- The literature [11] on sustainability is vast, as are substantive studies [1, 12] on digital transformation. Previous studies have indicated that there is a synergy between digital transformation and sustainability [5, 6, 10, 13]. Harel [14] argues that the capacity for digital transformation is essential for sustainable organizational performance due to its ability to integrate the processes that make up the business models among stakeholders, ensuring a sustainable competitive advantage. Thus, intensive investment in capabilities for digital transformation has become a crucial factor for industries [15] with sustainable prospects.
- In this way, we are facing a window of abundant opportunities for multidisciplinary research of factors that can influence the continuous and successful transformation towards sustainability [13]—including the various capacities for digital transformation directed towards sustainability. We argue that organizations need capabilities for digital transformation [12] towards sustainability. Thus, any discussion of sustainability must consider capabilities for digital transformation.

This study aims to capture this theme and advance the knowledge conceived so far, including some important implications. We offer a contribution at the forefront of one of the most dramatic transformations in the global business scenario for the coming seasons. We intend to promote the understanding of capabilities for digital transformation and sustainability intertwined in companies that intend to move forward with digital transformation. In this study, we aim to shed new light on how to govern for digitally enabled sustainability. This study is multivocal and is based on the opinion of employees of multinational corporations in an emerging economy that intends to advance with digital transformation and sustainability.

Thus, the key question guiding this article is: What are the outstanding capabilities for digital transformation towards sustainability? This study brings significant contributions: (a) it provides subsidies for managers and researchers in decision-making processes focused on digitally enabled sustainability; (b) generates insights to advance the debate on capabilities for digital transformation aimed at sustainability; and (c) advances the body of knowledge in relation to existing studies. This article is organized according to the following sections: The next section highlights capabilities for digital transformation and sustainability. Then, the methodology will be presented. In the next section we highlight the results, followed by the discussion, conclusions, implications, limitations and recommendations for future studies.

## 2 Capabilities for Digital Transformation and Sustainability

In the current geopolitical context, organizations face an imperative to achieve environmental sustainability, strengthen operations and have a positive impact on people and the planet [6]. It is at the center of the debate to accelerate the process of digital transformation to achieve sustainability [6]. The World Economic Forum's Global Risk Report 2021 found that issues related to the environment continue to feature in the top five global risks in terms of impacts. These environmental risks have challenged the strategies and operating models of all organizations. At the same time, COVID has rapidly accelerated digital transformation on a global scale across all industries. This represents an opportunity for managers and leaders to prepare for disruption through resilient business operations

supported by trusted data and digital technology [7]. As companies transform and create more efficient and agile business models, environmental sustainability is incorporated into the strategies of these models [7]. In short, companies consider sustainability as a strategy to be inserted in the core fabric of the organization and emerging technologies are fundamental to achieve these initiatives [5].

Significantly, companies are likely to take advantage of emerging technologies (such as AI, hybrid cloud and blockchain, etc.) to achieve sustainability, but are hampered by implementation [5]. Wang et al. [3] argue that digital technologies have a lot of potential to solve substantive problems of society, such as climate change, poverty and resource depletion. We argue that while global digital transformation and environmental action are some of the most powerful drivers for the future of business, these two agendas remain unconnected. Thus, a governance agenda that integrates sustainability and digital transformation is needed. Aligning these agendas has the potential to accelerate the resilience of all companies and improve the state of the world [7]. Digital transformation is a “process of fundamental change, made possible by the innovative use of digital technologies accompanied by the strategic leverage of key resources and capabilities, with the aim of radically improving an entity, redefining its value proposition for its stakeholders” [12]. This definition suggests that the potential of digital transformation is to improve results and should be a strategic imperative in leadership agendas [16].

Although the potential benefits of digital transformation have been widely reported in the literature, there are few conceptual and empirical studies that examine the potential of capabilities for digital transformation aimed at sustainability. Capacity for digital transformation is conceived by three micro-fundamental dimensions [17]: digital knowledge skills (individual dimension), digital intensity and context for action and interaction. We made an analogy of different literatures [12, 18–21] and adopted as capacities for digital transformation directed to sustainability: dynamic capacities; digital capabilities; digital technologies and data; and basic resources. Understanding the prominence of sustainability-driven capabilities for digital transformation is a priority for academics, practitioners and stakeholders [22] because digital transformation can generate economic, social and environmental value.

### 3 Research Methodology

This research was initially developed from a survey of the dimensions of capabilities for digital transformation highlighted in scientific articles, from the Web of Science, Scopus, Science Direct, etc. Using the prestigious literature [12, 18–21] on digital transformation and digital transformation capabilities, this study adopted the following dimensions to measure capabilities for digital transformation in relation to sustainability: “technologies and data”; “human and financial resources and knowledge”; “Dynamic capabilities and digital capabilities”. This study adopted global performance as a measure aimed at achieving sustainability. Based on these dimensions, a questionnaire (Likert scale, with measures: 0—totally disagree and 5—totally agree) was elaborated for data collection, structured in two parts: 1—general information of the respondents (position occupied and time of management experience); 2—prominence of capabilities for digital transformation (technologies, resources and capabilities) aimed at sustainability. Pre-tests

were carried out with two respondents from the information technology and business modeling area, from multinational companies. No significant changes were highlighted by respondents. The instrument’s internal reliability was measured by Cronbach’s alpha coefficient, with a value of 0.76, which can be considered substantial [23].

The identification of the respondents was carried out through a mapping using the professional social network LinkedIn. Professionals who work in the area of accounting oriented to decision-making were selected. Thus, financial managers, directors, controllers, accounting coordinators, accountants, supervisors, accounting business analysts, etc. were mapped. We chose this category of professionals due to the importance of accounting/controlling as an information system that integrates all internal areas of organizations, with the purpose of providing economic, social and environmental information and increasing transparency to internal stakeholders (employees, man-agers, shareholders, etc.) and external (investors, government, companies, etc.) [24]. Furthermore, the prestigious literature suggests that companies that incorporate digital technologies into their accounting systems can reduce fraud risks (machine learning, blockchain, etc.) [25]. The insights provided by the digital transformation will have substantial implications for the accounting ecosystem, leveraging accounting information systems [25] in providing information to meet stakeholder expectations.

Thus, using the Google Forms platform, questionnaires were sent to 205 professionals from Brazilian multinational corporations. We received 15 completed questionnaires (from 15 companies) and none were eliminated. Thus, the sample consists of 15 questionnaires. Based on this fact, it may not be possible to generalize the sample results to a wider population, but even so, we maintain a representative sample to strengthen our conclusions, since it is a specific group of respondents from multinational companies that are in the process of digital transformation and aims to meet the goals of sustainable development. The results of the respondents’ profiles (position held and time of management experience) are shown in Fig. 1a, b.

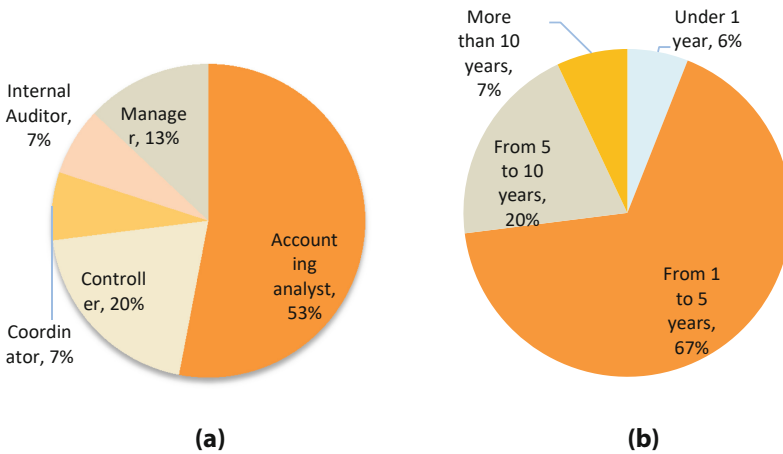


Fig. 1. a Position. b Time experience

The results suggest that the respondents of this research occupy positions of accounting analyst (53%), controller (20%), manager (13%), accounting coordinator (7%) and internal auditor (7%). In addition, the majority (94%) have more than 1 year of experience in the position.

## 4 Results and Analysis

Using descriptive statistics techniques, the results of sustainability-oriented capabilities for digital transformation are presented in the following groups: dynamic capabilities, digital capabilities, basic resources and technologies and data.

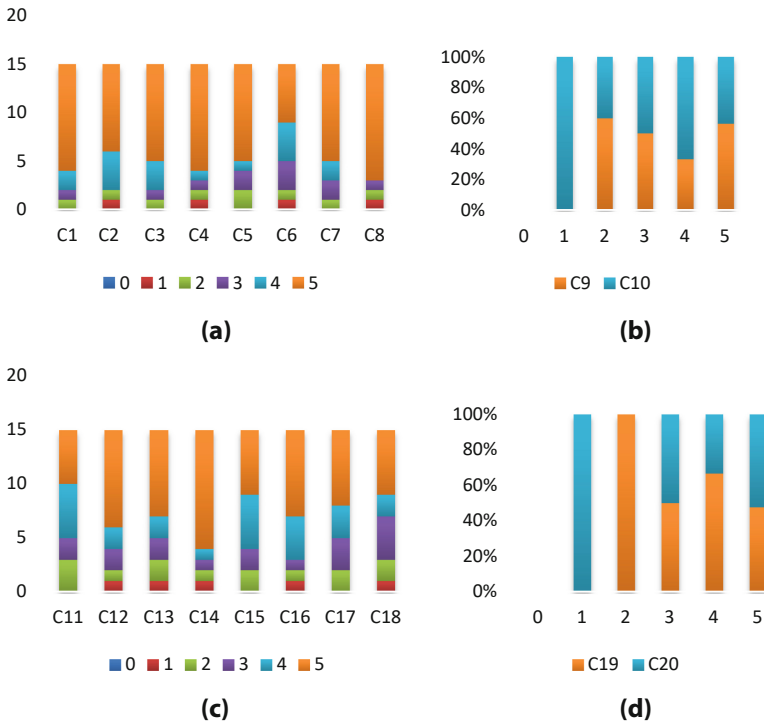


**Fig. 2.** Technologies/data, resources, dynamic capabilities and digital capabilities.

On average ( $M = 4.13$ ), the results presented in Fig. 2 (and Appendix A) suggest that the capabilities for digital transformation aimed at sustainability are quite substantive, with the following highlights: digital capabilities ( $M = 4.4$ ), dynamic capabilities ( $M = 4.31$ ), resources ( $M = 4.0$ ), and technologies ( $M = 3.83$ ). The most prominent dynamic sub-capabilities are adapting to environmental changes (C1— $M = 4.53$ ), strategic planning (C3— $M = 4.47$ ), strategic (C7— $M = 4.4$ ) and learning (C8— $M = 4.4$ ) capabilities. Regarding the most prominent technological sub-capabilities, we mainly highlight access to digital technologies (C9— $M = 4.0$ ). Regarding the most relevant resources, the results suggest human resources, particularly the role of leadership to encourage the search for knowledge and digital technologies in external sources (C14— $M = 4.33$ ) and meet the needs/interests of suppliers and customers in order to generate sustainable value for our stakeholders (C16— $M = 4.13$ ); and finally, financial investments (C12— $M =$



4.13) are considered essential for the digital transformation in businesses that seek sustainability. Finally, acquisition (C19) and implementation (C20) of digital technologies are considered substantially relevant for digital transformation in sustainable business ( $M = 4.4$ ). Our results indicate a greater intensity of responses concentrated at levels 4 and 5 (Fig. 3a–c).



**Fig. 3.** a Dynamics capabilities. b Technologies. c Resources. d Digital capabilities

Regarding the intensity of responses, dynamic capabilities concentrate 80% in levels 4 and 5 (Fig. 3a), distributed as follows: adaptation to environmental changes (C1) (86%); problem solving (C2) and strategic planning (C3) (87%); organization (C4) (80%); innovation (C5) (74%); management of environmental uncertainties (67%); strategic (C7) (80%), learning capacity (C8) (80%). Regarding technologies and data (Fig. 3b), the results indicated that 32% of responses are concentrated in levels 4 and 5, distributed as follows: capabilities to access digital technologies in order to generate sustainable value (C9) (33%) and capabilities to integrate internal and external data sources in order to generate sustainable value (C10) (31%). Regarding resources, the results indicate that 70% of the answers are concentrated in levels 4 and 5 (Fig. 3c), distributed as follows: teams with the necessary technical skills to implement digital transformation/digital technology projects in order to generate sustainable value—C11 (66%); financial resources needed to acquire/invest in digital transformation projects in order to create sustainable

value—C12 (73%); culture guided by digital technologies in order to generate sustainable value—C13 (70%); leadership that encourages the search for external knowledge and digital technologies in order to generate sustainable value—C14 (80%); managers lead digital transformation projects in order to create sustainable value—C15 (73%); digital transformation project managers understand and appreciate the business needs of suppliers and customers in order to generate sustainable value—C16 (80%); digital transformation project managers create opportunities to generate sustainable value—C17 (66%); and digital transformation managers are able to anticipate future business needs and create sustainable value—C18 (53%). Finally, regarding digital capabilities directed towards sustainability (Fig. 3d), the results indicate that 80% of the intensity of responses are concentrated in levels 4 and 5, with 80% for acquisition capabilities (C19) and 80% for implementation (C20) of new technologies to generate sustainable value. Putting all results together, our findings indicate that companies have substantial potential to see environmental opportunities, there is potential for financial and human resources, but to implement, but companies in this research have difficulties to deal with technologies and data integration for analyzes aimed at generating sustainable value.

## 5 Discussion and Conclusion

The unique contribution of this article lies in examining, through empirical evidence, the prominence of sustainability-driven digital transformation capabilities in an emerging economy. Thus, we shed light on the scenario of Brazilian multinational companies that intend to advance with digital sustainability by taking advantage of digital technologies. Going into detail, the conclusions of this study provide results and contribute to the following news in the field of digital transformation knowledge. The results of this study imply that capabilities for digital transformation are substantive for sustainability, with prominence of digital capabilities, dynamics and resources ( $M = 4.4$ ;  $M = 4.31$ ;  $M = 4.0$ ; respectively), but still bump into the technologies and integration of data for analysis with a focus on sustainability. Therefore, we suggest that these companies take steps to remedy these difficulties related to technologies and data integration, such as establishing strategic partnerships in digital ecosystems. Furthermore, we suggest that companies prioritize their efforts in capabilities and resources for digital transformation in a sustainable perspective. This finding is in line with other studies, which highlight the positive effect of digital transformation capacity on firm results [18] or in relation to sustainable performance [26]. These research results can be useful for academics and managers who intend to implement digital transformation projects in sustainability-oriented organizations in the context of emerging economies that share similar characteristics with Brazil. These findings mentioned above in this study have implications for both theory and practice.

### 5.1 Implications for Theory and Practice

The results of this study contribute to theoretical lenses in the field of sustainability-oriented digital transformation. Thus, the findings of this study suggest that companies adopt the dimensions of capabilities for digital transformation in an integrated manner:

technologies, resources, dynamic and digital capabilities, oriented towards sustainability in Brazilian multinational companies. The findings of this study advance the understanding of the theoretical lens of digital transformation capabilities that multinational companies can possess to improve their sustainability. In other words, our study indicates that companies efficiently and effectively manage the dimensions for digital transformation guided by sustainability. Finally, we suggest that managers better orchestrate data-driven technologies aimed at sustainability.

## **5.2 Limitations and Directions for Future Research**

Although our study has theoretical and practical implications, it is not free of limitations. The first limitation is related to the variables considered to measure digital transformation capabilities. Our choice for the variables was based on the proposal by [12] and [18]. So, the results of this study are inherent for this category of variables. Thus, we cannot generalize the results to other locations or realities. We suggest that future studies adopt other variables highlighted by other prestigious literature. Furthermore, this study is cross-sectional and applied in Brazil. Nor can we generalize the results to other countries. Longitudinal research would be interesting to understand the behavior of sustainability-oriented capabilities for digital transformation in organizations. Furthermore, future studies could be applied in other countries for the purpose of comparing results. Finally, another limitation concerns the sample size, considered small. Therefore, we call for future research to increase the sample size to investigate the prominence of capabilities for sustainability-oriented digital transformation. Despite the limitations, the results of this research generate insights and have relevant conclusions for theory, researchers and organizations about capabilities for digital transformation in sustainability-oriented organizations in emerging economies.

## **Appendix**

See Table 1.

**Table 1.** Robustness test—descriptive statistics

Digital transformation capabilities	Legend	Mean	SD	
Dynamics capabilities	Is our company capable of adapting to environmental changes in order to generate sustainable value in the digital age?	C1	4.53	0.68
	Is our company able to act quickly to solve problems in order to generate sustainable value in the digital age?	C2	4.27	0.88
	Do we have the capacity for strategic planning for digital transformation projects in order to create sustainable value?	C3	4.47	0.71
	Do we have the organizational capacity to implement digital projects in order to create sustainable value?	C4	4.33	0.97
	Do we have innovation capabilities in order to create sustainable value ?	C5	4.26	0.97
	Do we have the capacity to manage environmental uncertainties in order to generate sustainable value?	C6	3.86	0.97
	Do we have strategic capabilities aimed at digital transformation in order to create sustainable value?	C7	4.4	0.8
	Do we have learning capacity (access to external knowledge. incentive to innovation. etc.) oriented towards digital transformation in order to create sustainable value?	C8	4.4	0.96
Technologies	Do we have the ability to access digital technologies in a way that generates sustainable value?	C9	4	1.12
	Do we have the ability to integrate internal and external data sources in order to generate sustainable value?	C10	3.66	1.33
Resources	Do we have teams with the technical skills needed to implement digital transformation/digital technology projects in a way that generates sustainable value?	C11	3.8	0.93

*(continued)*

**Table 1.** (continued)

Digital transformation capabilities	Legend	Mean	SD	
	Do we have the necessary financial resources to acquire/invest in digital transformation projects in order to create sustainable value?	C12	4.13	1.04
	Do we have a culture guided by digital technologies in order to generate sustainable value?	C13	3.93	1.15
	Do we have a leadership that encourages the search for external knowledge and digital technologies in order to generate sustainable value?	C14	4.33	0.97
	Do our managers lead digital transformation projects in a way that creates sustainable value?	C15	4	0.8
	Do our digital transformation project managers understand and appreciate the business needs of suppliers and customers in order to generate sustainable value?	C16	4.13	0.92
	Do our digital transformation project managers create opportunities to generate sustainable value?	C17	4	0.93
	Are our digital transformation managers able to anticipate future business needs and create sustainable value?	C18	3.66	1.15
Digital capabilities	Are we able to acquire digital technologies in order to generate sustainable value?	C19	4.4	0.8
	Do we have the capacity to implement digital technologies in order to generate sustainable value?	C20	4.4	0.88

Alfa de Cronbach ( $\alpha$ ): 0.758422

## References

1. Lanzolla, G., Lorenz, A., Miron-Spektor, E., Schilling, M., Solinas, G., Tucci, C.L.: Digital transformation: What is new if anything? Emerging patterns and management research. *Acad. Manage. Discov.* **6**(3), 341–350 (2020)
2. Bonnet, D., Westerman, G.: The new elements of digital transformation. *MIT Sloan Manag. Rev.* **62**(2), 82–89 (2021)
3. Bohnsack, R., Bidmon, C.M., Pinkse, J.: Sustainability in the digital age: intended and unintended consequences of digital technologies for sustainable development. *Bus. Strat. Environ.* **31**(2), 599–683 (2022)
4. Elkington, J.: 25 Years Ago I Coined the Phrase ‘Triple Bottom Line’. Here’s Why It’s Time to Rethink It. *Harvard Business Review* (2018)
5. IBM.: Digital Transformation Can Turn Sustainability into Your Winning Business Strategy. IBM Institute for Business Value (2022). <https://www.ibm.com/blogs/blockchain/2022/02/digital-transformation-sustainability-ibv>
6. McKinsey.: A Digital Path to Sustainability. McKinsey Talks Operations (2022). <https://www.mckinsey.com/business-functions/operations/our-insights/a-digital-path-to-sustainability>
7. World Economic Forum.: Bridging Digital and Environmental Goals: A Framework for Business Action. Geneva
8. Le, T.T.: How do corporate social responsibility and green innovation transform corporate green strategy into sustainable firm performance? *J. Clean. Prod.* **362**, 132228 (2022)
9. Pan, S.L., Zhang, S.: From fighting COVID-19 pandemic to tackling sustainable development goals: an opportunity for responsible information systems research. *Int. J. Inform. Manage.* **55** (2021)
10. George, G., Schillebeeckx, S.J.D.: Digital sustainability and its implications for finance and climate change. *Macroecon. Rev.* **10**(1), 103–108 (2021)
11. Howard-Grenville, J., Davis, G., Dyllick, T., Miller, C., Thau, S., Tsui, A.: Sustainable development for a better world: contributions of leadership, management, and organizations. *Acad. Manage. Discov.* **5**, 355–366 (2019)
12. Gong, C., Ribiere, V.: Developing a unified definition of digital transformation. *Technovation* **102**(C), 102217 (2021)
13. Höllerer, M.A., Shinkle, G., George, G., Mair, J., Pan, S.L., Tim, Y., Collings, D.G.: Digital Sustainability: Addressing Managerial, Organizational, and Institutional Challenges, AMP Special Issue Papers: Digital Sustainability (2023)
14. Bianchi, I.: Technologies and Information Systems for Digital Transformation in Organizations and Society (2022). [http://scielo.pt/scielo.php?script=sci\\_arttext&pid=S1646-98952022000200001&lng](http://scielo.pt/scielo.php?script=sci_arttext&pid=S1646-98952022000200001&lng). Accessed 04 Jan 2023
15. Fischer, M., Imgrund, F., Janiesch, C., Winkelmann, A.: Strategy archetypes for digital transformation: Defining meta objectives using business process management. *Inform. Manage.* **57**(5), 103262 (2020)
16. Vial, G.: Understanding digital transformation: a review and a research agenda. *J. Strateg. Inf. Syst.* **28**(2), 118–144 (2019)
17. Felin, T., Foss, N.J., Heimeriks, K.H., Madsen, T.L.: Microfoundations of routines and capabilities: individuals, processes, and structure. *J. Manage. Stud.* **49**(8), 1351–1374 (2012)
18. Sousa-Zomer, T., Neely, A., Martinez, V.: Digital transforming capability and performance: a microfoundational perspective. *Int. J. Oper. Prod. Manag.* **40**, 1095–1128 (2020)
19. Fitzgerald, M., Kruschwitz, N., Bonnet, D., Welch, M.: Embracing digital technology: a new strategic imperative. *MIT Sloan Manag.* **55**(1) (2014)
20. Brown, A., Fishenden, J., Thompson, M.: Organizational structures and digital transformation. In: *Digitizing Government. Business in the Digital Economy*. Palgrave Macmillan, London. [https://doi.org/10.1057/9781137443649\\_10](https://doi.org/10.1057/9781137443649_10). Accessed 18 Jan 2023

21. Rowe, S.D.: Digital Transformation Needs to Happen Now, vol 21, issue (10) (2017). <https://www.destinationcrm.com/articles/editorial/magazine-features/digital-transformation-needs-to-happen-now-120789.aspx>. Accessed 18 Jan 2023
22. Zimek, M., Baumgartner, R.: Corporate sustainability activities and sustainability performance of first and second order. In: 18th European Roundtable on Sustainable Consumption and Production Conference (ERSCP) (2017)
23. Landis, J., Koch, G.: The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174 (1977)
24. Sciulli, N., Adharani, D.: The use of integrated reports to enhance stakeholder engagement. *J. Account. Organ. Change* (2022). Article publication date: 20 July 2022, ISSN: 1832-5912
25. Garanina, T., Ranta, M., Dumay, J.: Blockchain in accounting research: current trends and emerging topics. *Account Audit Account J* **35**(7), 1507–1533 (2022)
26. ElMassah, S., Mohieldin, M.: Digital transformation and localizing the sustainable development goals (SDGs). *Ecol. Econ.* **169**, 106490 (2020)



# Non-invasive Glucose Measurement with 940 nm Sensor Using Short Wave NIR Technique

M. Naresh<sup>1</sup>, Samineni Peddakrishna<sup>1</sup>(✉), and M. Thirupathi<sup>2</sup>

<sup>1</sup> School of Electronics Engineering, VIT-AP University, Amaravati 522237, India  
Krishna.saminen@gmail.com

<sup>2</sup> Department of Electronics and Communication Engineering, St. Martin's Engineering College, Secunderabad 500014, India

**Abstract.** Diabetes is one of the most common diseases in the world. A finger prick is typically required to obtain blood samples for diabetes testing. It is uncomfortable and prone to infection to undergo these treatments. To address this issue, a non-invasive shortwave near-infrared-based optical detection system with a 940 nm wavelength sensor operating in a transmission mode is proposed. For glucose estimation, the measured signal is converted to voltages and passed through a precision analog-to-digital converter. A relationship between voltage and predicted glucose is evaluated based on measurements of absorbance and transmittance. Using the proposed method for the sensor, the coefficient of determination ( $R^2$ ) and mean the absolute difference to 0.99 and 3.02 mg/dl, respectively, in real-time data analysis. Additionally, it is determined that the root mean square error (RMSE) is 4.53 mg/dl, while the mean absolute relative difference (MARD) is 1.60%.

**Keywords:** Blood glucose · Diabetes · Near-Infrared (NIR) · Noninvasive glucose · Short wave · Transmission mode

## 1 Introduction

According to the World Health Organization (WHO), the number of people with diabetes has doubled since 2015. The estimated global prevalence of diabetes in 2019 was 9.3% or 463 million people, and it is projected to increase to 10.2% (578 million) by 2030 and 10.9% (700 million) by 2045 [1]. Type 2 diabetes is the most common form of diabetes, and it is crucial for patients to self-monitor their blood glucose levels daily as part of their diabetes management program. This can help them make necessary lifestyle changes to improve their health.

In the past few years, several methods of monitoring glucose levels have been developed. Currently, the most commonly used method for checking blood glucose levels involves pricking the finger with a traditional glucose meter, which is an invasive approach [2]. However, this method is uncomfortable for patients, and even the smallest holes caused by pricking the skin can lead to infections,



making it difficult to incorporate into daily life. Additionally, invasive glucometers are not cost-effective because the single-use strips need to be replaced after use [3]. Alternatively, minimally invasive methods that cause less skin damage may be used, but these devices are expensive and require frequent calibration, making them unsuitable for regular monitoring of blood glucose levels [4]. For these reasons, researchers have been working on developing noninvasive, accurate, and cost-effective methods of measuring blood glucose [5]. By doing so, regular blood glucose monitoring can become a more comfortable and relaxed experience for people with diabetes.

One of the methods used to develop non-invasive glucose sensors is electromagnetic wave sensing, which interacts with compounds in the body through absorption, scattering, and transmission. However, the use of high-frequency EM wave sensing may result in ionizing radiation that can damage tissues, while low-frequency EM wave sensing is based on the impedance of tissue and is directly related to the properties of the material [6]. This measures the impedance of the skin using an alternating current of known intensity and is commonly referred to as impedance spectroscopy. This technology utilizes a resistor, inductor, and capacitor resonant circuit to measure impedance for glucose monitoring [7]. However, it is affected by changes in water content and temperature, as well as sweat.

As an alternative to high and low-frequency spectroscopy, microwave/millimeterwave spectroscopy and optical wave spectroscopy have been studied. In the microwave and mm-wave frequency ranges, EM wave radiation carries less photon energy and non-ionizing radiation, thereby does not cause tissue damage [8]. The dielectric properties of the skin are closely related to the dielectric properties of these sensors, and they vary with changes in glucose levels [9]. However, it is a challenging task to determine the relationship between glucose level and permittivity, which is one of the main factors restricting their development. Researchers are currently working on improving microwave sensors' selectivity and sensitivity [10].

In recent years, more attention has been given to the two optical techniques known as MIR and NIR. NIR signals have wavelengths between 750 and 2500 nm, while MIR signals have wavelengths between 2500 and 10,000 nm. Because MIR only penetrates a few micrometers into human tissue, it can only be used in the reflection mode, while NIR spectrometry is a suitable method for estimating blood glucose levels. NIR light can penetrate through multiple layers of the skin and reach subcutaneous vessels, regardless of skin pigmentation, making it superior to most other infrared wavelengths in terms of penetration depth [10,11].

NIR spectrometry can be categorized into long-wave NIR and short-wave NIR based on their bandwidths. While long-wave NIR is better for detecting glucose molecules during light absorption by biological tissue, it does not provide accurate results for in-vivo tests because of its shallow penetration [12]. Short NIR waves have less glucose molecule absorption, but they are well-suited for in-vivo testing because of their deep penetration [13]. In a study, NIR regions were used to estimate glucose levels, with three sensors operating at 940 and 1300 nm

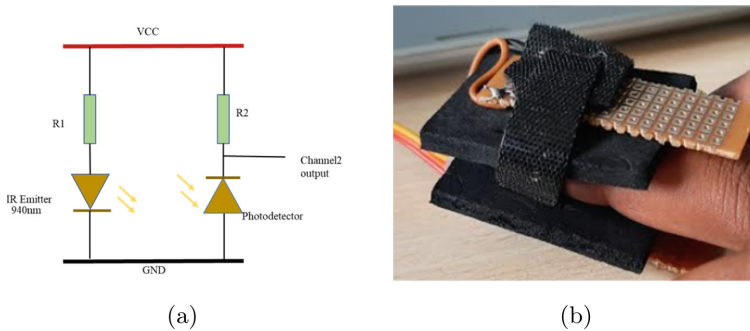
[14]. Two of the 940 nm sensors were used in absorbance and reflectance modes, and one 1300 nm sensor was used in absorbance modes. The results of this study showed that the short NIR regions are more suitable for the estimation of blood glucose levels and that the glucose absorption peaks for glucose isomers such as fructose, lactose, and galactose do not coincide with glucose absorption, indicating that these isomers do not affect glucose detection [15].

Thus, this work proposes the use of short-wave NIR transmittance spectroscopy at 940 nm to improve accuracy. The following sections discuss prior research and the novelty of the proposed approach. The proposed system employs a 940 nm transmission sensor and achieves a 99% accuracy by using a third-order polynomial. The accuracy of the proposed method has been determined by considering 54 subject samples. The relationship between glucose concentration and signal absorbance and transmittance is also discussed in the current work.

## 2 Proposed Methodology and Implementation

A new approach for non-invasive blood glucose measurement using a single wavelength. The technique utilizes a 940 nm short-wave NIR sensor in transmission mode to measure glucose levels in the blood. The system is implemented using an ADC, an infrared sensor, and a microcontroller, as shown in Fig. 1. To enable non-invasive measurement of blood glucose levels, a near-infrared TSAL6400 high-power infrared emitting Diode with a wavelength of 940 nm, GaAlAs, is used as the transmittance LED, along with an IR receiver LED to detect Infrared Rays. The anode and cathode of the IR receiver LED are connected in reverse to ground and supply voltage, respectively. When the IR receiver LED is struck by light, it allows current to flow, and the flow of current is proportional to the amount of light striking the IR receiver LED. The diameter of the IR receiver LED is 5 mm, with a power dissipation of 100 mW.

To ensure that the TSAL6400 is not damaged in the circuit, a current-limiting resistor  $R_1$  must be included in the series. The value of this resistance can be



**Fig. 1.** **a** Schematic diagram of 940 nm transmittance sensor and **b** image of prototype

calculated using Eq. 1 as shown below,

$$R_1 = \frac{V_{cc} - V_F}{I_F} \quad (1)$$

The resistance value  $R_1$  in ohms can be calculated using Eq. 1 with the DC power supply voltage  $V_{cc}$  in volts, forward voltage drops across the LED  $V_F$  in volts, and forward current  $I_F$  in amperes. According to the 940 nm datasheet,  $V_{cc}$  is 5 V,  $V_F$  is 1.35 V, and  $I_F$  is 100 mA. Thus, the calculated resistance is approximately 365 ohms, which can be rounded up to 380  $\Omega$ .

To prevent damage to the phototransistor, a resistor  $R_2$  must be included in the circuit design and connected in series with the cathode of the photodetector. This resistor limits the current in the circuit, and its value can be calculated using the following equation:

$$R_2 = \frac{V_{cc} - V_{CE(SAT)}}{I_C} \quad (2)$$

where  $R_2$  is the series collector resistance in ohms,  $V_{CE(SAT)}$  is the voltage dropped across the cathode and anode of the IR receiver in volts, and  $I_C$  is the collector current in amperes. Using the 940 nm sensor datasheet, with  $V_{CC} = 5 V$  and  $I_C = 100 \mu A$ , the resistance ( $R_2$ ) can be determined as  $R_2 = 47 k\Omega$ .

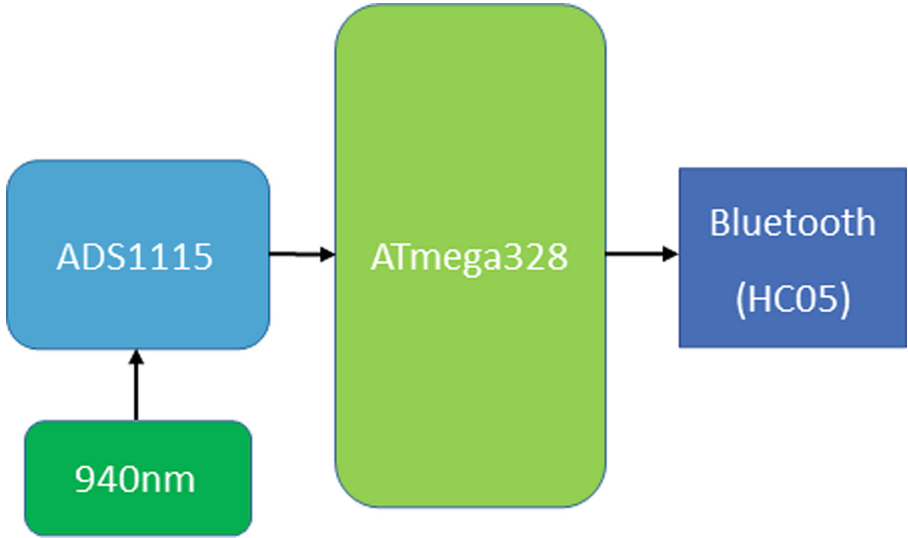
For processing the sensor data, the ADS1115 is selected as the high-precision ADC for this work due to its 860 samples/second rate. Two single-ended channels are used out of the four available, and the ADS1115 has a configurable gain amplifier with a maximum gain of x16 to amplify channel signals to their full range. It can operate from 3 to 5 V power and logic, measure a wide range of signals, and is user-friendly. The one channel is used in this project are A1, which are connected to the sensor outputs. The ADS1115 is interfaced with the ATmega328 using the I2C bus protocol. The microcontroller communicates with the ADS1115, receives data in the form of millivolts, and converts the voltages to compute the glucose value. The calculated glucose values are sent through Bluetooth and recorded in the system. The block diagram of the proposed approach is shown in Fig. 2. The millivolts are converted into volts to detect blood glucose concentration in the human finger in the transmission mode of sensor data. For this, the transmission mode sensor data is transferred to the microcontroller via ADC. For measuring T-transmittance and A-absorbance of measured glucose concentration of the respective subject are determined with help of Eqs. (3) and (4) respectively [13], by using algorithm in Table 1. The equation for T is given by:

$$T = \frac{V_t}{V_r} \quad (3)$$

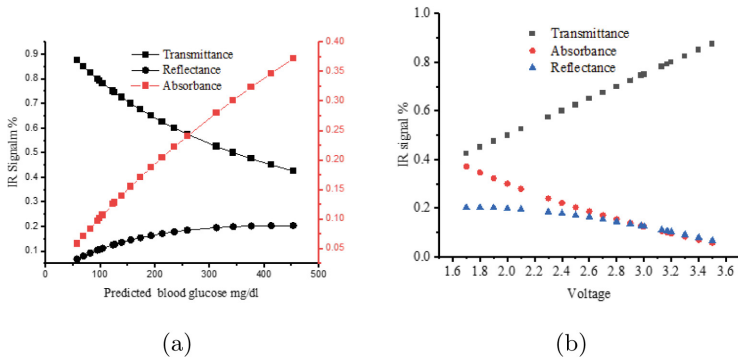
where  $V_i$  is the input voltage from ADC and  $V_r$  is the reference voltage. The term  $V_t$  is defined as  $V_t = V_i - V_r$ .

The equation for OD is given by:

$$OD = -\log_{10} T \quad (4)$$



**Fig. 2.** Block diagram of the proposed system



**Fig. 3.** Glucose range and its correlation **a** relation between IR signal and glucose, **b** relation between IR signal and voltage.

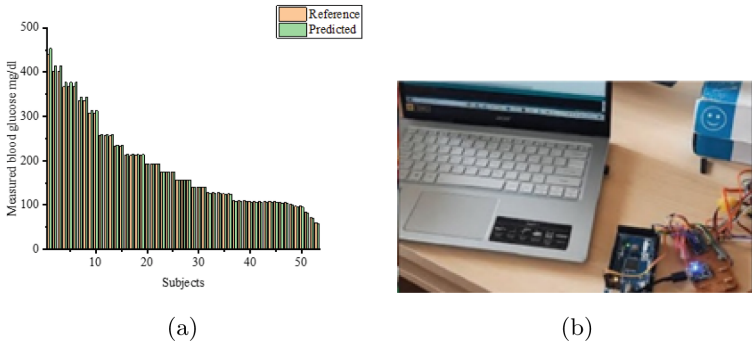
Figure 3 shows the predicted glucose range and its correlation with the absorbance, transmittance, and reflections of the sensor, which are obtained from the 54 subject's fingers using the transmission mode sensor. It provides information on the minimum and maximum values for various parameters such as Predicted glucose in mg/dl, Absorbance (A), Transmittance (T), Reflections (R), and Voltage (V).

As seen in Fig 3(a), an increase in predicted glucose correlates with an increase in signal absorbance and a decrease in transmittance. Additionally, the reflections of the signal display lossy behavior towards the transmitted signal. For the Predicted glucose in mg/dl, the minimum and maximum values are 57

and 453, respectively, with the near intersection being 259. The "Near intersection" column refers to the approximate point where the values of the parameters intersect. The Absorbance (A) values range from 0.05 to 0.37, with a near intersection of 0.57. The Transmittance (T) values range from 0.87 to 0.42, with a near intersection of 0.57. The Reflections (R) values range from 0.06 to 0.20, with a near intersection of 0.17. The sum of the Absorbance, Transmittance, and Reflections values for each row is always equal to 1.

Figure 3(b) displays the relationship between the output voltage obtained from the sensor output and the glucose concentration. It is observed that as the voltage obtained by the sensor increases, the glucose molecules absorb less signal. Conversely, if the glucose concentration absorbs more signals, the voltage decreases. The relationship between transmittance and voltage is proportional to the receiver part, while the absorbance is inversely proportional to the received voltage. For the Voltage (V), the minimum and maximum values are 1.7 and 3.5, respectively. The Absorbance (A) values range from 0.37 to 0.05, with a near intersection of 0.40. The Transmittance (T) values range from 0.42 to 0.87. The Reflections (R) values range from 0.20 to 0.06. The sum of the Absorbance, Transmittance, and Reflections values for each row is always equal to 1.

Furthermore, in order to assess the precision of the technique, a comparison was made between the sample values obtained from the proposed method and those from the reference glucometer provided by Dr. Trust, as depicted in Fig. 4(a). As illustrated in the figure, the measurements exhibit a high level of agreement with minimal deviation. Figure 4(b) presents a prototype setup for the measurement.



**Fig. 4.** Measurement comparison and setup **a** glucose measurement comparison with reference device and proposed approach, **b** prototype setup for the measurement.

### 3 Mathematical Modelling for Prediction of Blood Glucose Concentration

The current technique aims to establish a correlation between the voltages obtained from the sensor and the predicted blood glucose concentration in reference to the conventional intrusive glucometer. The output voltages were measured while subjects placed their finger in the sensor. The detector results in volts represent an independent variable corresponding to the expected glucose response from the 940 nm sensor. The proposed method was developed based on data collected from 54 subjects, aged 19–83, using the random blood glucose test mode. The overall precision of the proposed method was evaluated using various metrics such as mean absolute relative difference (mARD), mean absolute deviation (MAD), root mean square error (RMSE), and average error. The proposed method significantly reduces the overall error, and the average error from the expected glucose concentration is obtained to evaluate the system's performance. To assess the accuracy of the calculated blood glucose levels, metrics such as mARD, MAD, and RMSE are calculated using Eqs. (5), (6), and (7).

$$MARD = \frac{1}{n} \sum_{i=1}^n \left| \frac{BG_{Ref} - BG_{Pre}}{BG_{Ref}} \right| \times 100 \quad (5)$$

$$MAD = \frac{1}{n} \sum_{i=1}^n |BG_{Ref} - BG_{Pre}| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |BG_{Ref} - BG_{Pre}|^2} \quad (7)$$

In the above equations,  $BG_{Ref}$  and  $BG_{Pre}$  refer to the reference and predicted background values, respectively.  $n$  is the number of samples. MARD denotes mean absolute relative difference, MAD denotes mean absolute difference, and RMSE denotes root mean square error.

Further, Clarke error data analysis is used to quantify the diagnostic accuracy of the estimated concentration of glucose. Clarke (2005) investigates this, which has become the standard for validating the results. There is a small disparity between the projected and referenced blood glucose concentrations, therefore readings in zones A and B are preferable. During Clarke error grid analysis, it is determined that the proposed non-invasive technology locates all subjects in zone A. After statistical parameter analysis, the suggested noninvasive method is evaluated on 54 fresh individuals. These data come from individuals aged 18–83. As depicted in Fig. 5, all anticipated blood glucose concentration values fall within zone A of the Clarke error grid analysis in the following system. Where the mARD of 1.60 is superior to the 3.25 mARD of previous research. This represents a significant improvement. The improved MAD, average error and RMSE values are 3.02 mg/dl, 3.12%, and 4.53 mg/dl, respectively. The proposed approach for predicting glucose readings yielded a mARD of 1.60 based on

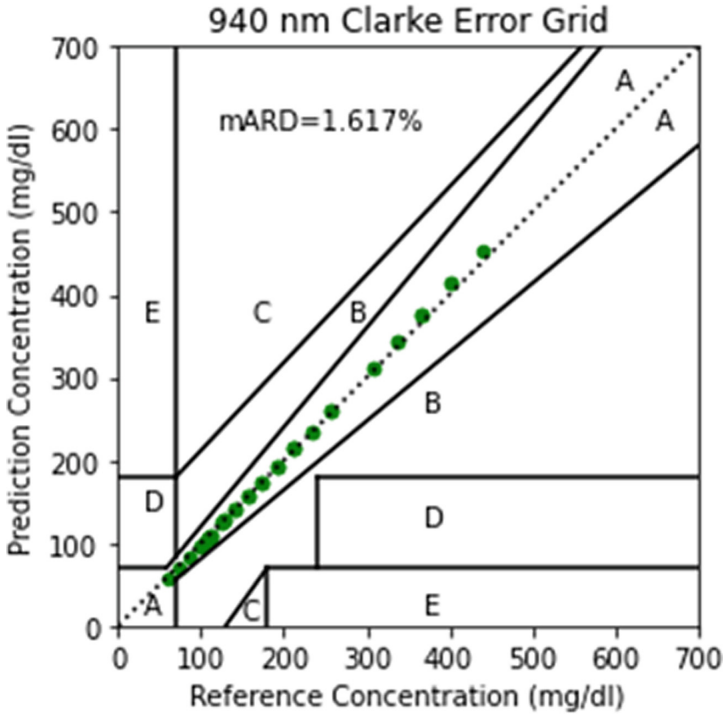


Fig. 5. Clarke error grid analysis between the reference and proposed sensor

the observation of these 54 participants. The proposed post-processing method resulted in a high correlation coefficient ( $R^2$ ) of 0.99. Maximum blood glucose detection is correlated with the least received light intensity. The precise range for measuring blood glucose in this experimental investigation is 80–420 mg/dl from 57–453 mg/dl for 940 nm sensors in the 700 mg/dl scale. Optical detection for noninvasive measurement of blood glucose is an effective method for general use. As the received light intensity is directly proportional to the concentration gradient of the glucose molecule.

#### 4 Conclusion

A NIR spectroscopy technique-based non-invasive glucose measurement system is proposed using a transmission sensor for the measurement system comparatively is accurate using NIR shortwave along with visible light blocking filter. During statistical analysis, the performance parameter such as mARD, average error, and RMSE are improved in comparison with previous proposed works. From proposed reflective sensor is better than the transmission mode sensor. The proposed computation became more efficient and optimized with a correlation coefficient of 0.99. Real-time validation of the device has been done through Dr. Trust’s invasive glucometer.

## References

1. Sun, H., Saeedi, P., Karuranga, S., Karuranga, S., Pinkepank, M., Ogurtsova, K., Duncan, B.B., Stein, S., Basit, A., Chan, J.C.N., et al.: IDF diabetes atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Res. Clin. Pract.* **183**, 109119–109132 (2022)
2. Gusev, M., Poposka, L., Guseva, E., Kostoska, M., Koteska, B., Simjanoska, M., Ackovska, N., Stojmenski, A.: Trends from minimally invasive to non-invasive glucose measurements. In: *2020 43rd International Convention on Information, Communication and Electronic Technology*, pp. 315–320 (2020)
3. Yeaw, J., Lee, W.C., Aagren, M., Christensen, T.: Cost of self-monitoring of blood glucose in the United States among patients on an insulin regimen for diabetes. *J. Manag. Care Pharm.* **18**, 21–32 (2012)
4. Chen, C., Zhao, X.-L., Li, Z.-H., Zhu, Z.-G., Qian, S.-H., Flewitt, A.J.: Current and emerging technology for continuous glucose monitoring. *Sensors* **17**, 182 (2017)
5. Van Enter, B.J., von Hauff, E.: Challenges and perspectives in continuous glucose monitoring. *Chem. Commun.* **54**, 5032–5045 (2018)
6. Zhang, R., Liu, S., Jin, H., Luo, Y., Zheng, Z., Gao, F., Zheng, Y.: Noninvasive electromagnetic wave sensing of glucose. *Sensors* **19**, 1151 (2019)
7. Hayashi, Y., Livshits, L., Caduff, A., Feldman, Y.: Dielectric spectroscopy study of specific glucose influence on human erythrocyte membranes. *J. Phys. D Appl. Phys.* **36**, 369–374 (2003)
8. Nakamura, M., Tajima, T., Ajito, K., Koizumi, H.: Selectivity-enhanced glucose measurement in multicomponent aqueous solution by broadband dielectric spectroscopy. In: *Proceedings of the 2016 IEEE MTT-S International Microwave Symposium (IMS)*, pp. 1–3, San Francisco, CA, USA (22–27 May 2016)
9. Hofmann, M., Fischer, G., Weigel, R., Kissinger, D.: Microwave-based noninvasive concentration measurements for biomedical applications. *IEEE Trans. Microw. Theory Tech.* **61**, 2195–2204 (2013)
10. Malik, B.H., Coté, G.L.: Real-time, closed-loop dual-wavelength optical polarimetry for glucose monitoring. *J. Biomed. Opt.* **15**, 017002 (2010)
11. Khalil, O.S.: Non-invasive glucose measurement technologies: an update from 1999 to the dawn of the new millennium. *Diabetes*
12. Jain, P., Maddila, R., Joshi, A.M.: A precise non-invasive blood glucose measurement system using NIR spectroscopy and Huber's regression model. *Opt. Quantum Electron.* **51**, 51 (2019)
13. Uwadaira, Y., Ikehata, A., Momose, A., Miura, M.: Identification of informative bands in the short-wavelength NIR region for non-invasive blood glucose measurement. *Biomed. Opt. Express.* **7**, 2729–2737 (2016)
14. Joshi, A.M., Jain, P., Mohanty, S.P., Agrawal, N.: iGLU 2.0: A new wearable for accurate non-invasive continuous serum glucose measurement in IoMT framework. *IEEE Trans. Consum. Electron.* **66**, 327–335 (2020)
15. Simeone, M.L.F., Parrella, R.A., Schaffert, R.E., Damasceno, C.M., Leal, M.C., Pasquini, C.: Near infrared spectroscopy determination of sucrose, glucose and fructose in sweet sorghum juice. *Microchem. J.* **134**, 125–130 (2017)





# Managing the Purchase-Sale Process of Digital Currencies Under Fuzzy Conditions

V. Malyukov<sup>1</sup>, B. Bebeshko<sup>2</sup>(✉), V. Lakhno<sup>1</sup>, O. Smirnov<sup>3</sup>, I. Malyukova<sup>4</sup>,  
and H. Mohylnyi<sup>5</sup>

<sup>1</sup> National University of Life and Environmental Sciences of Ukraine, Kyiv, Ukraine  
lva964@nubip.edu.ua

<sup>2</sup> State University of Trade and Economics, Kyiv, Ukraine  
b.bebeshko@knute.edu.ua

<sup>3</sup> Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine

<sup>4</sup> Rating Agency «Expert-Rating», Kyiv, Ukraine

<sup>5</sup> Luhansk Taras Shevchenko National University, Starobilsk, Ukraine

**Abstract.** The article discusses a model for managing the process of trading digital cryptocurrencies (DCC) in a fuzzy environment. It is demonstrated that the controllability of such a process can be described from a game theory perspective. This approach enables the development of constructive strategies for maintaining stability in relationships between players in the DCC market, particularly in cases where one of the players provides fuzzy information about the DCC volume of the counterparty player. Consequently, solving such problems will lead to a more accurate predictive evaluation of transaction effectiveness. The proposed model is novel in that it employs a constructive approach to solve bilinear multi-stage quality games with several terminal surfaces, specifically for the case of fuzzy information. A computational experiment is conducted, taking into account various ratios of parameters that describe the process of DCC trading operations. The results of this study provide interested parties, particularly players in the DCC market, with the means to maintain stability in both traditional currency and DCC markets. The model may also be useful for software products designed for trading platforms where DCC transactions are conducted.

**Keywords:** Digital cryptocurrencies · Game model · Buying and selling digital cryptocurrencies

## 1 Introduction

DCC is a special electronic means of payment, the exchange rate of which depends only on supply and demand [1]. Such electronic money is not regulated by any governmental systems. The function of observers and controllers in this case is bound to the network users and owners of the DCC [2], respectively. In the context of economic relations, DCC can act as both an investment and an object of investment activity. Although an extensive scientific literature is devoted to investment activity in DCC markets, as well as the analysis and evaluation of the effectiveness of investments in DCCs, the issue of

managing the procedure for buying and selling DCCs in a fuzzy setting based on the joint application of game theory and fuzzy mathematics has not yet been addressed. As shown below, game-theoretic modelling in the problems of managing the purchase and sale of DCCs, taking into account uncertainty, conflict, and the economic risk generated for investors, enables the assessment of the reliability level of potential transactions in the DCC market. This, in turn, reduces the levels of economic risk for investors in DCCs.

## 2 Literature Review

Considering the retrospective development of the DCC market, one can quickly see that interest in this market is growing rapidly over a relatively short period of time. Initially, interest in DCC was shown exclusively by enthusiasts [1–4], but this market has since attracted a wide range of investors [5–8]. Today, many states and their financial organisations—such as banks, credit and insurance companies—hold DCCs as assets [9–11]. Many of these organisations are also discussing the issue of full recognition of DCC as a legitimate means of payment, and promoting the idea of developing full-fledged payment systems for DCC at the state level. Specialised cryptocurrency exchanges are also actively operating worldwide [12–15], facilitating trading between private investors and companies that recognize DCC as a legitimate means of payment [15]. There is a certain analogy between classical exchange activities and exchanges dealing with DCC transactions. However, the dynamics of price changes on the DCC market cannot be compared to that of classical stock exchanges. This makes predicting DCC rates and analysing the risks of investing in DCC extremely relevant. Therefore, this study is devoted to considering this problem.

An analysis of scientific publications reveals that the attractiveness of investments in the DCC market has not been fully explored. Most research focuses on the issue of DCC volatility, as seen in works [11–15]. The authors of [8–17] consider BTC as an alternative to traditional money, highlighting its potential for widespread use as a global currency. However, most scientists and practitioners acknowledge the significant volatility of cryptocurrencies, making BTC and other DCCs a rather risky investment option. Although the global capitalization of the DCC market has experienced some decline compared to 2021, on December 22nd, three currencies showed growth. The leading currency of the day was DCC Helium HNT, which reached \$1.91. The second most profitable currency was DCC Toncoin (TON), which experienced active growth at 3.57% and traded at \$2.53. The third was DCC Ethereum Classic ETC, which traded at \$16.02 [18]. A stable currency is essential for the successful development of any state [8, 9], especially during the digital transformation of the economy. To maintain currency stability, various models are being developed that take into account the latest factors associated with the widespread use of technology. This study proposes a model that employs direct methods for maintaining currency stability. Note that for many tasks in various areas of human activity, forecasting, including economics or trade, is extremely important. Since a high-quality mathematically justified forecast is also important for the tasks of investing in DCC, the paper will consider issues related to forecasting methods for the financial market. All of the above predetermined the goals of our study and its relevance.

### 3 The Purpose of the Study

The purpose of the study: to develop a model of trading operations with digital cryptocurrencies in a fuzzy formulation based on the use of the apparatus of bilinear multi-stage quality games with several terminal surfaces.

## 4 Methods and Models

### 4.1 Model of Trading Operations with Cryptocurrencies in a Fuzzy Setting

The article considers a model of trading operations with DCCs in a fuzzy formulation. It is assumed that two players are involved in trading operations with DCCs. This is because, despite the large number of DCCs such as ADA, BTC, DOT, EOS, ETC, ETH, LINK, LTC, XRP, and others, trading sessions are conducted using DCC pairs selected for various reasons. The difference between the trade operations under consideration and similar operations with complete information lies in the fact that the first player does not know exactly the state  $y^\xi(0)$  (volume of DCC) of the second player. The first player has only information that the volume of the DCC of the second player belongs to the fuzzy set  $\{Y, m(\cdot)\}$ , where  $Y$ — the subset  $R_+$ ,  $m(\cdot)$  is the membership function of the state  $y^\xi(0)$  of the second player in the set  $Y$ ,  $m(y^\xi(0)) \in [0, 1]$  for  $y^\xi(0) \in Y$ . In addition, at every moment  $t$  ( $t = 0, 1, \dots, T$ ),  $T$ —a natural number, the first player knows his states  $x(\tau)$  (volumes of its DCC) for  $\tau \leq t$ . In this case, the following conditions are met:  $x(\tau) \geq 0$  if the condition is  $x(\tau) < 0$  met with certainty  $< p_0$  ( $0 \leq p_0 \leq 1$ ) and  $x(\tau) < 0$  if the condition is  $x(\tau) < 0$  met with certainty  $\geq p_0$ ; and also the values of realisations of the trading strategy of the first player  $u(\tau)$  ( $\tau \leq t$ ) during the trading session are known.

The reasoning is approached from the perspective of the first player, without making any assumptions about the knowledge of the second player, which is equivalent to assuming that the second player has no information. Both players make their moves simultaneously. Therefore, one player with digital cryptocurrency 1 (DCC1) buys digital currency 2 (DCC2) from the second player who has DCC2 and sells (or buys) DCC1. Before the start of the trading session, the spot rate of DCC1 in relation to DCC2 is set:  $k_{btc1}$ . At the moment  $t = 0$  (beginning of trading) player I has  $x$  (DCC1) to buy DCC2. The second player has  $y^\xi$  DCC2 to buy DCC1. Let us describe the model of trading operations with selected DCCs. At the moment  $t = 0$ , players I and II, replenish the volumes of DCC  $x(0)$  (DCC1) and  $y^\xi(0)$  (DCC2) they have and have the following volumes of DCC— $\alpha \cdot x(0)$  and  $\beta \cdot y^\xi(0)$ . Here, respectively,  $\alpha$  and  $\beta$  are the growth rates of  $u(0) \cdot \alpha \cdot x(0)$  ( $0 \leq u(0) \leq 1$ ) DCC1 and  $v(0) \cdot \beta \cdot y^\xi(0)$  ( $0 \leq v(0) \leq 1$ ) DCC2 volumes. Then the players allocate, respectively, DCC1 and DCC2 for the purchase of DCC2 and DCC1, respectively. It is assumed that at the time of the trading session, the DCC2 buying and selling rates were  $k_{pok}$  and  $k_{prod}$ . Then, the volumes of DCC1 and DCC2 for players I and II at the moment of time  $t = 1$  will be  $x(1)$  and  $y^\xi(1)$  respectively, where  $x(1)$  and  $y^\xi(1)$  are determined from the relations:

$$\begin{aligned} x(1) = & \alpha \cdot x(0) - u(0) \cdot \alpha \cdot x(0) \cdot [1 - (k_{btc1}/k_{prod})] \\ & + v(0) \cdot \beta \cdot y^\xi(0) \cdot [k_{btc1} - k_{pok}]; \end{aligned} \quad (1)$$

$$y^{\xi}(1) = \beta \cdot y^{\xi}(0) - v(0) \cdot \beta \cdot y^{\xi}(0) \cdot [1 - (k_{pok}/k_{btc1})] + u(0) \cdot \alpha \cdot x(0) \cdot [(1/k_{btc1}) - (1/k_{prod})]. \quad (2)$$

These ratios mean the following. Player I, managing DCC1, allocates a part of the DCC1 cryptocurrency  $u(0) \cdot \alpha \cdot x(0)$  for the purchase of DCC2. For the allocated value of the DCC1 cryptocurrency, he buys the amount of  $[u(0) \cdot \alpha \cdot x(0)/k_{prod}]$  DCC2, which is sold to him by player II at the selling rate of the DCC2 cryptocurrency  $k_{prod}$ . This rate has developed in this trading session. That is, player I, instead of the mass of the DCC1 cryptocurrency  $u(0) \cdot \alpha \cdot x(0)$ , which he allocated for the purchase of the DCC2 cryptocurrency, acquired DCC2, the volume of which is estimated in  $(k_{btc1}/k_{prod}) \cdot u(0) \cdot \alpha \cdot x(0)$  DCC1. As a result, the first player (hereinafter  $PL_1$ ), after carrying out the procedure for purchasing DCC2, has electronic financial resources (hereinafter FR) in the DCC1-equivalent equal to  $\alpha \cdot x(0) - u(0) \cdot \alpha \cdot x(0) \cdot [1 - (k_{btc1}/k_{prod})]$ .

In addition to the purchase of DCC2 by the player  $PL_1$ , during the trading session, the sale of DCC2 by the second player takes place (hereinafter  $PL_2$ ). For the purchase of DCC1  $PL_2$  allocates  $v(0) \cdot \beta \cdot y^{\xi}(0)$  DCC2. The first player ( $PL_1$ ) buys from  $PL_2$  at the buying rate  $k_{pok}$ . Therefore, after the procedure of selling DCC2 by the player  $PL_2$  in the amount of  $v(0) \cdot \beta \cdot y^{\xi}(0)$ , the player  $PL_1$  will have FR by the value of  $v(0) \cdot \beta \cdot y^{\xi}(0) \cdot [k_{btc1} - k_{pok}]$ , in the DCC1-equivalent. Then, after the trading session, the player  $PL_1$  FR, in DCC1-equivalent, will have:

$$\alpha \cdot x(0) - u(0) \cdot \alpha \cdot x(0) \cdot [1 - (k_{btc1}/k_{prod})] + v(0) \cdot \beta \cdot y(0) \cdot [k_{btc1} - k_{prod}]$$

The situation is similar with the player's  $PL_2$  FR.

For the amount of  $v(0) \cdot \beta \cdot y(0)$  DCC2 allocated for the purchase of DCC1, the player  $PL_2$  buys DCC1 in the amount of  $v(0) \cdot \beta \cdot y^{\xi}(0) \cdot k_{pok}$ . This leads to the fact that the player's  $PL_2$  FR, in DCC2, will decrease by  $v(0) \cdot \beta \cdot y^{\xi}(0) \cdot [1 - (k_{pok}/k_{btc1})]$ . In addition, taking into account the fact that the player  $PL_1$  «on his own» bought DCC2 from the player  $PL_2$ , the DCC2-equivalent of FR will increase by  $u(0) \cdot \alpha \cdot x(0) \cdot [(1/k_{btc1}) - (1/k_{prod})]$ . Thus, according to the results of the trading session, the player  $PL_2$  will have FR, in DCC2, in the amount that will be:

$$\beta \cdot y^{\xi}(0) - v(0) \cdot \beta \cdot y^{\xi}(0) \cdot [1 - (k_{pok}/k_{btc1})] + u(0) \cdot \alpha \cdot x(0) \cdot [(1/k_{btc1}) - (1/k_{prod})];$$

Conditions for the end of the trading session at the moment  $t = 1$  will be the fulfilment of conditions (3) or (4):

$$x(1) \geq 0, y^{\xi}(1) < 0; \quad \text{with certainty} \geq p_0 \quad (3)$$

$$x(1) < 0, y^{\xi}(1) \geq 0; \quad \text{with certainty} \geq p_0 \quad (4)$$

$$x(1) < 0, y^{\xi}(1) < 0; \quad \text{with certainty} \geq p_0 \quad (5)$$

If it turns out that condition (3) is satisfied, then we will say that at the trading session at the time moment  $t = 1$  the first player achieved the desired result with certainty  $p \geq p_0$  and the DCC trading procedure is completed.

If it turns out that condition (4) is satisfied, then we will say that on trading session at the moment of time  $t = 1$  the second player has achieved the desired result with certainty  $p \geq p_0$  and the DCC trading procedure is completed.

If it turns out that condition (5) is satisfied, then we say that on of the trading session at the moment of time  $t = 1$ , both players did not achieve the desired result with certainty  $p \geq p_0$  and the DCC trading procedure is over.

If neither condition (3), nor condition (4), nor condition (5) are met, then the trading procedure at the session continues further for the time points  $t = 1, 2, 3...$  Setting the procedure for trading operations of players using relations (1), (2) generates at each moment of time a set of pairs of fuzzy sets  $\{H_t, n_t(\cdot)\} \times \{F_t, m_t(\cdot)\}$ . These fuzzy sets will reflect the processes of transition from the initial states  $(x(0), y^{\xi}(0))$  of the players to subsequent states. Transition processes occur as a result of the application of control actions by the players. The conditions of awareness of the players, set when setting the problem under consideration, allow, during the continuation of the trading session, to be in a situation similar to the moment the trading situation began and, therefore, to find a solution in this game.

Let's define the following function:  $F(\cdot) : X \rightarrow R_+$ ,  $F(x) = \{\sup m(y), \text{for } y \leq x\}$ ,  $\phi(0) = \inf\{\phi'\}$ ,  $F(\phi') \geq p_0$ . The value  $\phi(0)$  will be called the characteristic of uncertainty  $PL\_2$ . Denote by  $\Phi$  is the set of such functions, and by is the set  $T^* = \{0, 1, \dots\}$ , of natural numbers, including zero.

That is, the strategy  $PL\_1$  is a rule that allows  $PL\_1$ , based on the available information, to determine the amount of the DCC1  $PL\_1$  volume allocated for the purchase of the DCC2 cryptocurrency of the second player in the trading session. The second player chooses his strategy  $v(\cdot)$  based on any information. The first player seeks to find the set of his initial states  $x(0)$  (volumes of DCC1) and the uncertainty characteristics  $\phi(0)$  of the second player. These characteristics have the following property.

Property: if the game starts from them, then the first player can ensure the fulfilment of condition (3) at one of the time  $t$  by choosing his strategy  $u_*(\cdot)$ . At the same time, this strategy, chosen  $PL\_1$ , contributes to preventing  $PL\_2$  the fulfilment of condition (4) at previous times.

The set of such states will be called the preference set ( $W_1$ ) for the first player. Then,  $u_*(\cdot)$ —strategies  $PL\_1$  with the indicated properties will be called its optimal strategies. The goal  $PL\_1$  is to find the preference set, as well as to find its strategies, by applying which it will obtain the fulfilment of condition (3).

According to decision theory classification, the formulated game model corresponds to the problem of decision-making under conditions of fuzzy information. Moreover, the model is a bilinear multi-stage quality game with several terminal surfaces and simultaneous moves. Finding the preference sets of the first player and their optimal strategies depends on various parameters. The solution to the problem is found using the tools of the theory of multi-stage quality games [19–21], which allows for finding a solution for any ratio of game parameters. The article provides a solution to the problem from the point of view  $PL\_1$ . The task is to find a set of «preferences»  $W_1$  and optimal

strategies  $u_*(\cdot)$  for all ratios of game parameters. The solution of the problem from the point of view  $PL\_2$  is similar. Note that  $W_1$  is the union of the sets  $W_1^i$ . This is a set of such initial states  $(x(0), \phi(0))$ , that if the game starts from them, then there is a strategy  $PL\_1$ , which, for any implementations of the strategy  $PL\_2$ , «brings», at the time  $t = i$ , the state of the system  $(x(0), \phi(0))$  to such that condition (3) will be satisfied. At the same time,  $PL\_2$  does not have a strategy that can “lead” to the fulfilment of condition (4) at one of the previous points in time. The solution of the problem is to find the preference set  $PL\_1$  and its optimal strategies.

## 4.2 The Solution of the Problem

Let us introduce the notation:  $s_1 = 1 - (k_{pok}/k_{prod})$ ;  $s_2 = (1/k_{btc1}) - (1/k_{prod})$ ;  $s_1' = s_1 \cdot k_{btc1}$ ;  $s_2' = s_2 \cdot k_{btc1}$ .

There are (potentially) four cases: (a)  $s_1 > 0, s_2 \leq 0$ ; (b)  $s_1 \leq 0, s_2 > 0$ ; (c)  $s_1 > 0, s_2 > 0$ ; (d)  $s_1 \leq 0, s_2 \leq 0$ ;

In case (a) and  $(\beta/\alpha + s_2' - 1) > 0$ , there is an infinite (countable) number of preference sets  $W_1^i$  for the first player-ally that have the property that if  $(x(0), \phi(0)) \in W_1^i$ , then the player  $PL\_1$  will be able to obtain the fulfilment of condition (3) in steps  $i$ , no matter how the player  $PL\_2$ . Moreover, the player  $PL\_2$  has a strategy that does not allow the first player to obtain the fulfilment of condition (3) in a smaller number of steps. The set  $W_1^i$  is written like this:

$$W_1^i = \{(x(0), \phi(0)): k(i-1) \cdot x(0) \leq \phi(0) < k(i) \cdot x(0)\},$$

where  $k(i) = (\alpha/\beta) \cdot [-s_2 - s_2' \cdot k(i-1) + k(i-1)]$ ,  $k(0) = 0$ .

The union of sets  $W_1^i$  defines the preference set for the first player  $W_1$ , which is written as follows:

$$W_1 = \{(x(0), \phi(0)): \phi(0) < q \cdot x(0)\}, \quad (6)$$

where  $q = (-s_2)/[\beta/\alpha + s_2' - 1]$ ; and from any state  $(x(0), \phi(0))$  of this set the player  $PL\_1$  can achieve the fulfilment of condition (3) in a finite number of steps.

Note that the reciprocal of  $q = (-s_2)/[\beta/\alpha + s_2' - 1]$ ; is the equilibrium exchange rate (with confidence  $p_0$ ) for the DCCs under consideration with such a ratio of parameters.

In case (a) and  $(\beta/\alpha + s_2' - 1) \leq 0$ , there is a finite number of preference sets  $W_1^i$  for the first player-ally that have the property that if  $(x(0), \phi(0)) \in W_1^i$ , then the player  $PL\_1$  can obtain the fulfilment of condition (3) in steps  $i$ , no matter how the player  $PL\_2$ . Moreover  $PL\_2$ , the player  $PL\_1$  has a strategy that does not allow the player to obtain the fulfilment of condition (3) in a smaller number of steps. The set  $W_1^i$  is written like this:

$$W_1^i = \{(x(0), \phi(0)): k(i-1) \cdot x(0) \leq \phi(0) < k(i) \cdot x(0)\}, \quad (7)$$

where  $k(i) = (\alpha/\beta) \cdot [-s_2 - s_2' \cdot k(i-1) + k(i-1)]$ ,  $k(0) = 0$ .

The set union defines the preference set  $W_1^i$  of the first player  $W_1$ , which is the same as  $R_+^2$ .

The optimal strategy of the first player is to allocate the entire amount of the DCC1 cryptocurrency to the purchase of DCC2. The second player is «instructed» to refrain from buying DCC1.

In case (b), the player  $PL_1$  cannot “build” his own preference set, since with such a ratio of parameters, the situation becomes preferable for the player  $PL_2$ .

In case (c), players do not have preference sets, since with such a ratio of parameters they will have both DCC1 and DCC2, for as long as they like. Case (c) is impossible because, by definition, the buying rate of DCC2 cannot be greater than its selling rate.

In view of the above, we can say that only in two of the four cases are situations that lead either to an unlimited increase in the ratio  $(x(0)/\phi(0))$ , or to an unlimited increase in the ratio  $(\phi(0)/x(0))$ . In the remaining cases, this does not happen. Since an unlimited growth of the ratio  $(x(0)/\phi(0))$ , or ratio  $(\phi(0)/x(0))$  is undesirable, for this it is necessary to “narrow” the area  $W_1$  in order to increase the area of stability of the ratio  $(\phi(0)/x(0))$ . As you can see, this becomes possible if the ratio  $(\beta/\alpha)$  is arbitrarily large, i.e. the growth rate of DCC2 in the foreign exchange market should significantly exceed the growth rate of DCC1.

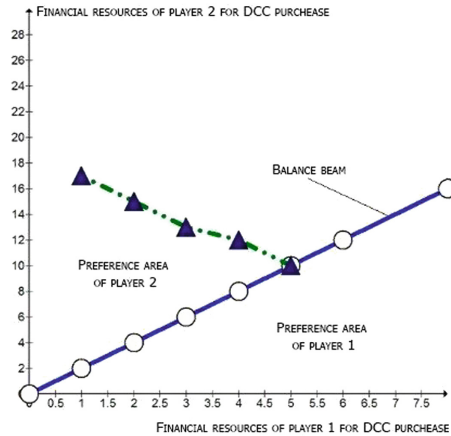
## 5 Computational Experiment

A computational experiment for the problem of finding the area of preference for an investor in DCC2 was carried out in the MathCad environment. The result is shown in Fig. 1.

The computational experiment reflects a situation where the ratio of initial parameters of the DCC partners is such that their cryptocurrencies are in the second player’s preference area with a given degree of certainty. However, the second player does not apply their optimal strategy while the first player does. In this case, the trajectory moves in the direction of the first player’s preference set. If such actions continue, the second player may suffer losses. The trajectory was constructed using a program created based on an algorithm for finding players’ strategies and preference sets. The algorithm was developed by applying the Bellman optimality principle, and selected statistical data from sources [12–14] were used. The paper presents the results of only one experiment due to space limitations. It should be noted that it is positive that the players’ strategies are found constructively in an explicit form for all ratios of interaction parameters. This reduces the complexity of the calculations required to obtain the results. Computation does not require high-power computers.

## 6 Discussion of the Results of a Computational Experiment

Figure 1 demonstrates the situation in which the player  $PL_2$  (buyer of DCC2, using the non-optimal behavior of the buyer of DCC1 at the initial moment of time, achieves what “brings” the state of the system to “its” terminal surface. If the buyer’s trajectory coincides with the balance beam, then this will correspond to situations of the equilibrium value of the DCC exchange rate. In this case (however, unlikely), the players, applying their optimal strategies, «move» along this ray. This «satisfies» both players simultaneously. If the movement trajectory is under the beam of balance, then this will



**Fig. 1.** Results of the computational experiment. The trajectory of the movement of the player (buyer DCC2)

illustrate the situation when the player  $PL_1$  (buyer of DCC1) has an advantage in the ratio of parameters, i.e. they are in the preference set  $PL_1$ . In this case  $PL_1$ , applying its optimal strategy, will achieve its goal, namely, bringing the state of the system to «its» terminal surface.

## 7 Conclusions

The game model for trading sessions in the digital crypto-currency (DCC) market with fuzzy information is discussed in this article. It is demonstrated that the controllability of the trading process can be described through a game approach by solving a system of discrete bilinear equations. This approach allows for considering cases that lead to instability in the ratio of financial resources of players, such as DCC1 to DCC2, and vice versa, and does not exclude the implementation of the game by a common collective counterparty. The model is novel in that it solves a bilinear multi-stage quality game with several terminal surfaces with fuzzy information, which distinguishes it from existing approaches. A solution is obtained for a new bilinear multi-stage quality game with dependent motions. The article also presents the results of a computational experiment that takes into account various ratios of parameters describing the process of buying and selling DCC. The presented results may be beneficial for preventing exchange rate instability in the market of investments in DCC under conditions of fuzzy information that typically occur in practice. Additionally, the model can be useful for predicting situations on trading floors that trade DCC and providing recommendations for the choice of control actions to maintain exchange rate stability in the market for investing in cryptocurrencies at the level of major banking players.

## References

1. Trimborn, S., Li, M., Härdle, W.K.: Investing with cryptocurrencies—a liquidity constrained investment approach. *J. Financ. Economet.* **18**(2), 280–306 (2020)



2. Petukhina, A., Trimborn, S., Härdle, W.K., & Elendner, H.: Investing with cryptocurrencies—evaluating the potential of portfolio allocation strategies (No. 2018-058). IRTG 1792 Discussion Paper (2018)
3. Angerer, M., et al.: Objective and subjective risks of investing into cryptocurrencies. *Fin. Res. Lett.* **40**, 101737 (2021)
4. Maiti, M., Vukovic, D., Krakovich, V., Pandey, M.K.: How integrated are cryptocurrencies. *Int. J. Big Data Manage.* **1**(1), 64–80 (2020)
5. Trimborn, S., Mingyang, L., & Härdle, W.K.: Investing with cryptocurrencies (2017)
6. Baur, D.G., Dimpfl, T.: Asymmetric volatility in cryptocurrencies. *Econ. Lett.* **173**, 148–151 (2018)
7. Peng, Y., Albuquerque, P.H.M., de Sá, J.M.C., Padula, A.J.A., Montenegro, M.R.: The best of two worlds: forecasting high frequency volatility for cryptocurrencies and traditional currencies with support vector regression. *Expert Syst. Appl.* **97**, 177–192 (2018)
8. Charfeddine, L., Maouchi, Y.: Are shocks on the returns and volatility of cryptocurrencies really persistent? *Financ. Res. Lett.* **28**, 423–430 (2019)
9. Caporale, G.M., Zekokh, T.: Modelling volatility of cryptocurrencies using Markov-switching GARCH models. *Res. Int. Bus. Financ.* **48**, 143–155 (2019)
10. Glas, T.N.: Investments in cryptocurrencies: handle with care! *J. Alternative Invest.* **22**(1), 96–113 (2019)
11. Bianchi, D.: Cryptocurrencies as an asset class? An empirical assessment. *J. Alternative Invest.* **23**(2), 162–179 (2020)
12. Böyükaslan, A., Ecer, F.: Determination of drivers for investing in cryptocurrencies through a fuzzy full consistency method-Bonferroni (FUCOM-F'B) framework. *Technol. Soc.* **67**, 101745 (2021)
13. Dempsey, M., Pham, H., Ramiah, V.: Investment in cryptocurrencies: lessons for asset pricing and portfolio theory. *Appl. Econ.* **54**(10), 1137–1144 (2022)
14. Bharadwaj, S., & Deka, S.: Behavioural intention towards investment in cryptocurrency: an integration of Rogers' diffusion of innovation theory and the technology acceptance model. *Forum Scientiae Oeconomia* **9**(4), 137–159 (2021, December)
15. Petukhina, A., Trimborn, S., Härdle, W.K., Elendner, H.: Investing with cryptocurrencies—evaluating their potential for portfolio allocation strategies. *Quant. Fin.* **21**(11), 1825–1853 (2021)
16. Bebeshko, B., Malyukov, V., Lakhno, M., Skladannyi, P., Sokolov, V., Shevchenko, S., Zhumadilova, M.: Application of game theory, fuzzy logic and neural networks for assessing risks and forecasting rates of digital currency. *J. Theor. Appl. Inf. Technol.* **100**(24) (2022). <http://www.jatit.org/volumes/Vol100No24/15Vol100No24.pdf>
17. Bebeshko, B., Khorolska K., Desiatko, A.: Analysis and modeling of price changes on the exchange market based on structural market data. In: 2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T), 2021, pp. 151–156. <https://doi.org/10.1109/PICST54195.2021.9772208>
18. Cryptocurrency Helium has become the most profitable for the day. Electronic resource. <https://noworries.news/kryptovalyuta-helium-stala-najprybutkovishoyu-za-dobu/>
19. Malyukov, V.P., Linder, N.V.: A multistep game of kind between two economic systems under complete information. *Cybern. Syst. Anal.* **30**(4), 545–554 (1994)
20. Lakhno, V., Malyukov, V., Kasatkin, D., Chubaieskyi, V., Rzaieva, S., Rzaiev, D.: Continuous Investing in Advanced Fuzzy Technologies for Smart City. *Lecture Notes on Data Engineering and Communications Technologies*, 142, pp. 313–327 (2023)
21. Kartbayev, T., Lakhno, V., Malyukov, V., Turgynbayeva, A., Alimseitova, Z.H., Malikova, F., Kashaganova, G.: Model for the decision support system during the procedure of investment projects assessment in the field of enterprise digitalization considering multifactoriality. *J. Theor. Appl. Inf. Technol.* **100**(7), 1684–1692 (2022)



# A Comparative Analysis of the Impacts of Traditional and Digital Billing Methods

Tasnim Faruki<sup>1</sup>, Rafa Tasnim<sup>1</sup>, Malyha Bintha Mabud<sup>1</sup>, Rashedul Amin Tuhin<sup>1</sup>(✉),  
Ahmed Wasif Reza<sup>1</sup>(✉), and Mohammad Shamsul Arefin<sup>2</sup>(✉)

<sup>1</sup> Department of Computer Science and Engineering, East West University, Dhaka 1212,  
Bangladesh

{mcctuhin, wasif}@ewubd.edu

<sup>2</sup> Department of Computer Science and Engineering, Chittagong University of Engineering and  
Technology, Chattogram 4349, Bangladesh

sarefin@cuet.ac.bd

**Abstract.** Bisphenol A or BPS is frequently used in thermal paper for POS machines to give receipts. Compared with many studies, this chemical substance impacts both the environment and human health. With current technology, this primitive system can easily be replaced. However, every change faces some obstacles. To overcome this, the main target is to identify its impacts and the benefits and problems of the alternate method. Mostly these findings will be done through the previous record and people's opinions. After analysis, several harmful effects of thermal paper are detected. Moreover, opinions are collected for the alternative to paper billing, digital billing. Furthermore, most of the opinions were biased due to the mindset of people. In Bangladesh, thermal paper is being used in POS machines almost everywhere. Millions of trees are getting cut off to produce thermal paper. Also, it causes a threat to human life. So, digitalizing it is essential. From the study, these subjects are thoroughly explored. Also, the debates between paper billing and digital billing are being analyzed. Because of the traditional mindset, people are hesitant to move from a paper-based billing process to a digital one, though it has positive practical impacts. The outcome of this study could benefit those who want to switch to digital billing in the future.

**Keywords:** Bisphenol A · Point of sale machine · Digital billing · Paperless

## 1 Introduction

Thermal paper is a specific type of paper that has a coating made of a substance that changes color when heated. Realizing the health and environmental effects that thermal paper is causing and trying to solve the problem by finding an eco-friendly solution is necessary. It will be very efficient and important to everyone as it is related to the environment, health and all above it is related to the future. Going green is about protecting and preserving the environment, habitats, and biodiversity. Moreover, surviving on earth and saving it is dependent on green computing. By reducing the usage of thermal paper,

the environmental and health impacts caused by it can be demolished, and, in this era, it is the most important thing to do for our earth. Looking at the past, our ancestors did not experience the enormous explosion of thermal paper; thus, it is a kind of modern trade. The local shops around us do not use thermal paper for their billing process, but super-shops use this. As modern Bangladeshi citizens, Shopping at the super shops where thermal paper is used is very common. So, it got our attention, and it is high time that everyone understands the negative impacts it is causing, mostly the younger generation being exposed to this daily. Anyone concerned about the environment and interested in going paperless with a process to save energy, consume fewer resources, and be eco-friendly will be interested in this topic. This study will reveal the amount of thermal paper usage that can be avoided, the reasonable amount of money that can be saved, and the drivers and barriers of the digital billing process. The benefits that will be provided and the difficulties that might occur with digital billing will be analyzed.

In traditional super shops, around 400 rolls of thermal paper are used annually in Bangladesh. In the UK, over 11 billion receipts are produced each year, yet 9.9 billion of them are wasted, which is the same as destroying 53,000 trees. In the USA, paper receipts are the reason for cutting down 10 billion trees. So the influence on the environment can be speculated. The thermal paper contains harmful substances like BPA (Bisphenol A) or BPS (Bisphenol S), which affect the human body. In Bangladesh, most super shops use Thermal paper in billing, which harms our health and environment. Most of the invoice paper is thrown away after purchase, so they are irrelevant. So, this study aims to understand the impacts on the environment and the economy of the traditional billing process and to analyze the advantages and disadvantages of digital billing. In this process, different kinds of challenges can arise. This study will be conducted to identify those challenges and their benefits. Digital billing has many benefits. For example, the bill can send through text messages to the customer's respective phone numbers, which requires a minimal charge. A database can be used for storing all the pieces of information of the customers. Furthermore, the transaction can be done relatively quickly without any turbulence. However, it can also have many challenges, like changing companies' policies, being unfamiliar with the digital system, not being comfortable enough to replace the as-is model, sharing contact information to shop employees, etc.

In the modern age, receipts play a vital role in trading goods as they keep an accurate record. Mostly thermal paper is used as cash receipts in the super shops because it is more convenient in the printing process and has low-cost maintenance than usual paper. This thermal paper is made of various substances. One of the crucial elements is Bisphenol A (BPA) or Bisphenol S (BPS) [1]. These chemical substances harm the human body and the environment [2]. About 90% of exposure to BPA or BPS is caused by these invoice papers [3]. Cash receipts are not only prejudicial to human health but also to the environment. In the USA, almost 10 billion trees and 1 billion gallons of water, and 250 million gallons of oil are used to make receipts per year [4]. It leaves an adverse impact on the environment which causes deforestation resulting in greenhouse gas emissions. In a calculation, 640000 tons of paper emit around 4,000,977,751 lb of CO<sub>2</sub>. It also wastes 13 billion gallons of water every year [5]. In Bangladesh, initiatives still need to be taken to reduce the usage of thermal paper. Mostly, people do not recognize the

damaging effects of thermal paper. So, there is a lack of information and initiatives in Bangladesh. This leads to the following Research Questions.

- What are the impacts of thermal paper usage?
- How much thermal paper is used in POS machines?
- Are there any benefits if the regular paper billing process is changed?
- Are the users aware of the problems with the regular billing process?
- What kind of problems can arise because of digital billing?

So, this study will be conducted to find some answers to these arising questions. Thus, initially, some objectives for this study have been associated. This study's primary focus is to analyze the impacts of the thermal paper used in the super-shops. Besides, the thermal paper usage in the super-shops will be estimated. Finally, the benefits and demerits of digitalization of the billing process will be identified. The primary contribution of this paper is that, earlier, there was no study on the environmental impact of thermal paper in Bangladesh. BPS and BPA have harmful effects, which can harm the environment and human health. They are widely used in Bangladesh and can be replaced by digital technology that will reduce the environmental impact.

The remaining parts of the study are structured as follows: Sect. 2 discusses the literature review. Section 3 contains illustrations of materials and methods. Next, in Sect. 4, the experimental results and discussion are represented. Finally, Sect. 5 is the last section of this study.

## 2 Literature Review

According to the paper [6], in 2015, the global BPA consumption was 7.7 million metric tons. It was 8 million metric tons in 2016 and is about to reach 10.6 million in 2022, with a CAGR of 4.8% between 2016 and 2022. In 2015 however, the EFSA declared that it does not create any health risk, so the EU (European Union) declared BPA as an approved product to be used, but a study on BPA and its potentially harmful effects is still ongoing. This study evaluates the problem by reviewing the current legislation on BPA in food items.

In [7], 95% of Canadians have BPA endocrine disruptors in their urine. This study aims to determine if the amount of BPA levels found in thermal paper receipts are high enough to generate a public health concern. The study will obtain the result by soaking the thermal paper and incubating 100(mg) of thermal paper in 10(mL) methanol for three hours at room temperature, then soaking it overnight at 4 °C. As a result, 13 out of 30 samples had BPA at a range of (0.124–871.17) mg BPA per kg. As the provisional daily intake set by Health Canada is (0.025 mg/kg body weight/day), the result shows that there can be enough BPA present in thermal paper to harm humans.

The paper [8] works on the concentrations of BPA and BPS found in thermal paper receipts. These invoices were procured in Italy. BPA was discovered in 44 instances, and the mean concentration level was 107.47  $\mu\text{g}/100\text{mg}$ . BPS was discovered in 31 instances, and the average concentration level was 41.97  $\mu\text{g}/100\text{mg}$ . BPA and BPS both were discovered in 26 instances. This study concludes that though the dermal intake is below tolerable, its existence in non-dietary and dietary sources could be deliberate as a health risk.

In [9], the authors show the principal disadvantages of thermal paper receipts as it contains BPA, which harms the environment and human health. To establish a paperless payment chain, it offers a solution integrating the NFC regulation in smartphones and a virtual receipt printing device.

In [10], Supposedly, BPS has been employed in consumer products as a BPA alternative. BPS has a chemical structure similar to BPA, which is an exogenous endocrine-disrupting chemical. There has been much study on BPA, but not enough study has occurred on BPS. This study discusses exposure to BPS in water, sludge, sediment, air, dust, consumer products, and human urine. The paper concludes that further investigation of BPS and its effects on human exposure and the environment is required.

In [11], the authors have compared the paper billing process with the electronic billing process. It has been said that the electric bill is increasing while paper-based billing is decreasing. The study compares the energy and Input material for paper-based and digital billing by evaluating 9 environmental impact categories. They have come out with the result that the environmental effects of an electric billing system are substantially less than a billing method that uses paper.

In [12], the environmental effects of implementing electronic invoicing have been explored. The electric and paper invoicing process has been mapped, and the steps of generating a carbon footprint have been identified for both parts. As a result, it was found that the incoming and the outgoing carbon footprint of paper invoicing is 214,62 gCO<sub>2</sub> and 336,66 gCO<sub>2</sub>, respectively, whereas, for electronic invoicing, the incoming and the outgoing carbon footprint is 48,39 gCO<sub>2</sub> and 154,82 gCO<sub>2</sub>.

### 3 Materials and Methods

#### 3.1 Data Collection Methods

Three methods can be employed in this study for assembling data. These are interviews, surveys, and document studies. An interview is a meeting in which one or more persons consult or question another person to gather information. A survey collects data from a specific individual through their responses. A document study is mainly analogizing previously published documents to interpret specific data. For this study, based on different criteria, a mixed method is going to be used.

In [13], the study was done by analyzing different documents related to it where it was conducted by collecting blood samples from an Italian marketplace. However, for this study collecting medical data is quite tricky. On the other hand, a document study is preferable for this study. So, to analyze the impacts of thermal paper, a document study will be conducted for this.

Again another study [14] was conducted by interviewing the financial manager and other participants regarding the digital invoice system. The author also took some open interviews to get ideas and opinions. As for this study, interviews are going to be held. The interview session will be conducted where POS machines are used, such as super shops, chain shops, etc. Both qualitative and quantitative will be collected because of the mutual exclusiveness. To estimate the cost related to thermal paper usage, the store manager and the employee interview will be taken. Economic analysis of the digital billing process and its comparison with paper-based billing will be shown through document studies.

Also, opinions about the digital billing process from the shops will be collected. Overall, all the qualitative and quantitative data are going to be collected & used for further study. And none of the data will be used for any harmful activities.

The questions of the interview were mainly designed in such a way that it fulfills our objectives 2 and 3. The different questions asked were:

- Q1. How much role usage per day?
- Q2. Per role price?
- Q3. Do you print the receipt for everyone?
- Q4. What is the minimum length for each receipt?
- Q5. After how many days does the POS machine need to be repaired?
- Q6. What types of problems do you face while using a POS machine?
- Q7. What do you do with the other part of the receipts?
- Q8. Do you record the purchase list in the database?
- Q9. What is your opinion about using a digital receipt instead of a POS machine?

### **3.2 Participants**

Numerous shops around Bashabo, Khilgaon, and Aftab Nagar were visited, which includes six grocery stores, four food chain shops, and two household commodity shops. Many other research papers have collected data on a larger scale. In [12], the study was done on a kitchen and furniture fittings company in Finland. In [15], data were collected from several companies and banks around Iraq. Although our data collection scale was much smaller than the other studies, still data from 12 different stores were collected. The shop managers and the employees were kind enough to participate in our study. They helped us by providing quantitative information and sharing their thoughts and opinions on the digital billing system.

### **3.3 Data Analysis**

The data analysis is based on paper receipts and their replacement. At first, The documentation study was done to analyze the impacts of thermal paper on the environment. This was qualitative data analysis. The data was collected from three kinds of chain shops. The analysis was made based on POS machines. The data was collected on the usage of thermal paper rolls every month and the problem they faced by using the POS machine. The information was gathered about the POS machines' supervision and the expenses of the thermal paper roll. The shop managers' data about the digital billing process and the current billing process were also noted.

Interviews were held at two grocery chain shops, and an overview of their billing process was recorded. Most shops use around 13–15 thermal paper rolls per month. No routine maintenance for the POS machine was followed. The information about customer purchases was collected on their database systems.

Data was collected from food chain shops and household commodity shops. The food chain shop requires 16–17 paper rolls in a month. The POS machines are not regularly checked there. And the household commodity shops use 12–16 rolls per month. Their machines get checked every two months.

In general, all three types of chain shops must print paper receipts whether customers want them or not. Moreover, the minimum length is the same for almost every type of shop. They have mixed reactions to the digital billing process and paper-based billing.

The data and the monthly expenditure for the paper rolls are computed for this study. The frequency of POS machine maintenance is also calculated. The data analysis is qualitative and quantitative as the comparison of paper-based billing with the digital billing process is committed.

### **3.4 Research Ethics**

All the interviewing ethics were followed throughout the process of data collection. All the participants were given a consent form with the study details, and the data usage was described. The interview was conducted only when the participants agreed with all the terms and conditions and signed the consent form. The contact numbers were also provided so they can contact if required. The information on the usage of the data and the purpose of this data collection was narrated to them. The safety of their given data was also confirmed.

## **4 Results and Discussion**

The data relating to the thermal paper used in POS machines was collected by physical inspection after visiting different chain shops for food, groceries, and home appliances. The concerned employees handling these POS machines, like the receptionists and managers, were asked in detail about the usage of the thermal paper. The managers of each shop helped in the collection of data.

While interviewing, details like how many customers can be served per roll of thermal paper, the length of the minimum receipt paper, etc. were asked. Also, it was asked whether they have a policy to print the invoice paper mandatory for every customer. The questionnaires also included the number of roles used a day, their prices, maintenance of POS machines, etc. Finally, their opinion about digital billing was asked through thorough discussion and observation.

### **4.1 Experimental Result and Analysis**

Thermal paper is one of the most stereotypical paper billing elements which has been used all over the world in POS machines for decades. It contains BPA or BPS, which has its demerits and is more acute in long-time usage. The study [16] found that from the sample of 619 ADHD symptom children, greater than 97% BPA and 42% BPS were found in 6 years old children. According to [17], long-time exposure to BPA and BPS results in potent endocrine disruptors. Similar issues are also found in the female reproductive system. Moreover, it was found in the [18] study. Their study states that the blood estradiol levels following gonadotropin stimulation, the number of retrieved oocytes (immature ovum), the number of adequately fertilized oocytes, and implantation have all been reported to be adversely correlated with BPA exposure. [19] says that it has been restricted in many countries due to the harmful effects of BPA. Even in the EU,

it has been banned. [20] states that thermal paper is one of the most color-developed papers. Furthermore, it can easily migrate onto the skin or be absorbed in the body while handing in receipts, tickets, etc. Even in the year 2020, there was no exact solution to it. Their findings also state that public health is also at risk due to the contamination of the BPA/BPS of thermal paper. From this study [21], emerging contaminants and climate change are two of today's major threats to soil organisms worldwide. Biosolids used in agriculture may contain xenobiotics (metabolic organisms) not controlled in soil, such as bisphenols. This known toxicity of Bisphenol A (BPA) has driven the research for substitutes. In this study [22], their findings say that BPA can easily be mixed up with the ecological system of aquatic plants and animals. Even it creates a chemical reaction with the ozone layer.[10] also states that millions of trees are being cut down yearly just to make thermal papers worldwide. So, the chemicals of thermal paper are not the only evil but also the reason for cutting down many trees, adversely affecting the environment. So, overall, it can be said that the impacts of the BPS or BPA of thermal paper are a concerning matter.

After interviewing 12 different shops, we collected data on the usage of thermal paper in each store, and we came up with the estimated cost for the traditional paper-based billing process. The price per 78x56 mm thermal paper roll is 60 taka. Here, we have excluded the POS machine price and maintenance costs.

The average thermal paper usage of 4 food chain shops is around 26.75 roles per month. There are around 200 outlets of this chain shop in Bangladesh.

**Table 1.** Thermal paper roll usage in food chain shops

Number of shops	Place	The total number of outlets
1	Aftab Nagar	15–18
2	Khilgaon	29–30
3	Bashabo	25–29
4	Taltola	25–30

This particular Food chain shop shown in Table 1 uses approximately 27 thermal paper rolls per month. So it concludes the usage of 324 rolls per year in one shop.

**Table 2.** Economic analysis in food chain shop

Estimated roll usage per month	Estimated roll usage per year	Estimated cost in one year per shop (TK)	Estimated total cost (TK)
27	324	19,440	3,888,000

We have interviewed a total of 4 different outlets. The collected data are given below. The average of 4 shops is considered, which is around 20.75, which is around 21 roles per month. There are 53 outlets of this shop in Bangladesh.



**Table 3.** Thermal paper roll usage in grocery shop-A

Number of shops	Place	The total number of outlets
1	Aftab Nagar	15–16
2	Aftab Nagar	22
3	Bashabo	30
4	Tilpapara	15

This particular Grocery shop(A) uses approximately 252 rolls per year in one shop. The monthly usage is 21 rolls.

**Table 4.** Economic analysis in grocery shop-A

Estimated roll usage per month	Estimated roll usage per year	Estimated cost in one year per shop (TK)	Estimated total cost (TK)
21	252	15,120	801,360

We have interviewed a total of 2 different shops. The collected data are given below. The average of 2 shops is considered, which is around 32.5 roles per month. There are 53 outlets of this shop in Bangladesh.

**Table 5.** Thermal paper roll usage in grocery shop-B

Number of shops	Place	The total number of outlets
1	Banasree	34–35
2	Tilpapara	28–30

This particular Grocery shop(B) shown in Table 5 uses around 33 thermal paper rolls per month. It concludes the usage of 396 rolls per year in one shop.

**Table 6.** Economic analysis in grocery shop-B

Estimated roll usage per month	Estimated roll usage per year	Estimated cost in one year per shop(TK)	Estimated total cost (TK)
33	396	23,760	1,259,280

We have interviewed a total of 2 different shops. The collected data are given below. The average of 2 shops is considered, which is around 12.5 or approximately 13 roles per month. There are 244 outlets of this shop all over Bangladesh.

**Table 7.** Thermal paper usage in household commodity shops

Number of shops	Place	The total number of outlets
1	Aftab Nagar	15–16
2	Khilgaon	8–9

This Household commodity shop uses around 156 rolls per year. And in one month it is 13 rolls per shop.

**Table 8.** Economic analysis in household commodity shops

Estimated roll usage per month	Estimated roll usage per year	Estimated cost in one year per shop (TK)	Estimated total cost (TK)
13	156	9360	2,283,840

**Table 9.** Economic comparison between traditional billing and digital billing

No of invoices	Cost in SMS (TK)	Cost in POS machine(TK)	Economic benefit (TK)
111,111	22,222	229,884	207,662

Replacing paper receipts will be economically beneficial, as from Tables 1, 3, 5, and 7. It can be seen that every year around 156–396 thermal paper rolls are needed just for one shop. And it costs a lot of money just on the thermal paper rolls. Tables 2, 4, 6, and 8 show that every year around 8 lakhs to 38 lakhs TK is spent on thermal papers only. POS machines cost around 16,000–20,000 TK, and maintenance cost varies on the repair process.

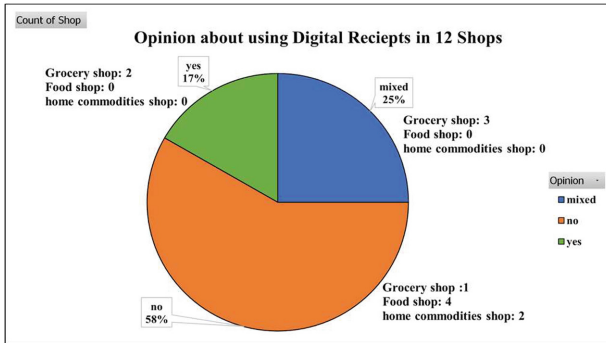
In Bangladesh, the standard price per SMS is 0.20 TK for any enterprise. According to this, an enterprise can buy 111,111 SMS with 22,222 TK from any telecommunication company.

On the other hand, the size of the thermal paper roll is 78x56 mm and the minimum serving size of invoice paper is 6 inches or 152.4mm. So, per thermal paper roll, around 29 people can be served with invoice paper. Now for 111,111 invoices, it requires around 3,832 thermal paper rolls which cost approximately 229,884 TK.

Table 9 illustrates the economic benefits of using digital billing. By using a digital billing process instead of paper-based billing, around 207,662 TK can be saved based on replacing 111,111 invoice papers.

The drivers and barriers of the digital billing process were also reflected in the minds of those who are involved with paper-based billing systems. Some of them are directly connected and some are indirect. But both party has influence regarding this.

Through questioning the shop employees, 17% of them are optimistic about digital billing. They did understand the environmental and economic benefits of using digital



**Fig. 1.** Opinion about driver barriers of digital billing

billing. They also feel that digital billing will be a much easier method to work with and will be more time and energy efficient. They were welcoming of the change and gave us a positive response.

From Fig. 1, 58% of the responses were negative. The reason behind their responses had so much to do with customer behavior. One shop employee said that it is related to the customer's security. Many female customers might feel uncomfortable providing their contact number or email address required for digital billing. Another employee responded by talking from the point of view of older people. Many older people might not be comfortable with using digital billing systems. They might need help understanding the process or finding it too complicated. Another response was about the customers not being interested in providing their contact numbers because of the promotional advertising through messages. Advertisement through mobile messaging is done by many super-shops nowadays, and Many people find it irritating. Many customers may think that providing their contact information might make them vulnerable.

Moreover, the final response was about the mindset of the company owners. An employee responded that it might be easy to think about 1 or 2 shops to be converted into a digital billing process. However, implementing this system might not be accessible if we consider the hundreds of outlets in these shops. It will need a much bigger initiative. The mindsets of the company heads and the customers must be changed. No one will be interested enough to fix something which is not broken.

From Fig. 1, 25% of them provided a mixed review. Though they understand the benefits of digital billing, they are not that welcoming of it. They are not sure about the customer responses to this change. It is not uncommon for people to be resistant to change, especially when it comes to something they are used to or comfortable with.

All the shops agreed that they printed receipts for everyone. The main reason for that is the company policy. Though many customers do not request a receipt, they must print it to follow the rules. One employee said he found many customers who did not take the receipt. In this case, discard those receipts. He also said that if the policies can be such that the customers can choose whether they want receipts or not, which can reduce this wastage.

## 4.2 Discussion

The research results showed that the BPS in thermal paper harms human health as it can quickly contaminate the blood flow and trigger some serious diseases. On the other hand, the environmental impacts are also severe as a considerable amount of wood is needed to make it. From an economic point of view, a huge amount of money can be saved if this paper billing system is annulled because, every year, a huge number of thermal papers are used on POS machines. According to this study, in the food chain shop, approximately 324 thermal paper rolls are being used yearly, and in the grocery shop, the number of users is lesser and stands at 312 thermal paper rolls a year. For the household commodity shops, it is 132 thermal paper rolls per year. So, it can be said that the number of thermal papers is used more at food shops or daily commodities shops. The digital billing process is a superior alternative to paper billing for health, environment, and profit. Not only that, but digital billing is also much faster and smoother, and harmless. However, the customers, as in users, are unaware of it. Furthermore, they are not even aware of the negative impacts of paper billing. And there is a lack of customer data and some social norms and beliefs against digital billing. However, some problems can arise through digital billing, such as customer security and comfort can be compromised. Also, the company owners have to invest a large amount of money in adapting and transforming the new billing system over time, it is cost-effective.

The study [6] showed BPA causes severe health risks. Also, the study [11] said that BPS has adverse environmental impacts. These studies have concluded that BPA and BPS have an unfavorable influence on human health and the environment.

The study [4, 7] showed that paper-based billing is more expensive than digital billing. From our findings, we also get the same results. Digital billing can be less costly than paper-based billing systems.

This study [5] shows the demerits of a paper-based billing system. With the development of technology and the IT sector, it is easy to change to digital billing, and nowadays, people are more interested in it. From this study, in our country, the customers are still into traditional receipt even though switching to a digital-billing process with the current set-up is possible.

This study evaluates the effects of the paper-based billing process. It also describes the economic benefits of digital billing and compares it with thermal paper-based billing. The results signify how BPS/BPA, present in the thermal paper, is destroying the environment and human health by entering our bodies. It also shows the economic benefits of replacing the paper-based billing process with a digital billing process. If the customers and the company owners understand the advantages of digital billing and change their mindset, applying digital billing in super-shops will be much more accessible and understandable.

This study explores the outcomes of using a thermal paper-based billing process and the economic expenses, it does not collect the opinion of customers' points of view regarding this issue directly from them. It also shows a cost estimation of digital billing. Besides that, the power consumption was not calculated manually for switching to the electronic billing system. For this study, this data has been collected from a documentation study.

Previously in our country, there were very few studies regarding this issue. It could be advantageous for the development of this sector.

### 4.3 Limitations and Future Works

The data have been collected from several super-shops. Therefore, shopping malls, ATM booths, and other places where POS machines are being used are not considered. The average use of thermal paper in those places may vary. Again, the data is collected from the shop managers' and employees' points of view, so there is a gap in information regarding the opinions of customers and owners. According to this study, most store managers are unwilling to switch to digital billing even if they understand the outcomes of using thermal paper. So, the opinion regarding these issues can have impurities. The scope of our research is limited to thermal paper-based receipt paper only. However, there are other paper-based billing methods like carbon paper, carbonless paper, wood paper, etc. Again, every POS machine needs a certain amount of electricity to run. These electrical costs are also excluded from this study. So, in the future, a study based on the medical data on BPA/BPS dilution in the blood can be collected from our country. Also, the software can be developed to digitize the billing process.

## 5 Conclusion

Thermal paper is used for the billing process in Bangladesh but it contains a crucial substance, BPA or BPS. It is not only harmful to human health but also to the environment. It causes various diseases by mixing with our blood through our skin. Every tremendous number of trees is being cut off to make these paper receipts which causes global warming. It is one of the reasons for deforestation and the increase in greenhouse gas emissions. These paper receipts have a more destructive impact on lives than blessings. Almost every super shop in our country uses thermal paper in their POS machines. Typical shops require about 360 to 400 paper rolls annually. A massive amount of paper rolls are being used in our country daily. The current paper-based billing process can be changed by digitizing it. It can protect many trees from being cut off and helps to reduce greenhouse gas emissions. It will also rescue us from getting in contact with BPA or BPS. It has fewer environmental impacts than paper-based processes. Digital Billing has many other advantages. It can cause economic benefits by saving money used on paper rolls and POS machines. Digital billing is an efficient and more secure process. It reduces errors, shortens the payment time, reduces the cost of print and postal, and reduces an employee's workload. Digital billing has advantages and disadvantages from our country's perspective. It can help reduce greenhouse gas emissions associated with paper production and transportation. However, it is also true that digital billing consumes energy and resources, such as electricity and data storage. It is high time we need to make greener choices. In our country, most people do not know about the effects of thermal paper. Its environmental damages and the resulting health hazards should be known to all.

## References

1. Serkan Yalcin, M., Gecgel, C., Battal, D.: 2 (4) (PDF) Determination of bisphenol A in thermal paper receipts 9 (2016). <https://doi.org/10.18596/jotcsa.21345>

2. BPA Is Found in Paper Receipts - The New York Times. <https://archive.nytimes.com/green.blogs.nytimes.com/2011/11/01/check-your-receipt-it-may-be-tainted/>. Last accessed 11 Jul 2022
3. Liao, C., Bisphenol, S., et al.: A new bisphenol analogue, in paper products and currency bills and its association with bisphenol a residues. *Environ. Sci. Technol.* **46**(12), 6515–6522 (2012). [https://doi.org/10.1021/ES300876N/SUPPL\\_FILE/ES300876N\\_SI\\_001.PDF](https://doi.org/10.1021/ES300876N/SUPPL_FILE/ES300876N_SI_001.PDF)
4. Going Paperless: The Hidden Cost of a Receipt | HuffPost Impact. [https://www.huffpost.com/entry/going-paperless-the-hidde\\_b\\_3008587?ec\\_carp=4408988367690251169](https://www.huffpost.com/entry/going-paperless-the-hidde_b_3008587?ec_carp=4408988367690251169). Last accessed 11 Aug 2022
5. Porter, B., Tamsamani, A.: *Skip The Slip*. Washington, DC 20006 (2018)
6. Almeida, S., et al.: Bisphenol A: food exposure and impact on human health. *Compr. Rev. Food Sci. Food Saf.* **17**(6), 1503–1517 (2018). <https://doi.org/10.1111/1541-4337.12388>
7. Ong, R.W. et al.: Bisphenol A (BPA) in thermal paper used for receipts. *BCIT Environ. Heal. J.* (2016)
8. Russo, G., et al.: Monitoring of bisphenol A and bisphenol S in thermal paper receipts from the Italian market and estimated transdermal human intake: a pilot study. *Sci. Total Environ.* **599–600**, 68–75 (2017). <https://doi.org/10.1016/j.scitotenv.2017.04.192>
9. Arva, M.C. et al.: Electronic receipts using near-field communication protocol as a solution for thermal paper receipts. In: *Proceeding 12th International Conference Electronic Computer Artificial Intelligent ECAI 2020* (2020). <https://doi.org/10.1109/ECAI50035.2020.9223257>
10. Wu, L.H., et al.: Occurrence of bisphenol S in the environment and implications for human exposure: a short review. *Sci. Total Environ.* **615**, 87–98 (2018). <https://doi.org/10.1016/j.scitotenv.2017.09.194>
11. (9) (PDF) Electronic billing vs. paper billing: dematerialization, energy consumption and environmental impacts. [https://www.researchgate.net/publication/261301335\\_Electronic\\_billing\\_vs\\_paper\\_billing\\_Dematerialization\\_energy\\_consumption\\_and\\_environmental\\_impacts](https://www.researchgate.net/publication/261301335_Electronic_billing_vs_paper_billing_Dematerialization_energy_consumption_and_environmental_impacts). Last accessed 04 Sep 2022
12. Tenhunen, M., Penttinen, E.: Assessing the carbon footprint of paper vs. electronic invoicing. In: *ACIS 2010 Proceeding*
13. Schug, T.T., Birnbaum, L.S.: Human health effects of bisphenol A. *Mol. Integr. Toxicol.* 1–29 (2014). [https://doi.org/10.1007/978-1-4471-6500-2\\_1](https://doi.org/10.1007/978-1-4471-6500-2_1)
14. Pessi, B.: The impact of implementation of the electronic purchase invoice system on a company on the exmple of hahle group (2018)
15. Shaban, M., et al.: Billing system design based on internet environment. *Int. J. Adv. Comput. Sci. Appl.* **3**(9) (2012). <https://doi.org/10.14569/IJACSA.2012.030934>
16. Kim, J.I., et al.: Association of bisphenol A, bisphenol F, and bisphenol S with ADHD symptoms in children. *Environ. Int.* **161** (2022). <https://doi.org/10.1016/J.ENVINT.2022.107093>
17. Molangiri, A., et al.: Prenatal exposure to bisphenol S and bisphenol A differentially affects male reproductive system in the adult offspring. *Food Chem. Toxicol.* **167**, 113292 (2022). <https://doi.org/10.1016/J.FCT.2022.113292>
18. Pivonello, C., et al.: Bisphenol A: an emerging threat to female fertility. *Reprod. Biol. Endocrinol.* **18**, 1 (2020). <https://doi.org/10.1186/S12958-019-0558-8>
19. Erkekoğlu, P., et al.: Toxic effects of bisphenols: A special focus on bisphenol A and its regulations. *Bisphenols* (2022). <https://doi.org/10.5772/INTECHOPEN.102714>
20. Reale, E., et al.: Skin absorption of bisphenol A and Its alternatives in thermal paper. *Ann. Work Expo. Heal.* **65**(2), 206–218 (2021). <https://doi.org/10.1093/ANNWEH/WXAA095>
21. Marcos, A., et al.: Effects of bisphenol S on the life cycle of earthworms and its assessment in the context of climate change. *Sci. Total Environ.* **781** (2021). <https://doi.org/10.1016/J.SCI.TOTENV.2021.146689>

22. Liu, J., et al.: Occurrence, toxicity and ecological risk of Bisphenol A analogues in aquatic environment – a review. *Ecotoxicol. Environ. Saf.* **208** (2021). <https://doi.org/10.1016/J.ECOENV.2020.111481>



# Chest X-ray Image Classification Using Convolutional Neural Network to Identify Tuberculosis

Fahmida Nusrat Promy<sup>1</sup>, Tasnia Afrin Chowdhury<sup>1</sup>, Omar Tawhid Imam<sup>2</sup>, Farhana Alam<sup>1</sup>, Ahmed Wasif Reza<sup>1</sup>(✉), and Mohammad Shamsul Arefin<sup>3,4</sup>(✉)

<sup>1</sup> Department of Computer Science and Engineering, East West University, Dhaka 1212, Bangladesh

wasif@ewubd.edu

<sup>2</sup> Department of Electrical and Electronics Engineering, Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh

<sup>3</sup> Department of Computer Science and Engineering, Daffodil International University, Dhaka 1341, Bangladesh

sarefin@cuet.ac.bd

<sup>4</sup> Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh

**Abstract.** Tuberculosis (TB) is a severe bacterial infection that can be spread by inhaling small droplets from an infected person's cough and sneeze. TB claimed the lives of 1.5 million individuals in 2020, including 214,000 HIV-positive people. TB is the world's second most widespread infectious lethal disease and the 13th leading cause of mortality. As a result, predicting whether someone has tuberculosis or not is critical. We experimented with chest X-ray images of healthy and tuberculosis patients. For our studies, we applied the CNN models VGG16, VGG19, Xception, ResNet50, InceptionResNetV2, DenseNet201, InceptionV3, and MobileNetV2. We also developed two models utilizing convolutional layers, max-pooling, and other techniques. In our study, VGG-16, Xception, and DenseNet201 provide any model's highest training and validation accuracy. Densenet201 has the highest accuracy, with 99.7% in validation and 99.7% in training. One model we have developed has good training and validation accuracy, with 90.7% in training and 90% in validation.

**Keywords:** Tuberculosis · CNN · VGG · Xception · ResNet50 · InceptionResNetV2 · DenseNet201 · InceptionV3 · MobileNetV2 · Chest X-ray

## 1 Introduction

Tuberculosis (TB) is a bacterial infection distributed by breathing tiny droplets from a highly infectious person's sneezes and coughs. Although they primarily affect the lungs, they can also significantly impact the stomach (belly), glands, joints, and nervous system. Pulmonary tuberculosis, the most common form of the disease, damages the lungs, but it



often spreads only far after prolonged or extensive contact with an infected ill person. In the maximum of healthy people, the body's immune system can destroy the bacterium, and there are no symptoms of harboring it.

People can contract tuberculosis from each other through the air. One can get this infection by inhaling just a few of these bacteria. People who have compromised immune systems are more likely to get sick, including those with HIV, diabetes, malnutrition, or smoking habits.

A chest X-ray is a medical imaging exam that employs electromagnetic radiation to provide images of the chest and internal organs such as the heart, lungs, blood vessels, and bones. This process is noninvasive, rapid, and inexpensive, making them an effective tool for identifying various chest-related diseases. A chest X-ray image can aid in diagnosing various disorders, including lung ailments, heart difficulties, foreign objects, and some types of cancer. To identify TB from a chest X-ray, healthcare workers search for symptoms of lung damage, such as white spots called "miliary pattern" or nodules or cavities in the lung tissue. They also check for signs of disease, such as swollen lymph nodes or fluid accumulation in the pleural cavity.

However, human-aided diagnosis of tuberculosis from a chest X-ray can be flawed due to the possibility of human error. The person interpreting the X-ray may have less training or expertise or be fatigued, which can impair their ability to interpret the X-ray images accurately. In the case of diagnosis, healthcare practitioners can also be very slow. As a result, a quick, reliable, and accurate diagnostic system is required.

To classify the disease or healthy chest from x-ray pictures, we developed some self-built models and several recognized models such as VGG16, VGG19, Xception, ResNet50, InceptionResNetV2, DenseNet201, InceptionV3, and MobileNetV2. We used a large amount of data to conduct studies on 6133 images. We preprocessed the data before employing the specified models to ensure effective results and accuracy.

The following describes how the paper is organized. Section 1 is the introduction to this research paper, Sect. 2 discusses previous works related to Chest X-ray image classification of tuberculosis or any other disease, Sect. 3 discusses system architecture, dataset description, data preprocessing, model architecture, and algorithms, Sect. 4 presents experimental setup and performance evaluation of our used models and self-built models, and Sect. 5 concludes the paper.

## 2 Related Work

Leu et al. [1] examined and evaluated various CNN architecture and formation parameters. They then applied the transferred learning techniques to chest radiography images. They refined their pre-trained CNN model from an image data set to their medical radiographic image data set to detect tuberculosis.

Pasa et al. [2] used five convolutional blocks and an average grouping layer and a fully connected SoftMax layer consisting of two outputs. Training was carried out using categorical transverse entropy as an error function and with small batches of 4 samples.

Vajda et al. [3] used an algorithm to segment the lung based on an atlas. It was a set of CXRs of multiple patients and their expert delimited the lung boundaries. The system would first select some models that are more of the same type as the patient's own X-ray

by measuring similarities in the shape of the lung. Then, it distorted the preferred models on the patient X-ray using a registration algorithm.

Norval et al. [4], aimed at examining the success rate of detection of 2 different methods. The methods were image preprocessing and a hybrid approach. The simple architecture was chosen because it takes less time to train and has fewer parameters.

Hooda et al. [5], They had proposed a CNN-based model for the perfect detection of TB. Between the LeNet and Alexnet architectures, they used a clear and simple architecture with a number of layers. The simple architecture was chosen because it takes less time to train and has fewer parameters.

Ureta et al. [6], segmented the pictures of the lungs in the CXR images, they used U-Net. The network was made up of 5 convolutional blocks, with 2 convolutional layers in each block, and ReLU serving as the activation function.

Ahsan et al. [7], used the VGG16 model. Each convolutional layer of VGG16 has a layer that is ReLU. It performs by converting the whole range from 0 to any number using in the convolution layer.

CXR pixel classification provided by Hogeweg et al. [8] with the basis for the detection systems considered in this study. Using labeled samples from training images, a classifier was trained. A final decision was made by combining the system and the clavicle detection output at the pixel level and combined with the final shape decision.

Oloko-Oba et al. [9], proposed a Deep (CNN) based model for the detection and also for the classification of disease tuberculosis. The stage for feature extraction and the stage for feature classification made up their proposed CNN structure.

Jaeger et al. [10], for lung segmentation, they used 3 types of masks. They implemented a collection of well-known shape and texture descriptors to measure patterns in the detected lung field.

Karargyris et al. [11] detected lung fields and identified anatomical-type structures of interest such as the ribs and heart. Their method simplified the problem by first identifying the general anatomy of interest (lung fields), then locating anatomical structures that were spatially close together (such as the heart and ribs), and finally, they used classification techniques to find specific abnormalities within these regions of interest.

van Ginneken et al. [12] used the Active Shape Model (ASM) Segmentation method. To identify corresponding areas within the lung fields, segments were generated. Each area of each image was given a texture feature vector because of texture analysis. The next step was to calculate each region's abnormality using these feature vectors.

Shen et al. [13] used two image databases. Three sets of images were generated: 20 images for the cavity set and others for non-cavity and normal type. Using all images from the three image sets, the proposed hybrid technique was tested.

van't Hoog et al. [14], A radiologist trained three clinical officers for two weeks in evaluating CXRs for quality and presence of common abnormalities, as well as in determining which abnormal conditions required further care. 1143 CXRs were included in the analysis; 1031 of them were randomly chosen from participants who did not have TB and 112 of them belonged to people who had been diagnosed with prevalent bacteriologically confirmed TB. The works in [15–20] considered image processing-related tasks for different purposes.

### 3 System Architecture and Design

In the first step, we have declared the necessary libraries which are needed to perform a different type of operation or visualization on our chest X-ray images. Then comes one of the essential stages: data preprocessing. It provides some crucial steps for developing successful models. Obtaining the dataset, importing libraries, importing datasets, encoding, and partitioning the dataset into training and testing sets are among the steps involved. We performed data augmentation, which included image flipping, rotation, scaling, and cropping. Then we developed some models with convolution layers and max pooling. Predefined models such as VGG16, VGG19, Xception, ResNet50, InceptionResNetV2, DenseNet201, InceptionV3, and MobileNetV2 were also used. Finally, we analyzed our model in the final stage to see how well it predicts and provides accuracy.

#### 3.1 Dataset Description

The dataset, Chest X-ray Database, was obtained via Kaggle. This massive database of chest radiograph-type images for positive and standard cases were by researchers from DU from Bangladesh and Qatar University. They have collaborated with Malaysia and Hamad Medical Corporation medical practitioners. In their recent release, there are 7000 images of TB, from that 2800 TB images can be downloaded from the NIAID TB portal, and 3500 normal images. Some data are from Shenzhen and Montgomery datasets.

Figures 1 and 2 show some sample pictures of an affected and a normal chest x-ray respectively.



**Fig. 1.** Affected tuberculosis

#### 3.2 Data Pre-processing

In this step, we have prepared our data for further steps like doing data splitting into training sets and testing sets. We considered two sets of normal or healthy images for each training and testing set and another for TB detected.

Our dataset consists of 5247 images from two classes in the training set and 886 from two classes in the testing set. We attempted to collect a significant amount of data. Using the NumPy library, we verified that our dataset contains images with a height



**Fig. 2.** Healthy chest radiograph

and breadth of  $512 \times 512$  pixels on average. As a result, we chose the shape of  $520 \times 520$  pixels for our model to avoid a significant problem if the image size is larger than expected. For better results, we augmented the data. We flipped in a horizontal direction and rotated the image by 20 degrees. We have also used scaling and shifting features. We adjusted the image width by a maximum of 5% and the image height by a maximum of 5%. Additionally, we utilized the zoom capability, where we only zoomed in by 10% on the images.

Figure 3 shows the changes that resulted in the data after data preprocessing. The image quality is substantially better for extracting features.



**Fig. 3.** Augmented dataset

### 3.3 Model Architecture and Algorithms

We have tried to propose some self-built models which have similar architecture to some existing predefined models, and this shows a massive accuracy increase which is greater than existing papers which are about tuberculosis. We also applied predefined models such as VGG16, VGG19, Xception, ResNet50, InceptionResNetV2, DenseNet201, InceptionV3, and MobileNetV2, as well as two self-built models that used appropriate convolutional layers, max pooling, and other techniques.

**Self-built Model-1:** We used our preferred model in this case. This simple model features a convolution layer, max pooling, and ReLU activation. We initially described it as a

sequential model. Then we add a 32-filter convolution layer with size (3,3), input shape (520,520), and activation ReLU. Then we assigned the max pooling layer a pool size of (2,2) and repeated the method twice.

Afterward, we applied a dense layer with a dropout of 0.5. Dropouts help to avoid overfitting by randomly turning neurons off during training. In our scenario, we randomly shut off 50% of the neurons. The sigmoid activation function was applied in the last layer because it is binary class-based. Finally, we compile our model with a loss equal to binary cross entropy and Adam as the optimizer.

The model description is shown in Table 1.

**Table 1.** Self-built Model-1 with only convolution layer and max pooling

Layer	Output shape	Parameter
Conv2d	(None, 518, 518, 32)	896 values
Max_pooling2d (MaxPooling2D) layer	(None, 259, 259, 32)	0
Conv2d_1 layer	(None, 257, 257, 64)	18,496 values
Max_pooling2d_1 (MaxPooling2) layer	(None, 128, 128, 64)	0
Conv2d layer	(None, 126, 126, 64)	36,928 values
max_pooling2d_2 (MaxPooling2 layer	(None, 63, 63, 64)	0
Flatten layer	(None, 254,016)	0
Dense layer	(None, 128)	32,514,176 values
Activation layer	(None, 128)	0
Dropout layer	(None, 128)	0
Dense_1 layer	(None, 1)	129 values
Activation_1 layer	(None, 1)	0

So, the total parameters are 32,570,625 from which trainable parameters are also 32,570,625.

**Self-built Model-2:** We initially declared it as a sequential model. We provided a convolution layer with 64 filters,  $11 \times 11$  kernel size,  $520 \times 520$  input image size, and activation ReLU. After that, we applied batch normalization. Then we applied max pooling 2D with a pool size of  $3 \times 3$  with a stride of (2,2). The process is repeated three times, with the size of the convolution layer changing each time, from (7,7), then (5,5), and finally (3,3). Batch normalization is applied after each convolution layer. Then repeated the max pooling. The model architecture is shown in Table 2.

So, here we can see that the total parameter is 27,527,553 and the trainable parameter is 27,526,145. We have also used predefined models such as VGG16, VGG19, Xception, ResNet50, InceptionResNetV2, DenseNet201, InceptionV3, and MobileNetV2.

**Table 2.** Selft-built Model-2 architecture and parameter description

Layer	Output shape	Parameter
Conv2d	(None, 510, 510, 64)	23,296 values
Batch normalization	(None, 510, 510, 64)	256 values
Max_pooling2D	(None, 254, 254, 64)	0
Conv2d	(None, 248, 248, 128)	401,536 values
Batch normalization	(None, 248, 248, 128)	512 values
Max_pooling2D	(None, 123, 123, 128)	0
Conv2d	(None, 119, 119, 256)	819,456 values
Batch normalization	(None, 119, 119, 256)	1024 values
Max_pooling2D	(None, 59, 59, 256)	0
Conv2d	(None, 57, 57, 256)	590,080 values
Batch normalization	(None, 57, 57, 256)	1024 values
Max_pooling2D	(None, 28, 28, 256)	0
Flatten layer	(None, 200,704)	0
Dense layer	(None, 128)	25,690,240 values
Activation layer	(None, 128)	0
Dropout layer	(None, 128)	0
Dense_1 layer	(None, 1)	129 values
Activation_1 layer	(None, 1)	0

## 4 Experimental Results

### 4.1 Experimental Setup

The proposed system has been experimented on a personal computer that has Windows 10, Core i3 10th generation. It has 16GB RAM and Python 3.8 (version) is used to develop it.

### 4.2 Experimental Result

Table 3 shows the training and validation accuracies and losses for each model used in this experiment.

Table 3 shows that VGG-16, Xception, and DenseNet201 provide very significant accuracy in terms of both training and validation accuracy. Densenet201 has the highest accuracy of all of these models, with 99.7% in validation and 99.7% in training. Validation accuracy was greater than 90% for all predefined models. In terms of training and validation accuracy, the self-built model -1 performs well, with 90.7% training accuracy and 90% validation accuracy.

**Table 3.** Result analysis of models

Model	Training accuracy (%)	Validation accuracy (%)	Training loss	Validation loss
VGG16	99	99.6	0.03	0.034
VGG19	99	97.9	0.029	0.013
Xception	99.4	99.2	0.0974	0.28
InceptionV3	99.8	98.1	0.06	0.49
InceptionResNetV2	98.9	97.5	0.18	0.76
ResNet50	82.6	90.6	1.96	0.81
DenseNet201	99.7	99.5	0.017	0.07
MobileNetV-2	99.7	96.7	0.06	1.47
Self-build 1	90.7	90	0.26	0.25
Self-build 2	86.7	88	1.02	2.12

From Table 4, we can see the amount of correct and wrong predictions. DenseNet gives the most correct answer among the validation data it only gives only 1 incorrect answer. VGG-19 and VGG-16 are also close in terms of giving most of the correct numbers.

**Table 4.** Number of corrected images detected

Model	Correctional images of tuberculosis	Incorrect images of tuberculosis	Corrected images of normal chest	Incorrect images of the normal chest
VGG16	496	1	380	9
VGG19	493	4	389	0
Xception	497	0	357	32
InceptionV3	497	0	366	23
InceptionResNetV2	497	0	344	45
ResNet50	494	3	276	113
DenseNet201	496	1	389	0
MobileNetV-2	497	0	363	26
Self-built Model-1	475	22	308	81
Self-built Model-2	472	25	300	89

### 4.3 Performance Evaluation

The training and testing accuracy of each of the models with a precision score, recall score, and f1 score are shown in Table 5.

**Table 5.** Training, testing accuracy, recall, precision, F1-score

Model	Training accuracy (%)	Testing accuracy (%)	Precision value (%)	Recall value (%)	F1-score (%)
VGG16	99	99.6	99	99	99
VGG19	99	97.9	99	100	100
Xception	99.4	99.2	97	96	96
InceptionV3	99.8	98.1	98	97	97
InceptionResNetV2	98.9	97.5	96	94	95
ResNet50	82.6	90.6	90	85	86
DenseNet201	99.7	99.5	100	100	100
MobileNetV-2	99.7	96.7	98	97	97
Self-build Model-1	90.7	90	89	87	88
Self-build Model-2	86.7	88	86	87	87

Table 5 shows that DenseNet201 has the highest accuracy in both training and validation. It has a perfect precision, recall, and f1 scores. VGG-19 comes in second with a training accuracy of 99% and a testing accuracy of 97.9%. It also has a perfect precision, recall, and f1 ratings. VGG-19 testing is less accurate than DenseNet201. Xception has better than 99% training and validation accuracy. However, its precision and recall values are lower than Densenet or VGG16.

Inception ResNet2 has a training accuracy of 98.9% and a validation accuracy of 97.5%, which is relatively low. Resnet-50 has the lowest training accuracy of any of these models.

Model-1 outperforms Model-2 in terms of self-built models. In training, Model 1 has a precision of 90.7%, while Model 2 has a precision of 86.7%. Model-1 received 90% validation, whereas Model-2 received 88% validity.

In Figs. 4, 5, 6, 7, and 8, we can see the loss and accuracy figures of VGG-19, Xception, DenseNet201, Inception\_Resent, and MobileNetV2 respectively. On the y-axis, it represents the loss or accuracy value and on the x-axis, it represents the epoch number. In each epoch, for each step, it takes a maximum of 15 s for VGG-19, 12 s for Xception, 10 s for DenseNet201, 15 s for Inception\_Resent, and 5 s for MobileNetV2.

We can conclude that VGG-16, Xception, and DenseNet201 provide excellent training and validation accuracy. Densenet201 has the highest accuracy of all of these models, with 99.7% in validation and 99.7% in training. Validation accuracy was greater than 90% for all predefined models. Model 1 from self-built models has a high accuracy of 90.7% in training and 90% in validation. Resnet-50 has the lowest accuracy of any of the predefined models.



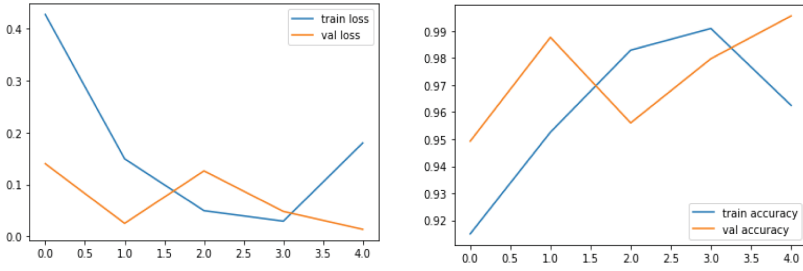


Fig. 4. Model loss and accuracy of VGG-19

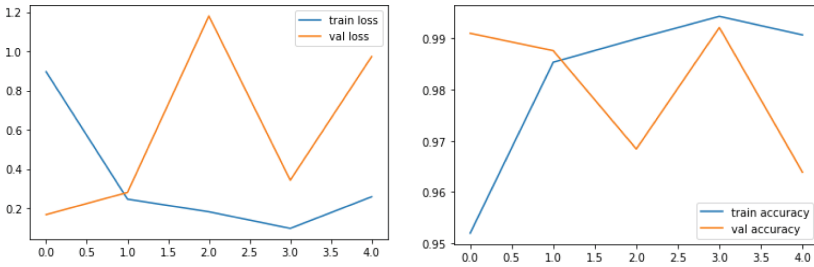


Fig. 5. Model accuracy and loss of Xception

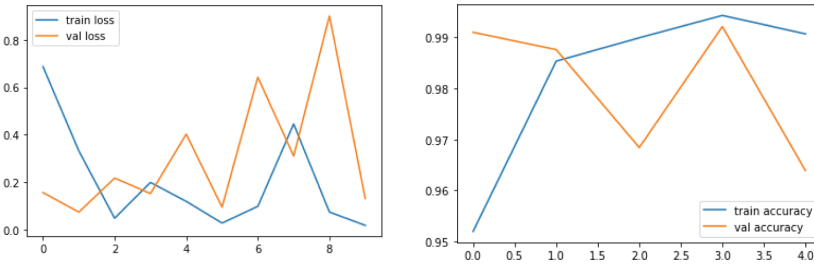


Fig. 6. Model accuracy and loss of DenseNet201

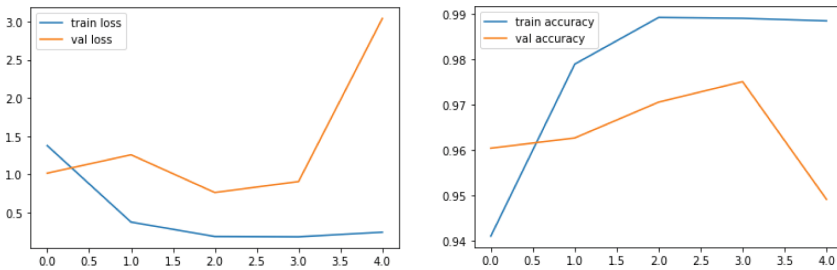
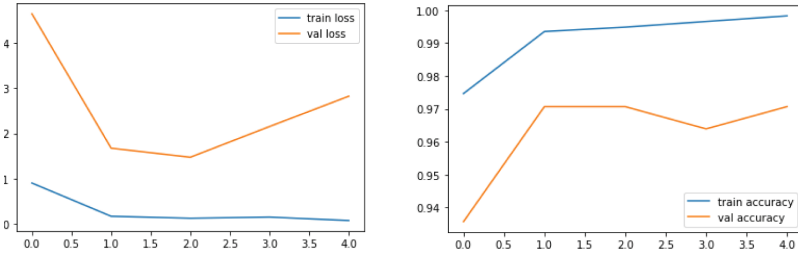


Fig. 7. Model loss and accuracy of Inception\_Resnet\_ver-2



**Fig. 8.** Model accuracy and loss of MobileNetV2

**Table 6.** Comparison with previous works

	Models or technique	Highest accuracy	Paper publication year
Our research paper	2 self-build models, VGG16, VGG19, Xception, ResNet50, InceptionResNetV2, DenseNet201, InceptionV3, and MobileNetV2	In self-build 90.7% Using predefined models 99.7%	
Paper [2]	Similar to AlexNet's self-proposed model	88% 95%	2019
Paper [4]	Preprocessing technique: Equalization of the histogram, enhancement contrast, the color channel sharpening	91.04% 92.54%	2020
Paper [21]	Densenet 201, Resnet 50, Efficientnet	92%	2023

So, from Table 6, we can see that our research works model's accuracy is larger than most of the previously worked research papers. We have tried to work with a large amount of data so that the result could be more accurate. We also built 2 new types of CNN models to propose new architectures for checking tuberculosis for the purpose of reducing computation power than existing models even though self-build models' accuracy is less than existing predefined models.

## 5 Conclusion

Tuberculosis is an infectious disease, and some infections do not produce symptoms. Thus, we attempted to develop an accurate model that could forecast the disease more precisely. Anyone who utilizes our approach can benefit since they can predict whether they have tuberculosis or not by using their chest X-ray. In the medical field, this technology has significant potential for diagnosing various diseases from X-Ray images. More trials with real-world datasets will allow the models to be trained and integrated into the healthcare industry to aid medical professionals.

Through our research, we have achieved a maximum of 99.7% accuracy in predicting tuberculosis and we have seen that it is higher than existing many research works. It is helpful for gaining accurate results. Medical professionals or normal people can use it to check their condition before any serious problems occur.

We also think this research can also help other researchers to think and work on building new types of CNN models for their own research instead of only using predefined models. As predefined models are proposed from doing huge research and those models are developed on millions of images and thousands of classes so, by building own or self-build models the training procedure takes very less time, and computational power and creates a huge influence on developing new models.




## References

1. Liu, C., et al.: TX-CNN: detecting tuberculosis in chest X-ray images using convolutional neural network. In: 2017 IEEE International Conference on Image Processing (ICIP), pp. 2314–2318 (2017). <https://doi.org/10.1109/ICIP.2017.8296695>
2. Pasa, F., Golkov, V., Pfeiffer, F., et al.: Efficient deep network architectures for fast chest X-Ray tuberculosis screening and visualization. *Sci Rep* **9**, 6268 (2019). <https://doi.org/10.1038/s41598-019-42557-4>
3. Vajda, S., et al.: Feature selection for automatic tuberculosis screening in frontal chest radiographs. *J. Med. Syst.* **42**(8), 1–11 (2018). <https://doi.org/10.1007/s10916-018-0991-9>
4. Norval, M., Wang, Z., Sun, Y.: Pulmonary tuberculosis detection using deep learning convolutional neural networks. In: Proceedings of the 3rd International Conference on Video and Image Processing, pp. 47–51. New York, NY, USA (2020). <https://doi.org/10.1145/3376067.3376068>
5. Hooda, R., Sofat, S., Kaur, S., Mittal, A., Meriaudeau, F.: Deep-learning: a potential method for tuberculosis detection using chest radiography. In: 2017 IEEE International Conference on Signal and Image Processing Applications (ICSIPA), pp. 497–502 (2017). <https://doi.org/10.1109/ICSIPA.2017.8120663>
6. Ureta, J., Shrestha, A.: Identifying drug-resistant tuberculosis from chest X-ray images using a simple convolutional neural network. *J. Phys. Conf. Ser.* **2071**(1), 012001 (2021). <https://doi.org/10.1088/1742-6596/2071/1/012001>
7. Ahsan, M., Gomes, R., Denton, A.: Application of a convolutional neural network using transfer learning for tuberculosis detection. In: 2019 IEEE International Conference on Electro Information Technology (EIT), pp. 427–433 (2019). <https://doi.org/10.1109/EIT.2019.8833768>
8. Hogeweg, L., Mol, C., de Jong, P.A., Dawson, R., Ayles, H., van Ginneken, B.: Fusion of local and global detection systems to detect tuberculosis in chest radiographs. In: Medical Image Computing and Computer-Assisted Intervention—MICCAI 2010, pp. 650–657. Berlin, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15711-0\\_81](https://doi.org/10.1007/978-3-642-15711-0_81)

9. Oloko-Oba, M., Viriri, S.: Diagnosing tuberculosis using deep convolutional neural network. In: Image and Signal Processing, pp. 151–161. Cham (2020). [https://doi.org/10.1007/978-3-030-51935-3\\_16](https://doi.org/10.1007/978-3-030-51935-3_16)
10. Jaeger, S., Karargyris, A., Antani, S., Thoma, G.: Detecting tuberculosis in radiographs using combined lung masks. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4978–4981 (2012). <https://doi.org/10.1109/EMBC.2012.6347110>
11. Karargyris, A., Antani, S., Thoma, G.: Segmenting anatomy in chest x-rays for tuberculosis screening. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 7779–7782 (2011). <https://doi.org/10.1109/IEMBS.2011.6091917>
12. van Ginneken, B., Katsuragawa, S., ter Haar Romeny, B.M., Doi, K., Viergever, M.A.: Automatic detection of abnormalities in chest radiographs using local texture analysis. In: IEEE Transactions on Medical Imaging, vol. 21, no. 2, pp. 139–149 (2002). <https://doi.org/10.1109/42.993132>
13. Shen, R., Cheng, I., Basu, A.: A hybrid knowledge-guided detection technique for screening of infectious pulmonary tuberculosis from chest radiographs. IEEE Trans. Biomed. Eng. **57**(11), 2646–2656 (2010). <https://doi.org/10.1109/TBME.2010.2057509>
14. van't Hoog, A.H., et al.: High sensitivity of chest radiograph reading by clinical officers in a tuberculosis prevalence survey. Int. J. Tuberculosis Lung Dis. **15**(10), 1308–1314 (2011). <https://doi.org/10.5588/ijtld.11.0004>
15. Curvo-Semedo, L., Teixeira, L., Caseiro-Alves, F.: Tuberculosis of the chest. Eur. J. Radiol. **55**(2), 158–172 (2005). <https://doi.org/10.1016/j.ejrad.2005.04.014>
16. Saha, R., Debi T., Arefin, M.S.: Developing a framework for vehicle detection, tracking and classification in traffic video surveillance. In: Vasant, P., Zelinka, I., Weber, G.W. (eds.) Intelligent Computing and Optimization. ICO 2020. Advances in Intelligent Systems and Computing, vol. 1324. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-68154-8\\_31](https://doi.org/10.1007/978-3-030-68154-8_31)
17. Fatema, K., Ahmed, M.R., Arefin, M.S.: Developing a system for automatic detection of books. In: Chen, J.I.Z., Tavares, J.M.R.S., Iliyasu, A.M., Du, K.L. (eds.) Second International Conference on Image Processing and Capsule Networks. ICIPCN 2021. Lecture Notes in Networks and Systems, vol. 300. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-84760-9\\_27](https://doi.org/10.1007/978-3-030-84760-9_27)
18. Rahman, M., Laskar, M., Asif, S., Imam, O.T., Reza, A.W., Arefin, M.S.: Flower Recognition Using VGG16. In: Chen, J.I.Z., Tavares, J.M.R.S., Shi, F. (eds.) Third International Conference on Image Processing and Capsule Networks. ICIPCN 2022. Lecture Notes in Networks and Systems, vol. 514. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-12413-6\\_59](https://doi.org/10.1007/978-3-031-12413-6_59)
19. Yeasmin, S., Afrin, N., Saif, K., Imam, O.T., Reza, A.W., Arefin, M.S.: Image classification for identifying social gathering types. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_10](https://doi.org/10.1007/978-3-031-19958-5_10)
20. Ahmed, F., et al.: Developing a classification CNN model to classify different types of fish. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_50](https://doi.org/10.1007/978-3-031-19958-5_50)
21. Park, M., Lee, Y., Kim, S.: Distinguishing nontuberculous mycobacterial lung disease and *Mycobacterium tuberculosis* lung disease on X-ray images using deep transfer learning (2023). <https://doi.org/10.1186/s12879-023-07996-5>



# Digital Wireless Mini-transduce of Plant Thermoregulation

A. Grishin<sup>1</sup> , A. Grishin<sup>1</sup>  , N. Semenova<sup>1</sup> , V. Grishin<sup>1</sup> ,  
and V. Panchenko<sup>1,2</sup> 

<sup>1</sup> Federal Scientific Agroengineering Center VIM, 1st Institutsky Passage 5, 109428 Moscow, Russia

5145412@mail.ru

<sup>2</sup> Russian University of Transport, Obraztsova St. 9, 127994 Moscow, Russia

**Abstract.** Digital technologies used in agriculture are based on data obtained from sensors and transducers providing primary information on the state of plants. In this research we developed and practically tested a leaf mini-transducer for digital assessment of thermoregulation processes that characterize the factors of plant productivity. The developed transducer has scientific and practical value; it allows obtaining data on thermoregulation processes individually for each plant. The tests carried out proved the existence of cooperative connection between the order parameter characteristic of the self-organization process, i.e. heating of leaves under the influence of thermal exergy, and the control parameter, i.e. cooling thermoregulation. We established that the process of thermoregulation is divided into phases, according to which we can determine the optimal temperature of photosynthesis for a particular plant in a particular phase of growth. The data obtained from the developed transducer can be used both for regulating climatic parameters in vertical farming to increase plant productivity and for collecting big data for scientific research and in-depth analysis.

**Keywords:** Digital technologies · Leaf transducer · Plant productivity · Thermoregulation · Closed artificial agroecosystems

## 1 Introduction

Agriculture cannot develop without the use of modern technologies. Modern multistorey vertical agricultural complexes already widespread in Japan, South Korea, and Australia are now built around the largest Russian cities, such as Moscow (RusEco), Petropavlovk-Kamchatsky (Eco-Vitamin), and Novosibirsk (iFarm) [1]. In such complexes, which are, in fact, modern closed agroecosystems, the latest technologies are used to optimize the production process, to increase number of products manufactured per unit area, and to reduce transportation costs. Sensors and transducers are always used on intelligent farms of an intensive type. They provide initial information on changes in the state of plants, including on deficiency states, as well as on environmental indicators, in which they are grown [2, 3]. Such information makes it possible to control the technological process of agricultural production by building an algorithm and making decisions in intellectualized control systems [4].

Until recently agriculturists working on open ground have been using non-invasive method for determining the local water absorption by roots using a soil moisture meter that can determine the local moisture content in the soil [5]. Thus, they can control the growth conditions and timely implement reclamation measures. Methods based on monitoring sap flow, which make it possible to estimate rather accurately the transpiration of plants both in open ground [6, 7] and in greenhouse conditions [8], can also be attributed to new, increasingly popular methods for determining water stress. To accurately calculate the moisture content at the biointerface we need a flexible and stable transducer. Moisture sensors using laser direct lithography technology allow eliminating complex and expensive conventional electrode preparation procedure and can be directly attached to plant leaves for long-term real-time tracking of transpiration [9]. These transducers record deficient plant-growing conditions, allowing monitoring plantings for making further decisions on adjusting agricultural techniques. However, such transducers do not allow maintaining plant productivity at an optimal level without extra costs to eliminate the consequences of stress, since they don't take into account the synergetic principle of energy conversion in plants.

An integrated multimodal flexible sensor system for plant growth management using stacked ZnIn4S4 (ZIS) nanoleaves as a sensing medium not only facilitates the creation of biointerfaces between plants and machines for precise control of their health status but also saves resources. A flexible transducer based on ZIS not only perceives light with a fast response but also controls humidity with a stable performance [10]. The most modern achievements in the field of plant stress monitoring include combined devices recording leaf reflection spectrum and chlorophyll fluorescence induction using lasers with wavelengths in the range of 405–470 nm [11]. Devices of this type allow detecting stress conditions at the initial stage of their emergence, including stresses caused not only by drought and temperature changes, but also by lighting conditions, the assessment of which was previously rather laborious [12]. However, since these systems have been developed recently, their cost is relatively high.

Timely monitoring of the plant condition is necessary to understand its physiological status and can help prevent negative effects of environmental stress and increase agricultural production. In connection with the intensification and digitalization of agricultural production, this task is relevant both for open ground and for protected ground, especially when using phase-wise regulation of the intensity and spectral composition of radiation.

The main indicators characterizing the physiological state of plants and associated with their productivity include transpiration [13, 14], the most important function of which is thermoregulation. To protect themselves from overheating, plants regulate their temperature by evaporation. With intensive transpiration, the leaf temperature can be several degrees below the temperature of the leaf, in which transpiration does not occur (wilting). The operation of 'the upper end engine' is also conditioned by leaf transpiration, and it is supported not only by metabolic energy, but also by the energy of the external environment [15].

The plant spends approximately 1% of consumed water for photosynthesis; approximately 5% of consumed water it spends for the transfer of nutrients to various organs, and the rest of consumed water it spends for thermoregulation [16]. Thermoregulation in

plants is of key importance in maintaining photosynthesis at its maximum temperature level.

The main function of thermoregulation is to reduce the temperature of the photosynthesizing leaf ( $t_l$ ), which captures the thermal energy of light radiation (ET), and to maintain the optimum temperature of photosynthesis ( $t_o$ ), by lowering the leaf temperature by  $\Delta T = t_l - t_o$ . Since the plant is a self-organizing structure, the main processes occurring in it are subject to the laws of synergetics and to the principle of extreme energetic self-organization according to the law of survival [17, 18]. The nonlinear dynamic process of transpiration occurs due to the external energy of the environment (thermal exergy) [19]. Thermoregulation occurs due to evaporative cooling in the stomatal apparatus of the plant and has a limited effect due to the limiting opening of the stomata themselves. In this regard, the process of thermoregulation can be represented in three phases:

- (1) The ambient temperature is below optimal; the ambient temperature and the temperature of the plant are equal, since the plant does not have internal sources that can increase its temperature.
- (2) The ambient temperature rises and becomes higher than optimal temperature, but the plant is able to cool itself and reduce its own temperature to optimal indicators due to the cooling evaporation of moisture from the leaf surface. The difference between the ambient temperature and the temperature of the plant has a limited growing value.
- (3) The ambient temperature is higher than optimal temperature, but the plant is no longer able to cool itself to optimal values. That is, the temperature is so high that the plant depletes all its cooling capacity and the plant temperature begins to rise following the ambient temperature.

In industrial plantings, to determine the parameters of plant transpiration, local methods for determining the amount of evaporated moisture are used in individual areas along the entire perimeter of the site with subsequent extrapolation. The weighting method has hitherto been considered the main method for determining the intensity of transpiration of herbaceous plants [20]; for woody plants, the water flow rate is determined by the linear velocity of water flow and the volume of water passing through a unit of water-conducting xylem [21]. These methods are rather laborious and do not allow for non-destructive monitoring of the transpiration intensity in vegetable plants directly in the place of growth. Measurements of turgor and the parameters of variable fluorescence of the plant leaves are much more informative for monitoring the stress status of crops [11, 22].

The purpose of the researches was to develop and test experimentally the transducer in the course of digital assessment of thermoregulation processes and to confirm experimentally the phasing of the thermoregulation process.

## 2 Materials and Methods

The leaf temperature depends on the physiological state of the plant, its species origin and environmental factors such as lighting, air temperature, and its circulation [23]. These temperatures were the subject of our researches.

The researches of the thermoregulation process were conducted on the early maturing tomato variety 'Boets' ('Buyan') bred in Siberia. The plant was set in a sealed container with a nutrient solution, while the root system was placed in the container with the leaves outside it (Fig. 1). This provided thermoregulation control without an error introduced by evaporation of another type. The plant itself was placed into a plexiglass container with LED illumination provided by a 20 W bispectral lamp, consisting of red (660 nm) and blue (460 nm) LEDs in the R/B ratio 10/2.



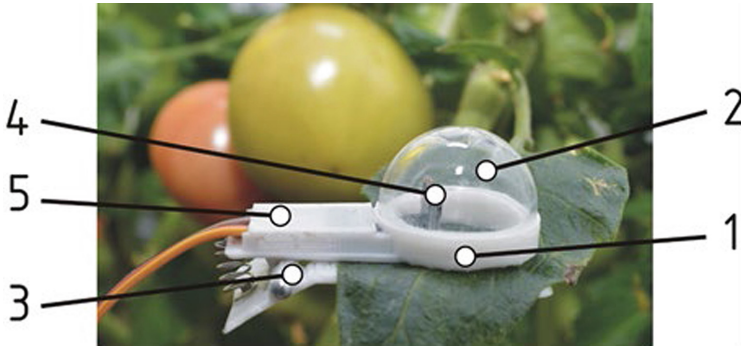
**Fig. 1.** The 'Boets' tomato variety plant in a sealed container with nutrient solution

To assess the thermoregulation process in leaves and control plant productivity by changing the control variable, we constructed a leaf mini-transducer. The transducer body was made of polylactide polymer, which is a biodegradable material obtained from renewable resources such as maize or sugar cane. The material is highly biocompatible. It is also used in medicine for the production of surgical sutures and pins. The body was equipped with a special pin for attaching the transducer to the plant leaf. The force of the pin preserved the leaf surface tissues. The upper part of the transducer was equipped with a spherical photoconductive dome. The dome was made of thin-walled glass, which provided sufficient light transmission to ensure the photosynthesis process inside the glass; the transducer was also light—with a wire it weighed 8.5 g; the dome diameter was 18 mm (Fig. 2). Such a design eliminated previously unresolved environmental problems such as dust, steam, etc. [23].

The developed device was equipped with two built-in digital sensors: air temperature sensor and infrared non-contact temperature sensor.

The air temperature sensing element was based on the Si7021-A20 chip manufactured by Silicon Labs. The Si7021 I2C humidity transducer is a monolithic CMOS integrated circuit that includes temperature transducer integrators, analog-to-digital converter, signal processor, calibration data, and I2C interface. The temperature transducers are factory calibrated; the calibration data are stored in the built-in non-volatile memory. This ensures that the transducers are completely interchangeable without the need for recalibration or software changes. The sensor is available in a  $3 \times 3$  mm DFN body. The sensor measures temperature from  $-10$  to  $+85$  C with an accuracy of  $\pm 0.4$  C.





**Fig. 2.** The leaf transducer structure: 1—transducer body; 2—photoconductive dome; 3—clamping element; 4—leaf temperature sensor; 5—air temperature sensor

The infrared sensor for measuring the surface temperature of the leaf runs on the MLX90614 chip with a 17-bit digital output signal using the I2C protocol. The sensor allows measuring the leaf surface temperature in the range from  $-40$  to  $+125$  C with an accuracy of  $0.2$  C.

The output signal of the transducer is digital. It is transmitted via the I2C protocol, which is compatible with all modern digital controllers, data transmitters, and other IoT equipment. This allows a large number of transducers to be integrated into the IoT system.

The leaf transducer can be connected to a processor (such as MCU Arduino, widespread in modern agriculture [24–26]), which processes input data for the chamber temperature and the surface temperature of the leaf. The data is then transmitted to the data logger via a serial port or radio transmitter. The radio transmitter operates via the LoRa protocol, which provides low-speed, long-range data transmission at a frequency of 433 MHz, at a rate of up to 32 kbit/s [27, 28]. This speed is sufficient for transferring low-volume data from the transducer at a given time interval.

The transducer was powered via a USB port from an external battery (powerbank) with a capacity of 10 A·h. Provided that the enhanced signal transmission is disabled, the consumption during data transmission is 4 mA.

The data obtained from the transducer were used to study the issues of plant thermoregulation from the bioenergetical standpoint, taking into account the thermoregulation process phases [29].

The endothermic process of evaporation, in which the heat of a stationary non-equilibrium phase transition “liquid-vapor” is absorbed (at constant temperature), is the key in the process of thermoregulatory cooling. The exergy of light radiation ( $E_T$ ) is a thermoregulation order parameter and a source of external energy, whereas the flow rate of thermoregulation ( $q_T$ ) is a control variable. The relationship between the order parameter and the control variable is expressed by the equation:

$$(t_l - t_o)M_l C_l = E_T \subset q_T r_w \quad (1)$$

where  $M_l$ —leaf mass;

$C_1$ —specific heat capacity of leaf mass;

$r_w$ —specific heat of phase transition (evaporation).

The characteristics of the processes were measured automatically several times using a recorder in the 2-week period. The measured data were archived on a SD card and then processed on a computer in Microsoft Excel (10).

The key component of the recorder was a programmable controller powered by an ATMega processor equipped with analog and digital inputs and outputs, real-time clock, device for storing on a micro-SD memory card for fixing data, and WiFi module for transferring them to cloud storage or an Internet database.

From the received data, we formed lines, which were separated by commas. Each line was written separately into a CSV file on a memory card and sent to the MySQL\_Connection library.

### 3 Results and Discussion

The developed leaf transducer has a number of advantages, including the use of modern digital communication protocols. The I2C/TWI data transfer protocol allows the transducer to be used with all modern IoT devices while the energy-efficient radio transmitter LoRa provides wireless connection of the transducer to system for controlling light-climatic parameters at distances of up to 200 m (without the use of repeaters). The cost of the transducer is approximately \$13 (2022). This makes it possible to massively use transducers of this type in closed artificial chambers and vertical farms, while receiving a large data array, which can be used for further analysis and development of control solutions for the automation system. If necessary, it allows for individual control of plant parameters.

However, the operating experience showed that the reliability of the transducer directly depends on the build quality of the sensor strapping boards. When using inexpensive components, the share of rejected sensors was approximately 5%. Service life of another 5% of sensors was no more than 30 days. Using the transducer on a plant for more than 7 days without changing its position led to an increase in the reading error due to possible damage to the leaf or adaptation of plant cells to a foreign structure. In particular, this led to disruption of the stomata system and, consequently, thermoregulatory properties of the leaf.

Data were registered with a 5-min interval; for each measured parameter we recorded a total of 4.896 values. The data obtained were combined into groups according to the heating stages from + 28 to + 34 C and the average values were determined (Table 1).

The obtained data were presented in the form of time series graphs of measured temperatures (Fig. 3). The graph distinguishes an area with a constant temperature of 29.2 C. The difference between the values did not exceed 1%, which is acceptable. Thus, there are 3 phases of thermoregulation:

- (1) Phase 1—values 1–15;
- (2) Phase 2—values 15–25;
- (3) Phase 3—values 25–39.

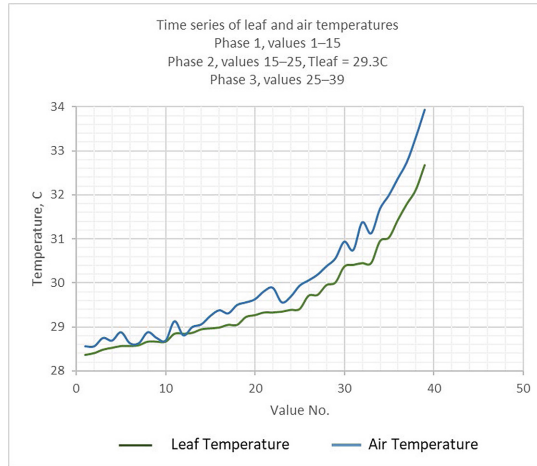
**Table 1.** Average values of plant leaf temperature (t) and air temperature measurements.

Value No	Air t, °C	Value No	Leaf t, °C	Value No	Air t, °C	Value No	Leaf t, °C
1	28.56	1	28.37	21	29.81	21	29.33
2	28.56	2	28.41	22	29.44	22	29.33
3	28.75	3	28.49	23	29.56	23	29.35
4	28.69	4	28.53	24	29.69	24	29.39
5	28.88	5	28.57	25	29.94	25	29.41
6	28.63	6	28.57	26	30.06	26	29.71
7	28.63	7	28.59	27	30.19	27	29.73
8	28.88	8	28.67	28	30.38	28	29.95
9	28.75	9	28.67	29	30.56	29	30.01
10	28.69	10	28.67	30	30.94	30	30.37
11	29.13	11	28.85	31	30.75	31	30.41
12	28.81	12	28.85	32	31.38	32	30.45
13	29.00	13	28.87	33	31.13	33	30.45
14	29.06	14	28.95	34	31.69	34	30.95
15	29.25	15	28.97	35	32.00	35	31.03
16	29.38	16	28.99	36	32.38	36	31.43
17	29.31	17	29.05	37	32.75	37	31.79
18	28.94	18	29.05	38	33.31	38	32.11
19	29.50	19	29.23	39	33.94	39	32.67
20	29.63	20	29.27	–	–	–	–

In Phase 1, an increase in plant leaf temperature was observed following an increase in air temperature. In Phase 2, the plant leaf temperature stabilized at an air temperature of 29.3 C. In Phase 3, the plant leaf temperature continued to grow after an increase in air temperature while maintaining a constant maximum temperature difference.

On the graphs, the leaf temperature in phase 2 has random deviations from a constant value, which is caused by objective reasons for plant growth and this randomness must be confirmed. For this, statistical processing of the measurement results was carried out by the method of “Successive differences” (Table 2), which made it possible to check the time series of the leaf temperature for random deviations of its values from the mathematical expectation, for which the factual  $\tau_f$  and theoretical  $\tau_t$  criteria were determined. If  $\tau_f > \tau_t$ , then the deviations are random and the temperature can be considered constant and equal to the mathematical expectation.

Experimental studies revealed three phases of plant thermoregulation. Further research should be aimed at theoretical and experimental confirmation of the hypothesis that the maximum productivity of the plant is provided at the optimum air temperature and maximum consumption of nutrient solution. If this hypothesis is confirmed, it is



**Fig. 3.** The results of measurements of the leaf and air temperature sequences by the leaf transducer in the controlled period.

**Table 2.** Statistical indicators by the ‘Successive differences’ method.

Statistical indicators	Value
Expected value	29.22
Standard deviation	0.17
Dispersion	0.03
The lower interval limit, °C	28.54
The upper interval limit, °C	29.89
$\tau_f$	0.58
$\tau_t$	0.26

necessary to develop a universal algorithm for controlling the air temperature in closed agroecosystems, which ensures the search for the temperature at the boundary of phases 2 and 3. This algorithm corrects the ambient temperature and controls the temperature difference between the leaf and the ambient air, determining their maximum difference ( $\Delta T_{max}$ ) at the lowest possible air temperature ( $T_{min}$ ) that provides a rational maximum consumption of the nutrient solution by the plant ( $R_{max}$ ):

$$\left\{ \begin{array}{l} T_{min} \\ \wedge \\ \Delta T_{max} \end{array} \right. \Rightarrow R_{max} \tag{2}$$

The use of the developed leaf mini-transducer, working in conjunction with a control system based on this algorithm, will provide intelligent control of the temperature in

closed artificial agroecosystems ensuring maximum plant productivity regardless of their type and growth phase.

## 4 Conclusion

As part of the recorder the digital mini-transducer of plant thermoregulation showed satisfactory functioning; its indicators corresponded to the indicators of the sensors used in it.

We studied the process of plant thermoregulation with an increase in air temperature and experimentally confirmed the presence of three phases. We noted that phase 2 corresponds to the constant leaf temperature, that is, to the temperature optimum equal to 29.3 C for tomato. Confirmation of the presence of thermoregulation phases allows developing an intelligent algorithm for controlling the climatic parameters of plant growth to ensure maximum productivity and economical use of resources.

The digital mini-transducer for thermoregulation of plants as part of the system for managing growing conditions can be used to control and maintain optimal productivity of plants in open and protected ground, which is especially important under changing lighting conditions.

## References

1. Hochberg, L.M. (ed.): Forecast of the scientific and technological development of the Russian Federation until 2030. Ministry of Education and Science of the Russian Federation. National Research University Higher School of Economics, p. 244 (2014)
2. Andreev, Y., Makeeva, T., Pukhova, E., Sevryugin, V., Sherstnev, G.: Technical Means of Digital Information Processing Systems Moscow. Ivan Fedorov Moscow State University of Printing Arts, p 74 (2015)
3. Anusha, K., Mahadevaswamy, U.B.: Automatic IoT based plant monitoring and watering system using Raspberry Pi. *Int. J. Eng. Manuf.* **8**(6), 55–67 (2018). <https://doi.org/10.5815/ijem.2018.06.05>
4. Bhavikatti, S., Sadanand, P., Patil, M., Vibhuti, P., Mudengudi, S.S.: Automated roof top plant growth monitoring system in urban areas. *Int. J. Eng. Manuf.* **9**(6), 14–23 (2019). <https://doi.org/10.5815/ijem.2019.06.02>
5. Van Dusschoten, D., Kochs, J., Kuppe, C., Sydoruk, V.A., Couvreur, V., Pflugfelder, D., Postma, A.J.: Spatially resolved root water uptake determination using a precise soil water sensor. *Plant Physiol.* **184**(3), 1221–1235 (2020). <https://doi.org/10.1104/pp.20.00488>
6. Venturin, A.Z., Guimarães, C.M., de Sousa, E.F., Machado Filho, J.A., Rodrigues, W.P., de Serrazine, Í., Araujo Serrazine, R., Bressan-Smith, C.R., Marciano, E.C.: Using a crop water stress index based on a sap flow method to estimate water status in conilon coffee plants. *Agric. Water Manage.* **241**, 106343 (2020). <https://doi.org/10.1016/j.agwat.2020.106343>
7. Barekova, A., Barek, V., Kovacova, M., Novotna, B., Kiss, V.: Climate conditions impact on the sap flow into plants and their dendrometric changes. *J. Ecol. Eng.* **21**(6), 224–228 (2020). <https://doi.org/10.12911/22998993/124077>
8. Nackley, L.L., de Sousa, E.F., Pitton, B.J.L., Sisneroz, J., Oki, L.R.: Developing a water-stress index for potted poinsettia production. *HortScience* **55**(8), 1295–1302 (2020). <https://doi.org/10.21273/HORTSCI14914-20>

9. Lan, L., Le, X., Dong, H., Xie, J., Ying, Y., Ping, J.: One-step and large-scale fabrication of flexible and wearable humidity sensor based on laser-induced graphene for real-time tracking of plant transpiration at bio-interface. *Biosensors Bioelectron.* **165**, 112360 (2020). <https://doi.org/10.1016/j.bios.2020.112360>
10. Lu, Y., Xu, K., Zhang, L., Deguchi, M., Shishido, H., Arie, T., Pan, R., Hayashi, A., Shen, L., Akita, S., Takei, K.: Multimodal plant healthcare flexible sensor system. *ACS Nano* **14**(9), 10966–10975 (2020). <https://doi.org/10.1021/acsnano.0c03757>
11. Smirnov, A., Proshkin, Y., Sokolov, A., Dorokhov, A.: Portable spectral device for monitoring plant stress conditions. *E3S Web Conf.* **210**, 05016 (2020). <https://doi.org/10.1051/e3sconf/202021005016>
12. Dorokhov, A.S., Smirnov, A.A., Semenova, N.A., Akimova, S.V., Kachan, S.A., Chilingaryan, N.O., Glinushkin, A.P., Yu Podkovyrov, I.: The effect of far-red light on the productivity and photosynthetic activity of tomato. *IOP Conf. Ser. Earth Environ. Sci.* **663**, 012044 (2021). <https://doi.org/10.1088/1755-1315/663/1/012044>
13. Graamans, L., Dobbelsteen, A., Meinen, E.: Plant factories; crop transpiration and energy balance. *Agric. Syst.* **53**, 138–147 (2017). <https://doi.org/10.1016/j.agsy.2017.01.003>
14. Li, L., Shiwang, C., Chengfei, Y., Fanjia, M., Sigrimis, N.: Prediction of plant transpiration from environmental parameters and relative leaf area index using the random forest regression algorithm. *J. Clean. Prod.* **261**, 121136 (2020)
15. Erzhapova, R.S.: Plant physiology. Water regime of plants. Grozny ChSU, **128** (2015)
16. Mudrik, V.A., Sventitsky, I.I.: Bioenergetic Aspects of Assessing the Moisture Supply of Plants. Pushchino, Pushchino Center for Biological Research of the Academy of Sciences of the USSR, p. 23 (1981)
17. Sventitsky, I.I.: Energy saving in the agro-industrial complex and energy extremity of self-organization. All-Russian Research Institute for Electrification of Agriculture, Moscow, p. 466 (2007)
18. Sventitsky, I.I., Grishin, A.P.: Definition of the term “energy-informational”. *Bulletin of All-Russian Research Institute for Electrification of Agriculture. Energy Elect. Technol. Agric.* **1**(4), 79–82 (2009)
19. Grishin, A.P., Grishin, A.A., Grishin, V.A.: The influence pattern of the transpiration process on plant productivity. *IOP Conf. Ser. Earth Environ. Sci.* **274**, 012126 (2019). <https://doi.org/10.1088/1755-1315/274/1/012126>
20. Voronin, P., Fedoseeva, G.P.: Stomatal control of photosynthesis in detached leaves of woody and herbaceous plants. *Plant Physiol.* **59**(2), 309–315 (2012)
21. Karaseva, V.N., Karaseva, M.A., Mukhortov, D.I.: Diagnostics of the physiological state of coniferous trees by bioelectric and temperature indicators. *Lesovedenie* **2**, 162–174 (2020). <https://doi.org/10.31857/S0024114820010088>
22. Domansky, V.P., Kozel, N.V.: Estimation of the magnitude of water deficit in a vegetative plant by the stiffness of the leaf blade and chlorophyll fluorescence variable. *News Belarusian Natl. Acad. Sci. Ser. Biol. Sci.* **3**, 50–56 (2011)
23. Yu, L., Wang, W., Zhang, X., Zheng, W.: A review on leaf temperature sensor: measurement methods and application. In: *International Conference on Computer and Computing Technologies in Agriculture, CCTA 2015: Computer and Computing Technologies in Agriculture IX*, vol. 478. IFIP Advances in Information and Communication Technology, pp. 216–230 (2016). [https://doi.org/10.1007/978-3-319-48357-3\\_21](https://doi.org/10.1007/978-3-319-48357-3_21)
24. Cruz, L.A.A., Griño, M.T.T., Tungol, T.M.V., Bautista, J.T.: Development of a low-cost air quality data acquisition IoT-based system using Arduino Leonardo. *Int. J. Eng. Manuf.* **9**(3), 1–18 (2019). <https://doi.org/10.5815/ijem.2019.03.01>
25. Akwu, S., Bature, U.I., Jahun, K.I., Baba, M.A., Nasir, A.Y.: Automatic plant irrigation control system using Arduino and GSM module. *Int. J. Eng. Manuf.* **10**(3), 12–26 (2020). <https://doi.org/10.5815/ijem.2020.03.02>

26. Saha, T., Jewel, M.K.H., Mostakim, M.N., Bhuiyan, N.H., Ali, M.S., Rahman, M.K., Ghosh, H.K., Khalid Hossain, M.: Construction and development of an automated greenhouse system using Arduino Uno. *Int. J. Inform. Eng. Electron. Bus.* **9**(3), 1–8 (2017). <https://doi.org/10.5815/ijieeb.2017.03.01>
27. García, S., Larios, D.F., Barbancho, J., et al.: Heterogeneous LORA-based wireless multimedia sensor network multiprocessor platform for environmental monitoring. *Sensors* **19**, 16 (2019). <https://doi.org/10.3390/s19163446>
28. Arlin, M.R.E., Niswar, M., Adnan, A., Fall, D., Kashihara, S.: LouPe: LoRa performance measurement tool. In: 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT), pp. 168–171 (2018). <https://doi.org/10.1109/EIConCIT.2018.8878525>
29. Grishin, A.P., Grishin, A.A., Grishin, V.A.: The uniformity of a random process structure of plants transpiration. *IOP Conf. Ser. Earth Environ. Sci.* **274**, 012127 (2019). <https://doi.org/10.1088/1755-1315/274/1/012127>

# **Convolution Neural Network, Deep Learning, and Machine Learning**





# Effective Fault Prediction Techniques for the Green Cloud Computing Environment Applying Machine Learning to Enhance Network Management

Hasnath Ahmed Tamim<sup>1</sup>, Md. Sagor Hossain<sup>1</sup>, Md. Asif Uzzaman Asif<sup>1</sup>,  
Borhan Uddin<sup>1</sup>, Ahmed Wasif Reza<sup>1</sup>(✉), and Mohammad Shamsul Arefin<sup>2</sup>(✉)

<sup>1</sup> Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh  
wasif@ewubd.edu

<sup>2</sup> Department of Computer Science and Engineering, Chittagong University of Engineering and  
Technology, Chattogram, Bangladesh  
sarefin@cuet.ac.bd

**Abstract.** Network errors are a significant issue in cloud computing. Occasionally, issues with cloud networks may cause productivity to drop considerably. The primary objective of this research is to employ machine learning techniques for predicting failures in Cloud Networks, enabling prompt identification of issues. The author attempted to determine the most accurate method, the Random Forest, that matches our model. We trained our model using gradient boosting, decision trees, K-neighbors, Random forests, SVC, and logistic regression. Then, by analyzing data as performance indicators, resource use, and energy consumption, random forests may be employed in a green cloud computing environment to forecast upcoming flaws. The system can recognize patterns and correlations in data that may indicate an issue and generate predictions for future forecasting based on these patterns, with an accuracy of 80.825479%.

**Keywords:** Cloud computing · Fault prediction · Data preprocessing · Random forest

## 1 Introduction

The issue of network errors poses a significant challenge for cloud computing. The objective of cloud computing is to provide a variety of computer services, including servers, storage, databases, networking, software, analytics, and intelligence, via the internet, with the aim of delivering rapid innovation, adaptable resources, and economies of scale. Pay-as-you-go pricing models allow users to pay only for the cloud services they use, which helps to reduce operational expenses, manage infrastructure more efficiently, and scale as the organization's needs evolve.

Cloud computing has three primary categories of services: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [1]. Each of these

services offers different levels of access and control over the underlying infrastructure. However, cloud network issues can sometimes have a negative impact on productivity, such as when using autopilot while driving or trading stocks. Despite this, cloud computing is a valuable technology for businesses of all sizes and offers many advantages [2]. By using cloud computing services, companies can reduce the risk of data loss and downtime, as advanced data backup and recovery solutions are often provided to simplify the process of safeguarding and retrieving data in case of an emergency. These are just a few of the many benefits of cloud computing.

This paper's main goal is to predict Cloud Network failures using machine learning so that we can spot issues right away. So, the authors attempted to discover the most accurate strategy here, which is the Random Forest that matches our model, as we trained our model using gradient boosting, decision trees, K-neighbors, Random forests, SVC, and logistic regression. Then develop a model for future forecasting. The accuracy of this model is given the best accuracy. This will be able to save time, money, and productivity if the problem can be solved quickly and affordably. Cloud computing makes it simple for companies to scale their resources up or down as necessary, which is crucial for companies that encounter swings in demand [3]. To guarantee that their services are reliable and performant, cloud providers invest in top-notch infrastructure and work with teams of specialists, which may reduce.

## 2 Literature Review

The use of cloud computing is growing in popularity. This is aimed at innovation in cloud services. To learn error detection and cloud technologies, designers researched a wide range of subjects. This study suggests a framework for prediction models in a green cloud environment. In order to better comprehend the ideas behind machine learning algorithms, from the literature.

It encompasses a broad variety of crucial sectors, including fuel assets, hybrid energy systems, virtual servers, and infrastructure as a design, according to Patel et al. [3]. It also covers packaging techniques, eco-labeling, and eco-labeling. The research begins with an overview of green technology and the cloud, which is followed by a survey of numerous green technology application domains. A quantitative evaluation of the applicability of many research initiatives on significant green technology subjects is presented.

An approach for defect detection using artificial neural networks is presented in this article by Amin et al. [4]. This approach will close the gaps left by previously established techniques and provide a fault-resilient model.

Zhao et al. have identified that the low-latency software comprises several components, including the socket low-latency communication protocol and virtual machine architecture [5]. The design mandates group situation ordering in addition to direct organization simulcast data dissemination. The joining process is effective, as it enables swift reconfiguration and recovery when a clone experiences a malfunction or joins/leaves a cluster. The virtual strategy utilizes identical sorting to facilitate extensive observation and analysis at storage sites while acquiring scheduling information at open locations. This software exhibits low latency throughput, high clone resilience, and a straightforward interface.

According to Amin et al. an appropriate response to the error tolerance will be produced using a fuzzy-based method in addition to a rigorous examination of the kind and detection of the mistake [6]. The process of applying for jobs can aid in load balancing and error tolerance following an error. Checkpointing techniques can enable re-execution or migration to address the mistake.

The researchers are interested in determining the most efficient way to switch from a network that isn't performing properly to one that is, according to Abro et al. [7]. A digital machine's failure prediction needs to be reasonable because of things like lost tools, time, and money. Virtual machines, which are frequently referred to as virtual computers, are known for their reliability, has raised several questions with the introduction of cloud computing. Reliability technology must include protective measures in order to guarantee system reliability. The recent work in [8–17] shows good contributions in the field of green computing.

### 3 System Architecture and Design

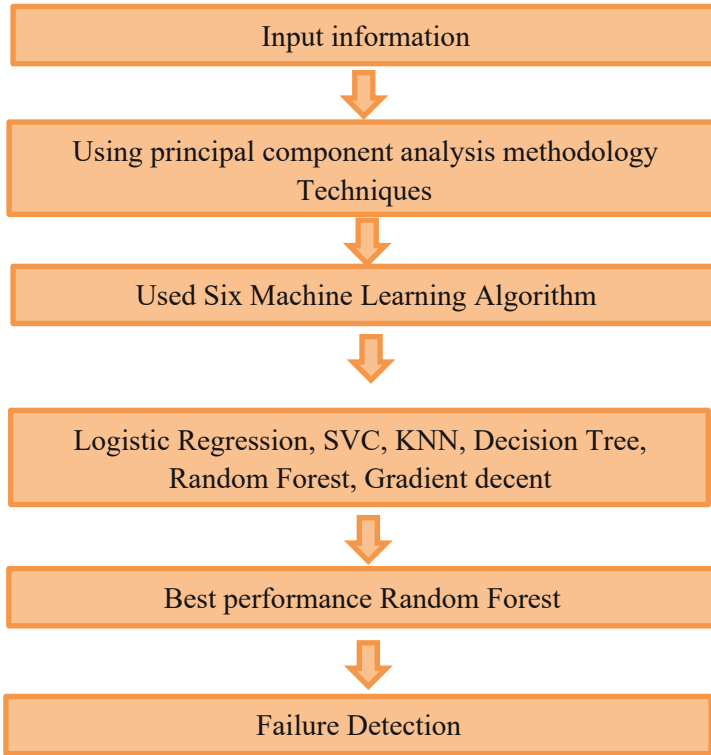
The suggested framework's process is shown in Fig. 1. The suggested method combines model building, information pre-processing, and a machine learning algorithm. This section provides an explanation of the developed model's operation for fault identification. The procedure is broken down into two stages: the single model and before data. Data from the sample dataset should be used first. Utilize principal component analysis than feature scaling techniques [18]. Afterward, six machine learning techniques were employed, namely gradient descent, decision tree, random forest, logistic regression, support vector classifier, and k-nearest neighbor's approach. Based on the highest accuracy, the next architecture finds the best-fitting model as a random forest. It also finds failures towards the end.

#### 3.1 Dataset Description

There are 12 characteristics in our data collection. Timestamp, CPU time, available CPU power, CPU usage in MHz and percentage, bandwidth and memory consumption, disk read and write throughput, network received and sent throughput, and status is the metrics we focus on. The data types are float64 and int64, and there are 8632 total rows and 12 total columns. A dataset that is appropriate for researching fault prediction in the environmentally-friendly cloud computing environment must comprise accurate and relevant information concerning the usage of cloud resources. Conversely, a dataset that is suitable for studying fault tolerance in the green cloud computing environment should include data on previous faults or failures as well as information on the system's response to such failures [19]. This data can be used to develop and test models that can predict the likelihood of failures and evaluate the effectiveness of different fault tolerance strategies.

Table 1 shows 12 features of the dataset used in the model and 5 rows of numerical details of the sample data set.

Table 2 gives overall information on CPU time, bandwidth consumption, memory use, and model status. The table displays the count and means of all numerical values.



**Fig. 1.** System architecture

There are quartiles of CPU time, bandwidth utilization, memory use, and model status displayed.

The correlation matrix for the data set is shown in Fig. 2. In network defect detection, a correlation matrix can be used to investigate the link between several network performance indicators.

Memory consumption on the y-axis and CPU utilization on the x-axis has a 0.6 connection. The relationship between disk read and write throughput is also 0.6. The colors in this heat map signify a  $-1$  to  $1$  scale. The heat map shows 0.0, suggesting that no relationship exists between the two features. The heat map shows 1.0 diagonally, indicating a positive correlation between the two features. When two variables have a positive correlation, it means that an increase in one variable is accompanied by an increase in the other. Conversely, when two variables have a negative correlation, an increase in one variable is accompanied by a decrease in the other [20].

### 3.2 Pre-processing of Data Sources

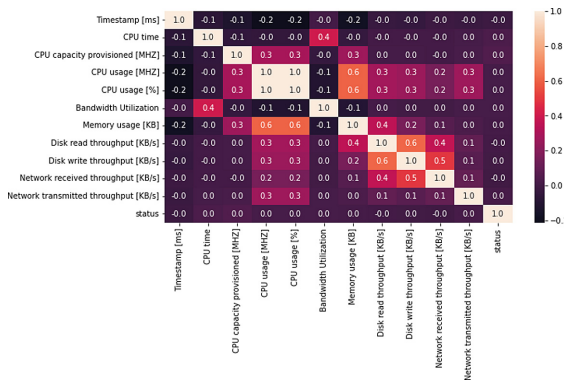
The pre-processing of the cloud information data may enhance the model's performance and accuracy. Figure 3 shows the steps of the data pre-processing.

**Table 1.** Information on the dataset

	Timestamp (ms)	CPU time (ms)	CPU capacity provisioned (MHZ)	CPU usage (MHZ)	CPU usage (%)	Bandwidth utilization (bps)
0	1,376,314,846	0.00	2925.999494	79.977319	2.733333	7.4
1	1,376,315,146	0.00	2925.999494	58.519990	2.000000	7.8
2	1,376,315,446	0.04	2925.999494	85.829318	2.933333	7.8
3	1,376,315,746	0.56	2925.999494	60.470656	2.066667	11.2
4	1,376,316,046	0.00	2925.999494	78.026653	2.666667	7.4
	Memory usage (KB)	Disk read throughput (KB/s)	Disk write throughput (KB/s)	Network received throughput (KB/s)	Network transmitted throughput (KB/s)	Status
0	331,348.8000	0.000000	3.866667	0.066667	0.200000	5
1	301,988.0000	0.066667	3.066667	0.000000	0.133333	5
2	352,320.8000	0.000000	3.533333	0.000000	0.133333	5
3	395,659.7333	0.066667	3.000000	0.133333	0.200000	6
4	300,590.1333	0.000000	3.800000	0.066667	0.266667	5

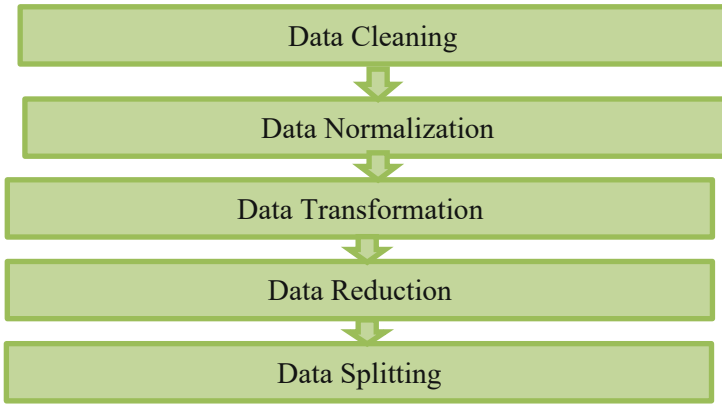
**Table 2.** Overall statistics

	Cpu time (ms)	Bandwidth utilization (bps)	Memory usage (KB)	Status
count	8.632000e^03	8632.000000	8632.000000	8632.000000
Mean	1.377611e^09	2.473569	360,548.862526	5.575765
Std	7.485339e^05	0.608882	61,056.668406	0.758259
Min	1.376315e^09	0.000000	88,078.400000	3.000000
25%	1.376963e^09	2.333333	341,134.666700	5.000000
50%	1.377612e^09	2.533333	362,106.400000	5.000000
75%	1.378259e^09	2.733333	384,475.466700	6.000000
Max	1.378907e^09	11.533333	893,384.800000	8.000000



**Fig. 2.** Correlation matrix

Any extra columns were taken out to increase classifier accuracy via data training. Additionally, we divided the data into training and testing to verify the accuracy of the model. Convert the Boolean vector to the desired value and address any imbalances. Keep the response target as a vector in ‘y’ and the feature matrix in ‘x’. Before performing feature, scaling, and supplying the model with data, the train-test split was complete. A dataset has now been used as a label. Data cleaning involves removing missing or irrelevant data, correcting inconsistent or incorrect data, and resolving any data quality issues. Data normalization involves transforming the data into a common scale to ensure that it is comparable across different data sources. The process of data transformation involves converting data into an analysis-friendly format, such as converting time-series data into numerical data. On the other hand, data reduction involves reducing the dimensionality of the data by eliminating redundant or irrelevant data points. Finally, data splitting involves dividing the data into training, validation, and testing sets to enable the development and evaluation of machine learning models.



**Fig. 3.** Pre-processing step

## 4 Implementation and Experimental Result

The thorough implementation procedure, experimental design, and model performance assessment are all presented in this part.

### 4.1 Experimental Set-Up

**Windows Operating system**—Every experiment in this project was carried out using the Windows operating system. Intel(R) Core(TM) i5 8400, 2.80–2.81 GHz CPU, and 32.0 GB of installed random access memory were the system requirements for this project. The system type is a 64-bit operating system, and the CPU is an x64-based processor. Here, using Jupyter Notebook as our software application.

#### Implementation

The kinds and accuracy of six different machine learning models were examined. Here, used a particular set of configurations is to examine and evaluate the concept. Accuracy is the most fundamental classification metric. It essentially establishes the accuracy % offered by various machine learning methods. However, accuracy is a bad statistic when data is skewed since it is unable to distinguish between various kinds of errors from (1).

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \quad (1)$$

Although a non-exact technique may uncover a substantial number of positives, its selection process is disruptive and often yields false positive findings [21]. Even while it may not classify all positives as positive, a precise model is highly tidy and is very likely to be correct when it does from (1) and (2).

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (2)$$

The F1 score is a measure of a model's accuracy that takes into account both precision and recall. It is the harmonic mean of precision and recall and indicates that both precision and recall are given equal importance [22].

$$\text{F1 score} = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (3)$$

## 4.2 Model Assessment

The logistic regression formula is used to model the probability of a binary outcome (e.g. success/failure or true/false) as a function of one or more input features. It is defined as:

$$p(y = 1|x) = \frac{1}{1 + e^{-z}} \quad (4)$$

Equation (4) shows that 'p(y = 1|x)' represents the probability of a positive outcome (y = 1) based on the input features 'x'. Here, 'e' represents the exponential function, and 'z' represents the linear function of the input features, also known as the log odds or logit,

$$z = w_0 + w_1 \times x_1 + w_2 \times x_2 + \dots + w_n \times x_n \quad (5)$$

Here,  $w_0$ ,  $w_1$ ,  $w_2$  and  $w_n$  are the model parameters from (5), also known as weights, and  $x_1$ ,  $x_2$ , and  $x_n$  are the input features. The weight values are determined via maximum likelihood estimate from the training data.

A supervised machine learning approach that may be used for binary classification is the support vector classifier (SVC). Finding a border (or hyperplane) that broadly divides the two classes of data points is the fundamental tenet of the SVC. This hyperplane's equation is provided by,

$$w_0 + w_1 \times x_1 + w_2 \times x_2 + \dots + w_n \times x_n = 0 \quad (6)$$

where  $w_0$ ,  $w_1$ ,  $w_2$  and  $w_n$  are the model parameters, also known as weights  $x_1$ ,  $x_2$  and  $x_n$  are the input features from (6).

Unlike other methods that use a formula to determine the similarity between two data points, the KNN method relies on a distance metric. The most commonly used distance metric in KNN is the Euclidean distance. It is defined as the square root of the sum of the squares of the differences between the coordinates of two points in n-dimensional space. The formula for the distance metric is:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (7)$$

Here x, and y are two points of n space Euclidean. Now  $x_i$  and  $y_i$  are Euclidean vectors starting from the origin of the initial point and n is space from (7).

Instead of using a formula in the traditional sense, a decision tree employs a set of decision rules to divide the data into classes and give each data point a value. Finding



the characteristic and the related decision threshold that produces the most homogenous and pure subsets of data is the aim of the decision tree method. The decision tree method measures the quality of a split using specific criteria, such as information gain, Gini impurity, or gain ratio. These criteria assess the decrease in uncertainty or impurity after a split and are based on the probability distribution of the classes or values in the data. Starting at the tree's root, the decision tree method divides the data recursively according to the feature and threshold that maximize the criteria at each node until a stopping condition is satisfied, such as a minimum number of data points at a leaf or the tree's maximum depth. A collection of decision rules may be used to categorize or forecast the value of incoming data points as the decision tree's ultimate output.

An ensemble learning method called Random Forest is based on the decision tree algorithm. The final forecast is generated by averaging or majority voting the predictions of the many decision trees, each of which was trained on a distinct subset of the data. The way the data is divided up throughout the training phase is the fundamental distinction between the decision tree and the random forest. By averaging or majority voting the predictions of numerous decision trees, the Random Forest method may assist to avoid overfitting that can occur in decision trees, producing a more robust and accurate model [23].

With the use of the ensemble learning approach known as gradient boosting, many weak models are combined to create a strong model. It doesn't have a set formula, but it may be seen as an iterative procedure to boost a single decision tree's performance. The ensemble's forecasts are combined using a weighted sum or majority vote to get the final prediction.

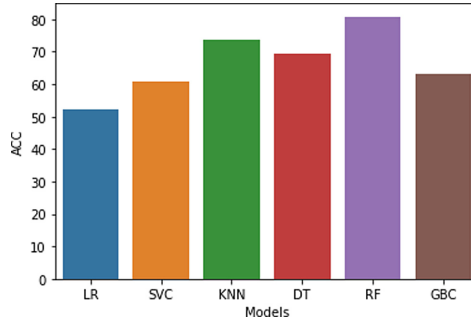
$$f(x) = f_0(x) + \sum_{n=1}^{\infty} (f_n(x)) \quad (8)$$

From (8), where  $f_0(x)$  is the initial tree,  $f(x)$  is the final ensemble, and  $f(x)$  represents the new tree added in iteration  $t$ . The learning rate controls the contribution of each tree in the ensemble.

Figure 4 depicts the results of all six machine-learning approaches. This will save the Random Forest approach in our save model because it clearly outperforms all other algorithms in terms of accuracy.

### 4.3 Performance Analysis

Table 3 compares all algorithms in this area. As from the table, it can be seen, the accuracy is 52.226, the f1 score is 0.506, and the precision score is 0.520. Now obtained a precision score of 0.608, f1 scores of 0.507, and accuracy of 60.87 when this was utilized as the support vector classifier. When this was employed k-nearest neighbors, the precision, f1, and accuracy scores were 0.699, 0.759, and 73.83 respectively. The experiment obtained a precision score of 0.682, an f1 score of 0.699, and an accuracy of 69.38 when this was applied to decision trees. When this was employed in a random forest, it obtained an accuracy of 80.83, a precision score of 0.807, and an f1 score of 0.806. Gradient boosting classifier results in scores of 0.622 for precision, 0.638 for f1, and 63.15 for accuracy.



**Fig. 4.** Compares all algorithms (model vs accuracy)

**Table 3.** Performance assessment from the dataset

Machine learning models	Precision score	F1 score	Accuracy
Logistic regression	0.520330	0.506712	52.226194
Support vector classifier	0.608709	0.506713	60.870978
KNN	0.699779	0.759054	73.838154
Decision tree	0.682638	0.699617	69.385765
Random forest	0.807465	0.806937	80.825479
GBC	0.622747	0.638623	63.145921

#### 4.4 Result Analysis

The use of logistic regression is subject to various restrictions, including the linearity assumption, the use of only binary classification, and overfitting. Limitations in support vector classifiers include interpretability, time complexity, and sensitivity outliers. K-nearest neighbors are subject to computational complexity, large dimensionality, and noisy data restrictions. Limitations of the decision tree include overfitting, instability, a short dataset, and a bias toward features with several outcomes. Additionally, the random forest has certain drawbacks, including overfitting, the ability to only handle categorical variables, and the potential to perform poorly in cases of extreme data imbalance. Gradient-boosting classifiers have several drawbacks, including poor interpretability, large memory requirements, and sensitivity to hyperparameters. These issues make it difficult to understand and fine-tune the model. In contrast, Random Forest is considered one of the best machine learning methods overall because it overcomes many of the shortcomings of other algorithms such as Logistic Regression, K-Nearest Neighbors, Decision Trees, Support Vector Classifiers, and Gradient Boosting Classifiers. It offers better interpretability, requires less memory, and is less sensitive to hyperparameters, making it easier to use and optimize [24]. The prediction values on the new dataset are presented in Table 4.

**Table 4.** Prediction of a new dataset

Features	Values
CPU time	1,376,314,846
CPU capacity provisioned (MHZ)	0.65
CPU usage (MHZ)	2925.99
CPU usage (%)	1.2
Bandwidth utilization	0.065
Idle	15.0
Memory usage (KB)	21.0
Disk read throughput (KB/s)	0.9946
Disk write throughput (KB/s)	3.39
Network received throughput (KB/s)	0.47
Network transmitted Throughput (KB/s)	10.2

In Table 4, a machine learning method is used to anticipate faults in a green cloud environment. The new prediction dataset has twelve characteristics. The outcome demonstrates good defect prediction for green cloud computing using a machine learning system. It can make predictions using our save model. Using an ensemble of decision trees to lower the variance of the model makes Random Forests less prone to overfitting than decision trees. Unlike decision trees, which can only handle categorical data, Random Forests can manage both continuous and categorical variables. Decision trees may not be as effective at managing enormous datasets as Random Forest. When determining which traits are most crucial for accurate predictions, Random Forest may provide an estimate of feature relevance. Using distributed computing frameworks, the Random Forest technique may be readily scaled to huge data applications. Therefore, it is evident from Fig. 3 and Table 3 analysis that random forest outperforms all other algorithms in terms of accuracy, precision score, and f1 score. So, to anticipate faults in a green cloud computing environment, we chose to include random forest in our model.

## 5 Conclusion

In situations where a virtual machine encounters issues such as resource constraints and server unavailability, existing solutions often prioritize migration over recovery. However, there is a need to identify the most effective method for turning a failing system into a successful one. To achieve this goal, algorithms with a single purpose, such as failure prediction, migration, and fault tolerance, can be utilized. Additionally, accepting imperfections can also be considered. Anticipating the failure of a virtual

machine in advance is crucial to avoid issues such as wasted resources, energy, and money. Therefore, a fault tolerance system should include preventative measures to ensure uninterrupted service [25, 26]. Activities may increase the accuracy and speed of defect identification by adding machine learning into network fault detection, decreasing downtime, and boosting overall network dependability. The performance of machine learning-based defect detection, on the other hand, is dependent on the quality of the data used to train the models as well as the availability of labeled data for supervised classification. Machine learning models are regarded as black boxes, which means it is impossible to comprehend why the model generated a specific prediction. Machine learning models are frequently only tested and validated on the same distribution of data that was used to train them, resulting in poor performance when confronted with new and unknown data. Therefore, it is essential to highlight and improve virtual machine proactive failure prediction. When a virtual machine was expected to die, a method was used to securely move its resources to another VM. The relocation was finished quicker because of the compression approach, which markedly boosted resource usage. In order to improve resource efficiency, this study offers machine learning-enabled defect prediction methods for cloud computing.

## References

1. Nidhi, J.K., Jain, C.I.: Cloud load balancing techniques: a step towards green computing. *IJCSI Int. J. Comput. Sci. Issues* **9**(1), 238–246 (2012)
2. Xiong, N., Vandenberg, A., Han, W.: Green cloud computing schemes based on networks: a survey. *IET Commun.* **6**(18), 3294–3300 (2012)
3. Patel, Y.M.N.S.: Green cloud computing: a review on Green IT areas for cloud computing environment. In: 2015 International Conference on Futuristic Trend in Computational Analysis and Knowledge Management, India (2015)
4. Zeeshan, A., Harshpreet, S., Nisha, S.: Review on fault tolerance techniques in cloud computing. *Int. J. Comput. Appl.* **116**(18), 11–17 (2015)
5. Peter, M.-S., Louise, M., Zhao, W.: Fault tolerance middleware for cloud computing. In: 2010 IEEE 3rd International Conference on Cloud Computing, pp. 67–74 (2010)
6. Amin, R., Musa, M., Ahad, F.: Providing a new approach to increase fault tolerance in cloud computing using fuzzy logic. *Int. J. Comput. Appl.* **44**(1), 1–9 (2020)
7. Chunlin, L., Muhammad, S., Vishnukumar, A., Mewada, S., Malpani, L., Jonathan, O.-O., Abro, H.J.: Artificial intelligence enabled effective fault prediction techniques in cloud computing environment for improving resource optimization. *Sci. Program.* **2022**, 1–7 (2022)
8. Yeasmin, S., Afrin, N., Saif, K., Reza, A.W., Arefin, M.S.: Towards building a sustainable system of data center cooling and power management utilizing renewable energy. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_67](https://doi.org/10.1007/978-3-031-19958-5_67)
9. Liza, M.A., Suny, A., Shahjahan, R.M.B., Reza, A.W., Arefin, M.S.: Minimizing E-waste through improved virtualization. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_97](https://doi.org/10.1007/978-3-031-19958-5_97)

10. Das, K., Saha, S., Chowdhury, S., Reza, A.W., Paul, S., Arefin, M.S.: a sustainable e-waste management system and recycling trade for Bangladesh in green IT. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_33](https://doi.org/10.1007/978-3-031-19958-5_33)
11. Rahman, M.A., Asif, S., Hossain, M.S., Alam, T., Reza, A.W., Arefin, M.S.: A sustainable approach to reduce power consumption and harmful effects of cellular base stations. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_66](https://doi.org/10.1007/978-3-031-19958-5_66)
12. Ahsan, M., Yousuf, M., Rahman, M., Proma, F.I., Reza, A.W., Arefin, M.S.: Designing a sustainable e-waste management framework for Bangladesh. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_104](https://doi.org/10.1007/978-3-031-19958-5_104)
13. Mukto, M.M., Al Mahmud, M.M., Ahmed, M.A., Haque, I., Reza, A.W., Arefin, M.S.: A sustainable approach between satellite and traditional broadband transmission technologies based on green IT. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_26](https://doi.org/10.1007/978-3-031-19958-5_26)
14. Meharaj-Ul-Mahmmud, Laskar, M.S., Arafin, M., Molla, M.S., Reza, A.W., Arefin, M.S.: Improved virtualization to reduce e-waste in green computing. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_35](https://doi.org/10.1007/978-3-031-19958-5_35)
15. Banik, P., Rahat, M.S.A., Rafe, M.A.H., Reza, A.W., Arefin, M.S.: Developing an energy cost calculator for solar. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_75](https://doi.org/10.1007/978-3-031-19958-5_75)
16. Ahmed, F., Basak, B., Chakraborty, S., Karmokar, T., Reza, A.W., Arefin, M.S.: Sustainable and profitable IT infrastructure of Bangladesh using green IT. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_18](https://doi.org/10.1007/978-3-031-19958-5_18)
17. Ananna, S.S., Supty, N.S., Shorna, I.J., Reza, A.W., Arefin, M.S.: A policy framework for improving e-waste management in Bangladesh. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_95](https://doi.org/10.1007/978-3-031-19958-5_95)
18. Sandhya, M., Sharmila, S., Ganesh, A.: A study on fault tolerance methods in cloud computing. In: 2014 IEEE International Advance Computing Conference (IACC), pp. 844–849 (2014)
19. Jhavar, R., Universit, I., Piuri, V.: Fault tolerance management in IaaS clouds. In: *Fault Tolerance Management in IaaS Clouds*, pp. 1–6 (2012)
20. kaggle, workloadtrace, [Çevrimiçi]. <https://www.kaggle.com/datasets/ashikhassan007/workloadtrace>. [Erişildi: 2020 2020 2020]
21. Sivagami, K.E.V.M.: An improved dynamic fault tolerant management algorithm during VM migration in cloud data center. *Future Gener. Comput. Syst.* **98**, 35–43 (2019)
22. Arfaeinia, M.M.H., Hagshenas, N.: A fuzzy approach to fault tolerant in cloud using the checkpoint migration technique. *Int. J. Intell. Syst. Appl.* **14**(3), 18–26 (2022)

23. Lei, X., Xiaodong, T., Xiurong, H.: Scientific decision support system of marine environmental management in China's Yellow Sea and Bohai Sea based on cloud computing mode. *J. Intell. Fuzzy Syst.* **37**, 5877–5886 (2019)
24. Jayaram, N.: Green cloud computing. *Int. J. Eng. Comput. Sci.* **11**(5), 25532–25534 (2022)
25. Chee, S.Y., Rajkumar, B., Saurabh Kumar, G.: Green Cloud Framework For Improving Carbon Efficiency of Clouds, vol. 1, pp. 492–502. Springer-Verlag, Berlin, Heidelberg (2011)
26. Radu, L.-D.: Green cloud computing: a literature survey. *Symmetry* **9**(12), 1–20 (2017)



# Transforming the Financial Industry Through Machine and Deep Learning Innovations

Sweta Singh<sup>1</sup>, Rahul Bhagat<sup>2</sup>, S. H. Preeti<sup>3</sup>, and G. P. Girish<sup>4</sup>(✉)

<sup>1</sup> Department of Marketing, IBS Hyderabad, IFHE University, (a Deemed to-be-University Under Sec 3 of UGC Act 1956), Hyderabad, India  
swetasingh@ibsindia.org

<sup>2</sup> Prestige Institute of Management and Research, DAVV Indore, Indore, Madhya Pradesh, India

<sup>3</sup> GITAM School of Business, Hyderabad, GITAM University (a Deemed to-be-University Under Sec 3 of UGC Act 1956), Hyderabad, India

<sup>4</sup> Department of Finance, IBS Bangalore (Off-Campus Centre), IFHE University (a Deemed to-be-University Under Sec 3 of UGC Act 1956), Hyderabad, India  
pggirish.ibs@gmail.com

**Abstract.** The domain of finance has been one of the most extensively delved into application areas for Machine Learning (ML). The application of ML and DL in finance has garnered substantial consideration from both academic researchers and financial industry practitioners over the past few decades. A plethora of studies have been conducted, yielding various models. Within ML domain, Deep Learning (DL) has lately gained a prodigious deal of curiosity, principally due to its superior performance comparative to classical models. DL has taken on many diverse forms, and curiosity in its potential applications lingers to breed. While DL models have instigated to gain traction in the financial sector, there remains much unexploited research potential in this field. In this study we present and review state-of-the-art ML and DL models that have been developed for financial applications. The review will help in identifying possible future applications in related areas of stock market forecasting, algorithmic trading, credit risk assessment, portfolio allocation, asset pricing, and derivatives market.

**Keywords:** Machine learning · Finance · Deep learning · Application

## 1 Introduction

Machine Learning (ML) is an intricate sub-discipline of Artificial Intelligence (AI) that facilitates computers to acquire knowledge and enhance their performance without the need for explicit programming. It is achieved through the construction of complex algorithms that possess the capacity to automatically assimilate and predict outcomes from voluminous datasets. The three forms of Machine Learning are (a) Supervised Learning: The algorithmic model is trained on labeled datasets. The model learns to establish an input-output mapping based on the provided labels. Illustrations include tasks such as

image classification and stock price prediction. (b) Unsupervised Learning: The algorithmic model is trained on unlabeled datasets. The model uncovers patterns and correlations within the data without any explicit instruction. Illustrations include tasks such as anomaly detection and clustering. (c) Reinforcement Learning: The algorithmic model learns by interacting with the surrounding environment. The model receives feedback in the form of rewards or penalties based on its actions and learns to maximize its cumulative reward over time. Illustrations include tasks such as robots and game-playing agents [1–4].

Machine Learning models are extremely useful in finance domain for (a) Fraud Detection: For spotting fraudulent transactions and unearth patterns of fraud. (b) Risk Management: For identifying potential investment risks and assist investors in making informed decisions. (c) Algorithmic Trading: For develop trading strategies that enable the analysis of substantial amounts of data and carry out trades in real-time. (d) Customer Segmentation: For segregating customers based on their unique behavior and preferences. (e) Credit Scoring: For gauging credit risk and determining credit scores with greater accuracy. Deep Learning is a sub-domain of Machine Learning. It constitutes a neural network that is intricately designed to acquire knowledge and extract increasingly intricate and abstract representations of data. Deep Learning and Machine Learning are interconnected since both domains make use of algorithms that possess the ability to learn and enhance their performance based on previous experience. Nevertheless, Deep Learning specifically refers to the use of deep neural networks, consisting of multiple layers of interconnected nodes, to assimilate exceedingly complex representations of data. In essence, Deep Learning embodies an advanced and complex facet of Machine Learning that utilizes deep neural networks to model and solve multifaceted problems. The process involves training these networks on voluminous datasets and enabling them to learn attributes and patterns through numerous layers of abstraction [1, 5–8].

Deep Learning, owing to its exceptional ability to handle complex and large datasets and extract meaningful insights from data, has found numerous applications in the field of finance. In this regard, some of the most prominent applications of Deep Learning in finance are: (a) Fraud Detection: To identify fraudulent transactions and patterns in financial data with a remarkable level of precision. This is a vital application in the finance industry where the detection of fraudulent activities is of utmost importance to protect the interests of financial institutions and their customers. (b) Trading and Investment Decisions: Deep Learning models can effectively analyze vast amounts of financial data to facilitate more informed trading and investment decisions. These models are also adept at predicting market trends and prices, thus helping to identify profitable trading opportunities. (c) Credit Risk Assessment: Deep Learning algorithms can meticulously scrutinize voluminous credit data to determine the creditworthiness of borrowers more accurately. This enhances the precision of credit scoring models, thereby assisting lenders in making more informed credit decisions. (d) Financial Forecasting: Deep Learning models are capable of generating more accurate financial forecasts by analyzing historical financial data and market trends. This is particularly advantageous in investment banking and asset management, where precise financial forecasts are indispensable for decision-making. (e) Customer Segmentation: Deep Learning models can segment customers based on their behavior and preferences, providing financial institutions with valuable insights to



personalize their offerings and marketing campaigns. This, in turn, leads to improved customer engagement and loyalty [1, 9–11].

In this study we present and review state-of-the-art ML and DL models that have been developed for financial applications. The review will help in identifying possible future applications in related areas of stock market forecasting, algorithmic trading, credit risk assessment, portfolio allocation, asset pricing, and derivatives market. The rest of the paper is structured as follows. In Sect. 2 we present various ML and DL models. In Sect. 3 we review the literature from the perspective of application in finance domain. In Sect. 4 we present the application of ML and DL models in Finance, summarize and conclude our study highlighting the way forward.

## 2 ML and DL Models

Advanced ML techniques can help financial institutions better analyze and understand complex financial data, make more informed trading decisions, manage risk more effectively, and identify fraudulent activities more efficiently. DMLP denotes Deep Multilayer Perceptron, which is a feedforward neural network consisting of multiple layers of perceptrons. It is primarily employed for supervised learning tasks, like speech recognition, natural language processing, and image classification. CNN refers to Convolutional Neural Network, which is a specialized neural network architecture for processing data with grid-like topology, such as images or time-series data. It is usually used for tasks such as object detection, image recognition, and video analysis. RNN stands for Recurrent Neural Network, which is a neural network architecture with feedback connections enabling information to be passed from one step of the network to the next. It is primarily used for tasks like speech recognition, time-series analysis, and language modeling. LSTM denotes Long Short-Term Memory, which is an RNN designed to solve the vanishing gradients problem in traditional RNNs. It is typically utilized for tasks like natural language processing, speech recognition, and time-series analysis [12–14].

RBM refers to Restricted Boltzmann Machine, which is a generative stochastic artificial neural network that learns a probability distribution over its input set. It is usually used for tasks such as dimensionality reduction, unsupervised learning, and feature learning. DBN denotes Deep Belief Network, which is a neural network architecture comprising multiple layers of RBMs. It is typically utilized for tasks like dimensionality reduction, unsupervised learning, and feature learning. AE stands for Autoencoder, which is a neural network architecture intended to learn a compressed representation of input data. It is primarily used for tasks like dimensionality reduction, feature learning, and data compression. In finance, these neural network architectures are typically used for tasks like stock price prediction, algorithmic trading, credit risk assessment, and fraud detection. For illustration, CNNs can be utilized to analyze financial time-series data, while LSTMs can be used to model the long-term dependencies in financial data. RBMs and DBNs can be utilized for unsupervised feature learning, which can be valuable for detecting patterns in financial data. AE can be utilized for dimensionality reduction, which simplifies the data and makes it easier to analyze [14–17].

Deep Reinforcement Learning (DRL) has multiple applications in finance, including portfolio optimization, algorithmic trading, and risk management. DRL algorithms

enable learning to make trading decisions based on historical market data and adapt to new market conditions continually. For illustration, DRL can be employed to develop trading agents that learn to optimize trading strategies, manage risk, and maximize profit. Generative Adversarial Networks (GANs) have found applications in finance, such as fraud detection and synthetic data generation. GANs can be trained to produce realistic financial data like stock prices or credit card transactions, which can be used to train machine learning models or simulate financial scenarios. GANs can also be used to identify anomalies or fraudulent transactions in financial data. Capsule Networks are a neural network architecture that models hierarchical relationships between objects more effectively. Capsule Networks have several applications in finance, such as sentiment analysis, credit risk assessment, and fraud detection. For illustration, Capsule Networks can be employed to model the connections between different credit risk factors and detect high-risk borrowers. Deep Gaussian Processes (DGPs) are a probabilistic machine learning model that can be used for regression and classification tasks. DGPs have applications in finance, such as forecasting stock prices, credit risk assessment, and portfolio optimization. For illustration, DGPs can be utilized to model the volatility and correlation structure of financial data and make more accurate predictions of future prices or returns [1].

### 3 Literature Review

Several investigations have utilized sophisticated machine learning techniques to prognosticate the values of cryptocurrencies like Bitcoin, Dash, Ripple, and Litecoin. The datasets implemented in these studies incorporate intricate technical indicators, interest rates, exchange rates, and sentiment analysis. The criteria utilized to assess these predictions encompass an array of performance metrics such as accuracy, F1-measure, accumulative portfolio value, maximum drawdown, Sharpe ratio, F1-score, mean squared error, sensitivity, specificity, precision, and root mean squared error. Diverse machine learning models, including Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), Multilayer Perceptron (MLP), and Reinforcement Learning (RL), have been applied to make these projections [18–20]. Additionally, Bayesian optimization has been leveraged to enhance the performance of LSTM and RNN models.

Bruno [21] used ML techniques such as Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN) and Multilayer Perceptron (MLP) to predict the prices of different cryptocurrencies such as Bitcoin, Dash, Ripple, Litecoin etc. The dataset used in their study included technical indicators such as Moving Average (MA), Bollinger Bands (BOLL), and OCHLV data, as well as interest rates and exchange rates. The performance criteria used were accuracy and F1-measure. In [22] authors used Convolutional Neural Networks (CNN) to predict the prices of Bitcoin and other cryptocurrencies. The dataset used in this study included price data from 2014 to 2017. The performance criteria used were Accumulative Portfolio Value, Maximum Drawdown (MDD) and Sharpe Ratio (SR). In [23] authors used CNN and Reinforcement Learning (RL) to predict the prices of 12 most-volume cryptocurrencies. The dataset used in this study included price data from 2015 to 2016. The performance criteria used were SR, portfolio value, and MDD.

Lihao and Dacheng [24] study used graph embedding and deep Autoencoder (AE) techniques to analyze Bitcoin data. The dataset used in this study included transaction IDs, input/output addresses, and timestamps. The performance criterion used was F1-score. Gonçalo [25] used ML techniques such as CNN, LSTM, and State Frequency Model to predict the prices of Bitcoin, Lite coin, and other cryptocurrencies. The dataset used in this study included OCHLV data, technical indicators, and sentiment analysis. The performance criterion used was Mean Squared Error (MSE). Sean et al. [26] study used Bayesian optimization and ML techniques such as RNN and LSTM to predict the prices of Bitcoin. The dataset used in this study included price data from 2013 to 2016. The performance criteria used were sensitivity, specificity, precision, accuracy, and RMSE.

Numerous investigations have scrutinized varied transactional data types to forecast results through machine learning techniques. For instance, certain studies have employed credit card transactional data to prognosticate transactional accuracy or to identify fraudulent activity, while others have examined foreign trade data or actual-world data procured from an automobile insurance provider. Abhimanyu et al. [27] used credit card transaction data from retail banking in 2017, along with several derived features. They applied LSTM and GRU models to predict the accuracy of transactions. Jon et al. [28] analyzed card purchases' transactions made by customers from 2014 to 2015, and they focused on features related to fraud. They used ANN models to predict the probability of fraud, with AUROC as their performance criteria. Ishan et al. [29] used European cardholders' credit card transactions in 2013 and focused on personal financial variables. They applied ANN and RF models to predict recall, precision, and accuracy. Johannes et al. [30] analyzed credit card transactions in 2015 and used transaction and bank features. They applied LSTM models to predict AUROC. Ebberth et al. [31] used foreign trade data from the Secretariat of Federal Revenue of Brazil in 2014, focusing on eight features related to foreign trade, tax, transactions, employees, and invoices. They applied AE models to predict MSE. Thiago et al. [32] analyzed open data from the Secretariat of Federal Revenue of Brazil and used 21 features, including Brazilian state expense, party name, type of expense, etc. They applied deep Autoencoder and LDA models to predict MSE, RMSE, and other performance criteria. Yibo and Wei [33] used real-world data from an automobile insurance company and labeled it as fraudulent. They used DNN and LDA models to predict TP, FP, accuracy, precision, and F1-score. Longfei et al. [34] analyzed transactions from a giant online payment platform in 2006, focusing on personal financial variables. They applied GBDT + DNN models to predict AUROC.

Numerous investigations have leveraged machine learning algorithms to forecast stock prices using a variety of techniques and features. Some studies have utilized LSTM and CNN + LSTM models with OCHLV features, whereas others have implemented technical indicators and deep learning approaches such as stacked Autoencoder and DNN + RL. To assess these predictions, the studies have employed several performance metrics including MSE, RMSE, MAE, RSE, MAPE, correlation coefficient, and Sharpe Ratio. Furthermore, the performance of the models has been assessed across various computing environments such as Spark, Python, and Keras. Sercan and Ugur [35] used data from GarantiBank in Turkey and applied the LSTM method with OCHLV, spread, PLR, and volatility features to predict stock prices. The study measured the

performance of the model using MSE, RMSE, MAE, RSE, and correlation criteria in the Spark environment. Wei et al. [36] used data from several stock markets, including CSI300, Nifty50, HSI, Nikkei 225, S&P500, and DJIA, to predict stock prices. The study applied a combination of technical indicators and deep learning methods such as WT, stacked Autoencoder, and LSTM to predict stock prices. The study measured the performance of the model using MAPE, correlation coefficient, and THEIL-U in an unknown environment. Shuanglong et al. [37] analyzed Chinese stocks data from 2007 to 2017 using the CNN + LSTM method with OCHLV features to predict stock prices. The study measured the performance of the model using annualized return and mxm retracement criteria in the Python environment. Liheng et al. [38] used price data from 50 stocks listed on the NYSE from 2007 to 2016 and applies the SFM method to predict stock prices. The study measures the performance of the model using MSE criteria. Yue et al. [39] used price data from 300 stocks from SZSE and commodity data from 2014 to 2015. The study applied the FDDR and DNN + RL methods to predict the stock prices and measure the performance of the model using profit, return, and Sharpe Ratio in the Keras environment. Bang et al. [40] analyzed the Singapore Stock Market index using OCHL data of the last 10 days of the index from 2010 to 2017. The study applied the DNN method to predict stock prices and measures the performance of the model using RMSE, MAPE, profit, and Sharpe Ratio criteria. David [41] analyzed GBP/USD price data from 2017 using the Reinforcement Learning + LSTM method with NES. The study measured the performance of the model using SR, downside deviation ratio and total profit.

Multiple studies have applied machine learning models to stock data to predict various performance measures, using different features and environments and using various criteria conducted in different environments. Omer et al. [42] studied the performance of DMLP (Deep Multilayer Perceptron) with a genetic algorithm for predicting the future returns of stocks in Dow30 based on the RSI (Relative Strength Index) feature. The study used annualized return as the performance criteria and the Spark MLlib, Java as the environment. Ariel and Yosi [43] studied the use of FFNN (Feedforward Neural Network) on price data of 10 stocks from S&P500 in SPY ETF to predict cumulative gain. The study used Matlab and MatConvNet as the environment. Luigi et al. [44] studied the use of LSTM (Long Short-Term Memory) on close data of Dow30 stocks and several technical indicators using TALIB to predict accuracy. The study used Python, Keras, and TensorFlow as the environment. Justin and Rama [45] studied the use of LSTM on high-frequency records of all orders and transactions with price data of various assets to predict accuracy. Avraam et al. [46] studied the use of LSTM on price and volume data in the LOB (Limit Order Book) of Nasdaq Nordic stocks to predict precision, recall, F1-score, and Cohen's k. Gudelek et al. [47] studied the use of CNN (Convolutional Neural Network) on price data and technical indicators of 17 ETFs to predict accuracy, mean squared error (MSE), profit, and AUROC (Area Under the Receiver Operating Characteristic Curve). Omer and Ahmet [48] studied the use of CNN with feature imaging on price data and technical indicators of stocks in Dow30 and 9 top volume ETFs to predict recall, precision, and F1-score of annualized return. Guosheng et al. [49] studied the use of CAE (Convolutional Autoencoder) on price data of FTSE100 to predict total return, Sharpe ratio (SR), maximum drawdown (MDD), and mean return. Avraam

et al. [50] studied the use of CNN on price and volume data, and 10 orders of the LOB of Nasdaq Nordic stocks to predict precision, recall, F1-score, and Cohen's  $k$ . Hakan et al. [51] studied the use of CNN on 75 technical indicators and OCHLV (Open, Close, High, Low, and Volume) data of 100 Borsa Istanbul stocks to predict accuracy. Omer and Ahmet [52] studied the use of CNN with feature imaging on price data of ETFs and Dow30 to predict annualized return.

## 4 Application in Finance and the Way Forward

The domain of finance has been one of the most extensively researched application areas for Machine Learning (ML). The application of ML and DL in finance has garnered substantial consideration from both academic researchers and financial industry practitioners over the past few decades. A plethora of studies have been conducted, yielding various models. Within ML domain, Deep Learning (DL) has lately gained a prodigious deal of curiosity, principally due to its superior performance comparative to classical models. DL has taken on many diverse forms, and curiosity in its potential applications lingers to breed. In this study we presented and reviewed ML and DL models that have been developed for financial applications. The review will help in identifying possible future applications in related areas of stock market forecasting, algorithmic trading, credit risk assessment, portfolio allocation, asset pricing, and derivatives market to all researchers and stakeholders from industry as well as academia.

It is feasible to assert that the LSTM model is the dominant DL model that most researchers prefer due to its established structure for financial time series data forecasting. As financial data usually has time-varying data representations that require regression-type approaches, LSTM and its derivatives are well-suited due to their ability to adapt to problems easily. DRL-based implementations, particularly those coupled with agent-based modeling, are another model that is gaining interest. Although algorithmic trading is the preferred implementation area, it is possible to develop working structures for any problem type. Most studies in literature shows that hybrid models are preferred over native models for better accomplishments. However, creating more complex hybrid models that are not easy to build and interpret can also pose a danger.

Through performance evaluation results, it is generally possible to claim that DL models outperform ML counterparts when working on the same problems. DL models have the advantage of being able to work on larger amounts of data. With the growing expansion of open-source DL libraries and frameworks, DL model building and development processes are easier than ever. Price/trend extrapolation and algo-trading models have the maximum curiosity amongst all financial solicitations that use DL models in their applications. Risk assessment and portfolio management have always been popular, and it looks like this is also valid for DL researchers.

Financial text mining is particularly gaining more attention than most other financial applications. The streaming flow of financial news, tweets, statements, and blogs has opened up a whole new world for the financial community, allowing them to build better and more versatile prediction and evaluation models integrating numerical and textual data. The general approach nowadays is to combine text mining with financial sentiment analysis, which is likely to achieve higher performance. Many researchers are

working on this particular application area, and the next generation of outperforming implementations will likely be based on models that can successfully integrate text mining with quantified numerical data. Cryptocurrencies are one other hot area within DL research, and block chain research is also gaining attention. Cryptocurrency price prediction has the most attraction within the field, but since the topic is fairly new, more studies and implementations will likely keep pouring in due to the high expectations and promising rewards.

## References

1. Ahmet, M.O., Mehmet, U.G., Omer, B.S.: Deep learning for financial applications: a survey. *Appl. Soft Comput.* **93**, 106384 (2020)
2. Sendhil, M., Jann, S.: Machine learning: an applied econometric approach. *J. Econ. Perspect.* **31**, 87–106 (2017)
3. Keke, G., Meikang, Q., Xiaotong, S.: A survey on fintech. *J. Netw. Comput. Appl.* **103**, 262–273 (2018)
4. Boris, K., Evgenii V.: *Data Mining in Finance: Advances in Relational and Hybrid Methods*. Kluwer Academic Publishers (2000)
5. Bo, K.W., Yakup, S.: Neural network applications in finance: a review and analysis of literature (1990–1996). *Inf. Manag.* **34**, 129–139 (1998)
6. Yuhong, L., Weihua, M.: Applications of artificial neural networks in financial economics: a survey. In: 2010 International Symposium on Computational Intelligence and Design. IEEE (2010)
7. Blake, L.: *Handbook of Computational Economics*, vol. 2. Elsevier, pp. 1187–1233 (2006)
8. Stephan K.C., Andreas, M.: Kernel methods in finance. In: *Handbook on Information Technology in Finance*, pp. 655–687. Springer, Berlin Heidelberg (2008)
9. Rafik A.A., Bijan F., Rashad R.A.: Soft computing and its applications in business and economics. In: *Studies in Fuzziness and Soft Computing* (2004)
10. Anthony, B., Michael, O.: *Natural Computing in Computational Finance*. Springer, Berlin Heidelberg (2008)
11. Ludmila, D.: *Soft Computing in Economics and Finance*. Springer, Berlin Heidelberg (2011)
12. Yann, L., Yoshua, B., Geoffrey, H.: Deep learning. *Nature* **521**, 436–444 (2015)
13. Sepp, H., Jürgen, S.: Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997)
14. Xueheng, Q., Le, Z., Ye, R.P.S., Gehan, A.: Ensemble deep learning for regression and time series forecasting. In: 2014 IEEE Symposium on Computational Intelligence in Ensemble Learning (CIEL), pp. 1–6 (2014)
15. Yoshua, B.: Deep learning of representations for unsupervised and transfer learning. In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pp. 17–36 (2012)
16. Geoffrey, E.H., Simon, O., Yee, W.T.: A fast learning algorithm for deep belief nets. *Neural Comput.* **18**, 1527–1554 (2006)
17. Pascal, V., Hugo, L., Yoshua, B., Pierre, A.M.: Extracting and composing robust features with denoising autoencoders. In: *Proceedings of the 25th International Conference on Machine Learning*, pp. 1096–1103 (2008)
18. Intelligent computing & optimization. In: *Proceedings of the 4th International Conference on Intelligent Computing and Optimization 2021 (ICO2021)*. <https://doi.org/10.1007/978-3-030-93247-3>
19. Intelligent computing & optimization. In: *Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022)*. <https://doi.org/10.1007/978-3-031-19958-5>

20. Intelligent computing & optimization. Special Issue. <https://link.springer.com/journal/40305/volumes-and-issues/10-4>
21. Bruno, S.: Deep neural networks for cryptocurrencies price prediction. Master's thesis, Humboldt-Universität zu Berlin, Wirtschaftswissenschaftliche Fakultät (2018)
22. Zhengyao, J., Dixing, X., Jinjun, L.: A deep reinforcement learning framework for the financial portfolio management problem (2017)
23. Zhengyao, J., Jinjun, L.: Cryptocurrency portfolio management with deep reinforcement learning. In: 2017 Intelligent Systems Conference IEEE (2017)
24. Lihao, N., Dacheng, T.: Bitcoin mixing detection using deep autoencoder. In: 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC). IEEE (2018)
25. Gonçalo, D.L.F.L.: Deep learning for market forecasts (2018)
26. Sean, M., Jason, R., Simon, C.: Predicting the price of bitcoin using machine learning. In: 2018 26th Euromicro International Conference on Parallel, Distributed and Network-Based Processing IEEE (2018)
27. Abhimanyu, R., Jingyi, S., Robert, M., Loreto, A., Stephen, A., Peter, B.: Deep learning detecting fraud in credit card transactions. In: 2018 Systems and Information Engineering Design Symposium (SIEDS). IEEE (2018)
28. Jon, A.G., Juan, A., Roberto, P., Jordi, N.: End-to-end neural network architecture for fraud scoring in card payments. *Pattern Recogn. Lett.* **105**, 175–181 (2018)
29. Ishan, S., Rameshwar, P., Ullas, N.: Ensemble learning for credit card fraud detection. In: Proceedings of the ACM India Joint International Conference on Data Science and Management of Data—CoDS-COMAD18. ACM Press (2018)
30. Johannes, J., Michael, G., Konstantin, Z., Sylvie, C., Pierre, E.P., Liyun, H.G., Olivier, C.: Sequence classification for credit-card fraud detection. *Expert Syst. Appl.* **100**, 234–245 (2018)
31. Eberberth, L.P., Marcelo, L., Rommel, N.C., Thiago, M.: Deep learning anomaly detection as support fraud investigation in Brazilian exports and anti-money laundering. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE (2016)
32. Thiago, A.G., Rommel, N.C., Ricardo, S.C.: Identifying anomalies in parliamentary expenditures of Brazilian chamber of deputies with deep autoencoders. In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE (2017)
33. Yibo, W., Wei, X.: Leveraging deep learning with lda-based text analytics to detect automobile insurance fraud. *Decis. Support Syst.* **105**, 87–95 (2018)
34. Longfei, L., Jun, Z., Xiaolong, L., Tao, C.: Poster: practical fraud transaction prediction. In: ACM Conference on Computer and Communications Security (2017)
35. Sercan, K., Ugur, A.: A deep learning approach for optimization of systematic signal detection in financial trading systems with big data. *Int. J. Intell. Syst.* **Special Issue**, 31–36 (2017)
36. Wei, B., Jun, Y., Yulei, R.: A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PLoS ONE* **12**, e0180944 (2017)
37. Shuanglong, L., Chao, Z., Jinwen, M.: CNN-LSTM neural network model for quantitative strategy analysis in stock markets. In: *Neural Information Processing*, pp. 198–206. Springer International Publishing (2017)
38. Liheng, Z., Charu, A., Guo, J.Q.: Stock price prediction via discovering multi-frequency trading patterns. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining—KDD17. ACM Press (2017)
39. Yue, D., Feng, B., Youyong, K., Zhiquan, R., Qionghai, D.: Deep direct reinforcement learning for financial signal representation and trading. *IEEE Trans. Neural Netw. Learn. Syst.* **28**, 653–664 (2017)

40. Bang, X.Y., Mohd. R.A.R., Ahmad, S.A.: A stock market trading system using deep neural network. In: *Communications in Computer and Information Science*, pp. 356–364. Springer Singapore (2017)
41. David, W.L.: Agent inspired trading using recurrent reinforcement learning and LSTM neural networks (2017)
42. Omer, B.S., Murat, O., Erdogan, D.: A deep neural-network based stock trading system based on evolutionary optimized technical analysis parameters. *Proc. Comput. Sci.* **114**, 473–480 (2017)
43. Ariel, N., Yosi, K.: Financial time series prediction using deep learning (2017)
44. Luigi, T., Elena, M.V., Vincenzo, L.: Replicating a trading strategy by means of LSTM for financial industry applications. *IEEE Trans. Indus. Inform.* **14**, 3226–3234 (2018)
45. Justin, S., Rama, C.: Universal features of price formation in financial markets: perspectives from deep learning. *SSRN Electron. J.* (2018)
46. Avraam, T., Nikolaos, P., Anastasios, T., Juho, K., Moncef, G., Alexandros, I.: Using deep learning to detect price change indications in financial markets. In: *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE (2017)
47. Gudelek, M.U., Boluk, S.A., Ozbayoglu, A.M.: A deep learning based stock trading model with 2-d CNN trend detection. In: *2017 IEEE Symposium Series on Computational Intelligence*. IEEE (2017)
48. Omer, B.S., Ahmet, M.O.: Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. *Appl. Soft Comput.* **70**, 525–538 (2018)
49. Guosheng, H., Yuxin, H., Kai, Y., Zehao, Y., Flood, S., Zhihong, Z., Fei, X., Jianguo, L., Neil, R., Timpathy H., Qiangwei, M.: Deep stock representation learning: from candlestick charts to investment decisions. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE (2018)
50. Avraam, T., Nikolaos, P., Tefas, A., Juho, K., Moncef, G., Iosifidis, A.: Forecasting stock prices from the limit order book using convolutional neural networks. In: *2017 IEEE 19th Conference on Business Informatics*. IEEE (2017)
51. Hakan, G., Yusuf, Y., Zehra, C.: Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations. *Knowl. Based Syst.* **137**, 138–148 (2017)
52. Omer, B.S., Ahmet, M.O.: Financial trading model with stock bar chart image time series with deep convolutional neural networks (2019)





# MRI-Based Brain Tumor Classification Using Various Deep Learning Convolutional Networks and CNN

Md. Saiful<sup>1</sup>, Sakib Haider<sup>1</sup>, S. M. Arafat Rahman<sup>1</sup>, Nahid Reza<sup>1</sup>,  
Ahmed Wasif Reza<sup>1</sup>(✉), and Mohammad Shamsul Arefin<sup>2,3</sup>(✉)

<sup>1</sup> Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh  
wasif@ewubd.edu

<sup>2</sup> Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh  
sarefin\_406@yahoo.com

<sup>3</sup> Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh

**Abstract.** Yearly, brain tumors cause many fatalities and a significant portion of these victims come from rural regions. However, beginning brain tumor diagnosis technology is not as effective as anticipated. We thus set out to develop an accurate approach that would aid doctors in recognizing brain tumors. Even though there have been several types of research on this topic, we tried to develop a classification approach that is significantly more accurate and error-free and is trained using a sizable amount of authentic datasets rather than an enhanced data-modified version of the VGG-16 convolutional neural network architecture was used to analyze a dataset of 6328 MRI images that were categorized into three different types: Pituitary, Glioma, and Meningioma. The results were highly impressive, with the model achieving an overall accuracy of 99.5%. The precision rates for each type were also outstanding, with a precision rate of 99.4% for gliomas, 96.7% for meningiomas, and 100% for pituitaries. These results suggest that the modified VGG-16 architecture is highly effective in accurately classifying MRI images of the brain into these three distinct categories. Additionally, it outperformed various other current CNN designs and cutting-edge research in terms of outcomes.

**Keywords:** VGG-16 · MRI · Pituitary · Glioma · Meningioma

## 1 Introduction

Brain tumors can be deadly and infrequent. They can appear anywhere and have different picture intensities [1], exist in a range of sizes and shapes, and can take many different forms. Brain tumors can either be primary or metastatic, depending on their place of origin. In contrast to metastatic cancer cells, which, before traveling to the brain, form cancer in any other part of the body, primary cancer cells develop in brain tissue [2]. For the best possible planning of a patient's care and treatment, quick detection of a brain tumor is essential.

To gather information about tumors, several medical imaging modalities are employed. More thorough information about tumors can be generated by combining several technologies. On the other hand, due to its several advantages, MRI is the most commonly used technology. MRI is a nonsurgical Intra vivo imaging technique that stimulates target tissues with radiofrequency radiation, allowing them to produce internal pictures when magnetic magnets are applied. Without utilizing ionizing radiation, the scan produces dozens of 2D picture layers with excellent soft tissue contrast in MRI collections [3].

Based on their education and expertise, radiologists manually classify the anomalies of the brain MRI into different categories of tumors [4]. The evaluation procedure requires early brain tumor identification and classification. The potential for misdiagnosis of the tumor, which might result in patients receiving inappropriate medication, is one of the biggest challenges of manual detection [5]. Additionally, manual diagnosis offers little benefit in this situation because time is a crucial factor. Consequently, it makes sense that a speedy and automated detection method would be required [6].

Convolutional neural networks (CNNs) employ various types of layers, including pooling and convolution layers, for tasks such as feature extraction, classification, segmentation, and dimensionality reduction of feature representations [7].

The work's main objective was to build a reliable model that can operate reliably and quickly to identify different types of tumors. The suggested method uses an altered CNN model that is modeled after the Visual Geometry Group (VGG) architecture, which outperforms several other architectures substantially. Due to the sensitivity of the results and the possibility of a life-threatening disease from a mistaken diagnosis, Working with medical images is harder than normal. We focused primarily on attaining a high precision that can surpass the effects of other widely used architecture as a result.

Following the intro, we go into relevant studies that pertain to our topic and how their research may be evaluated in the Comparable Studies segment. The training of the model and performance evaluations are described in the Result section. In the Results part, we have examined the outcome of our suggested design; in the Discussion part, we have evaluated the outcomes and evaluated our proposed architecture's performance with other approaches. In the Conclusions portion of the paper, we also put everything to rest.

## 2 Related Work

To detect brain cancer on MRI scans, we have many techniques on our hands. Neural Networks (NN) and SVM are the most applied approaches [6].

The decision-making procedure is separated into two parts: key component inquiry-based extraction and CNN-based classification. To distinguish between various types of brain tumors, we have developed a CNN-based deep neural net model that achieves an accuracy rate of almost 96% [8].

A small kernel Convolution layer, which uses  $3 \times 3$  kernels for every convolutional layer with a single phase, can also be utilized for categorization. The result in the larger dataset still displays a 90.67% accuracy [9].

Deep learning is utilized to extract characteristics from two distinct CNN models. The robust features produced by this method are categorized by ELM [10].

To accurately segment the true lesion, the Grabbing cut method is used. In this work, segmentation is demonstrated using Unit architecture using ResNet50 as a standard. The identification of brain malignancies using evolutionary approaches (such as MobileNet V2, ResNet50, transfer learning, and so on) is effective when used with reinforcement learning. to 91.8%, 92.8%, 92.9%, 93.1%, and 99.6%, correspondingly. Networks like Inception V3, Resnet 50, and VGG-16 provide trained parameters, followed by a multi-level CNN model constructed, achieving a classifier accuracy of 99.89% through its use [11].

The given DeepTumorNet hybrid deep learning model is built by modifying the GoogleNet architecture's layers and integrating a leaky ReLU activation function. This architecture's precision was 99.66% [7, 12].

The proposed architecture of a differential deep neural network uses differential operators to create the differential feature maps of CNN, yielding an accuracy of 99.25% [13, 14]. The works in [15–19] also focused on image analysis for performing different important tasks.

### 3 System Architecture and Design

The procedure of our suggested model is shown in Fig. 1. Before The dataset is then divided into testing and training halves, images should be first loaded and then go through numerous important pre-processing steps.

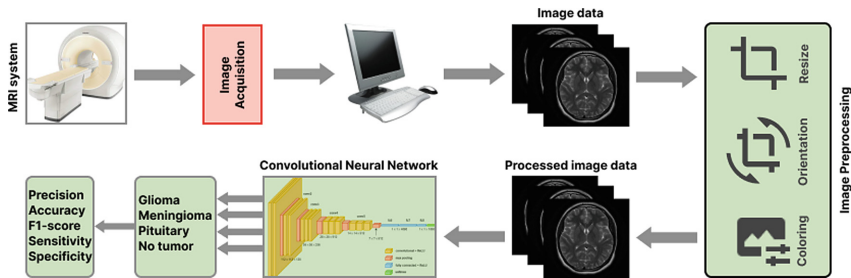
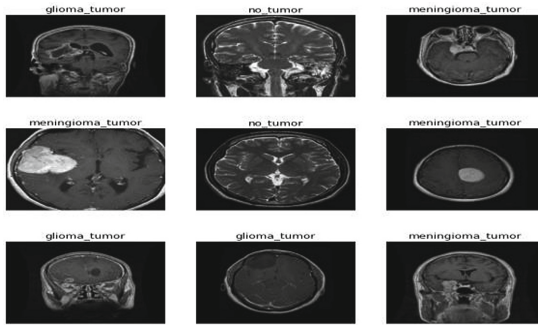


Fig. 1. Design of the proposed model

#### 3.1 Description

We were seeking a relatively large dataset for this study because we did not want to use the data augmentation method. We also required a dataset with different subtypes of normal and pathological brain pictures. We used two public-access Kaggle datasets [20, 21] to solve this issue. To extend the dataset, we uploaded a little amount of data from several sources. Ultimately, a dataset of 10,153 MRI pictures was acquired, and Fig. 2 displays the samples from each class. There are 10,153 photos in total. For the classifications Meningioma, Glioma, Pituitary, and No Tumor, there are 2582, 2547, 2658, and 2396 data points, respectively. The data range in size from 2.3 to 2.6 k. We did not need to utilize any extra procedures to cope with this little data imbalance because there are no serious issues with data imbalance in the dataset.



**Fig. 2.** Image of the healthy brain and the three further forms of brain tumors-pituitary, glioma, and meningioma

### 3.2 Data Preprocessing

The dataset should be preprocessed before being used to train the proposed model. The graphics were of various sizes. We resize the image to  $200 \times 200 \times 1$  pixels. It offers higher performance and simpler computation as a consequence. The MRI picture size was reduced from  $256 \times 256 \times 1$  (the biggest was  $500 \times 500 \times 1$ ) to  $200 \times 200 \times 1$  du. The appropriate size is selected to ensure that the entire skull is visible, and after cropping and resizing, the photographs have a centering effect. The raw data were all oriented the same way. We've chosen grayscale images, and before moving on, we made sure that the images were grayscale for the betterment of the implementation. We shuffled the data set before dividing it into training and testing portions, with training comprising 80% of the dataset and testing 20%. In this instance, the dataset is split into 20% for testing and 80% for training. We used the 20% testing data as validation even though we didn't do any independent data validation.

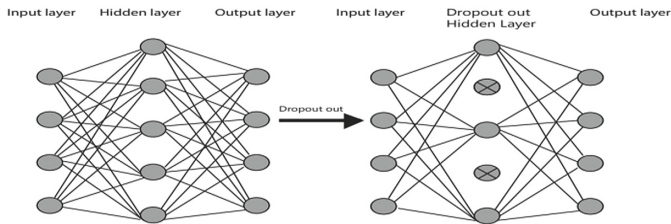
### 3.3 Proposed Model

We can observe from Fig. 1, the architectural layout of the suggested model. In this scenario, a structure similar to the VGG-16 design was used. There are numerous levels in the VGG-16 deep CNN architecture. In order to use a restricted perceptron, we used a number of (deeper architecture) convolution layers. To reduce the number of pixels in the output of the convolutional layer, we incorporated a max pooling layer, following the example of the VGG model. The purpose of this was to decrease the pixel density.

Overfitting is a common issue when constructing a robust model with many parameters on a small training dataset. In order to deal with overfitting, dropout regularization is used. The use of dropout layers is intended to lessen overfitting. For our particular model, a 20% dropout has come with the best results. We can observe the nodes from Fig. 3 which all have connections to the output layer. To help our model overcome overfitting, some of the nodes are overlooked after dropping out.

The algorithms listed below illustrate the protocol required to put our suggested system into practice. Initially, the steps for putting our full model into practice:

Procedure:



**Fig. 3.** Nodes are connected with a full network and dropout hidden layer

1. Install Google Drive to access the RGB image dataset.
2. Create an empty NumPy array for the labels.
3. Create a 4D NumPy array to store the RGB images.
4. Read data from the dataset folder and add photos with RGB values to the array. Assign a label to each image.
5. Split the dataset into training and test sets.
6. Encode the labels.
7. Build a CNN model.
8. Train the model on the training dataset.
9. Evaluate the model using the holdout test dataset.

We must be able to load the dataset into memory in order to use it to train a model. In order to achieve this, we employed Algorithm 1, which loads the preprocessed photos from our dataset together with the proper class labels into NumPy arrays.

**Algorithm 1:** After image processing the process image must be in a single array. We need to train the model on all images. The next procedure loads the images as a NumPy array into memory.

1. Begin the process.
2. Load the dataset containing images from a given directory.
3. Create an empty numpy array to hold the images and name it 'image\_array'.
4. Retrieve the list of all classes present in the dataset.
5. For each class in the list:
6. Get a list of all the image files belonging to that class.
7. Create a temporary numpy array with dimensions of  $4 \times 4$  to hold grayscale values of each image.
8. Loop through each image file in the list and read the image into the temporary array. Stack the temporary array onto the 'image\_array' numpy array.
9. Add the corresponding class label to an array of labels.
10. End the loop for all classes.
11. Split the 'image\_array' and label array into training and testing sets with a ratio of 0.25 using a built-in function.
12. Convert the training and testing labels to one-hot encoding format.
13. Build a unique sequential CNN model for the dataset.
14. End the process

The second algorithm (Algorithm 2) uses our suggested model after that. This uses a unique sequential CNN model.

**Algorithm 2:** Implementation of a customized CNN model. The layers of our model are constructed using the following algorithm.

**Input:** Data Image Directory

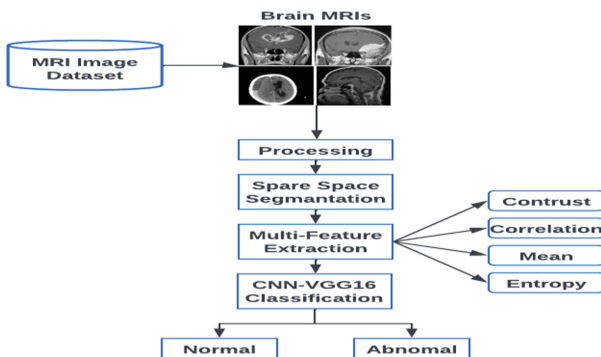
**Output:** Model Training

1. Begin the process.
2. Set up the model as a sequential one. Set the input size to (100,100,3).
3. Add the first convolution layer with 48 channels, ReLU activation, and a  $3 \times 3$  kernel.
4. Add the first max pooling layer with a pool size of  $3 \times 3$ .
5. Add the second convolution layer with 78 channels, ReLU activation, and a  $4 \times 4$  kernel.
6. Add the second max pooling layer with a pool size of  $3 \times 3$ .
7. Add a dense layer with 2000 neurons and ReLU activation.
8. Add a dropout layer with a rate of 0.5.
9. Add a second dense layer with 700 neurons and ReLU activation.
10. Add a second dropout layer with a rate of 0.5.
11. Add a third dense layer with 250 neurons and ReLU activation.
12. Add a dense layer with five neurons and ReLU activation as the output layer.

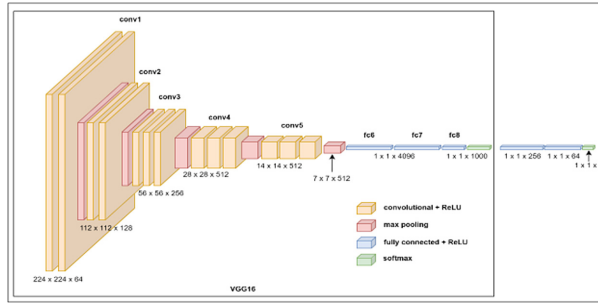
Compile the model using Adam optimizer, categorical cross-entropy loss, and accuracy as the assessment matrix. End the process.

### 3.4 Architecture and Design Proposed Model

For the brain tumor classification challenge, we have created a number of models. One is a modification of the VGG16 CNN models to fit our needs for our increased brain tumor detection jobs, and the other is our own (I) unique bespoke model. Both models' structures have been graphically represented in Figs. 4 and 5, respectively.



**Fig. 4.** The modified VGG-16 model



**Fig. 5.** Proposed unique VGG model for classifying brain tumors

In our proposed model, we utilized the stochastic gradient descent (SGD) optimizer, which is a popular and well-established optimization algorithm for training neural networks. This optimizer multiplies the gradient of the loss function with the learning rate and subtracts the resulting value from the weights of the model. Despite being a simple and straightforward approach, SGD has a strong theoretical foundation and is still commonly used in training convolutional neural networks, especially for edge devices [13].

An efficient CNN training method is batch normalization, which handles the input for each layer for each mini-batch [14]. In this case, batch normalization serves the following purpose.

In the final layer of our suggested model, we used the softmax activation function. We also implemented a normalization technique for the B channel of a mini-batch. This involved multiplying a randomly generated scaling value and adding a randomly generated shift value to the B channel. Initially, the scaling and shift values were set to either 0 or 1, but they were updated after each epoch to reduce the issue of changing input values and accelerate the training process. Our batch size consisted of 42 images. Overall, we aimed to stabilize the input values and improve the training efficiency of our model.

## 4 Model Evaluation Result

### 4.1 Hypothetical Setup

The 13 gigabytes of RAM and 108 Gigabytes of disk space Google Collab laptops system, that has been made available to us, has implemented fully all desired systems. The backend is Python 3 on Google Computing Cloud, and the TPU option is activated for speedier training. The supervised classification frameworks TensorFlow and Keras' built-in functions and built-in layers are used to build the models.

### 4.2 Hypothetical Result

For the suggested system, we used a modified VGGNet-CNN architecture. With a  $200 \times 200$  input vector size, a total of 4456 MRI image samples were employed. Eighty percent

of the dataset was divided into training and twenty percent into testing sets. Eighty percent of the testing set’s data were retained for validation, while the remaining twenty percent were used for testing. Three epoch settings (30, 40, and 50) were used to test the model, and it was evaluated with three different learning rates (0.001, 0.05, and 0.01) for each epoch.

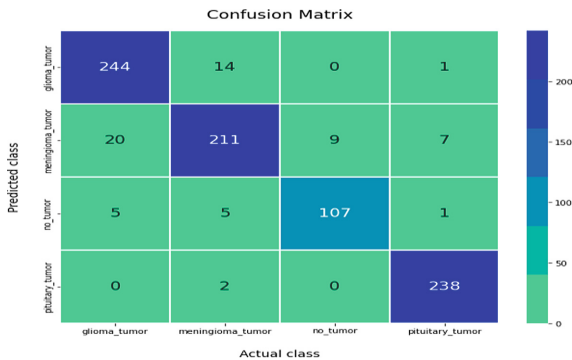
**Table 1.** Performance accuracy of the model compared for Various epochs and learning rates used.

Epoch	40	50
LR	Accuracy	Accuracy (%)
0.001	99.2%	99.3
0.01	<b>99.7%</b>	99.02
0.05	98.9%	99.4

With the following hyperparameters, the best precision of 99.7% was discovered: epoch as 20, momentum as 0.9, and learning rate as 0.01. Table 1 presents the outcomes.

### 4.3 Confusion Matrix

In Fig. 6, the x-axis represents predicted values, while the y-axis represents actual or ground truth values. Evaluation measures including True Positive, True Negative, False Positive, False Negative, Precision, Recall, Sensitivity, Specificity, Accuracy, and F-1 Score were constructed to assess the performance of this model.



**Fig. 6.** A confusion matrix evaluates how accurately the model performs.

The proposed system has been evaluated using a confusion matrix and various performance metrics such as F1 score, sensitivity, specificity, accuracy, precision, and recall. The terms True Positive (TP), False Negative (FN), True Negative (TN), and False Positive (FP) have been used to calculate these metrics. In Table 2, the highest scores for



precision, recall, and specificity have been highlighted in bold. The accuracy for glioma is 98.98%, meningioma is 99.13%, pituitary is 99.95%, and no tumor is 99.81%.

**Table 2.** The proposed model evaluated for its performance.

Classified	TP	TN	FP	FN	Precision	Recall	Specificity	Accuracy	F1-score
Glioma	244	1510	3	18	0.994	0.965	0.958	98.94	0.99
Meningi-oma	211	1532	17	1	0.965	<b>0.999</b>	<b>0.998</b>	99.14	0.98
No Tumor	107	1601	2	2	0.995	0.996	0.997	99.3	0.99
Pituitary	238	1516	0	1	<b>1.00</b>	<b>0.996</b>	<b>0.998</b>	99.75	0.99

#### 4.4 Performance Evaluation

We can infer from earlier conversations that VGG16 operates more effectively in our system. In the I verdict in favor, VGG16 provides an accuracy of 99.2% while the customized CNN model provides an accuracy of 98.4% (Table 3). VGG16 provides us with less value in validation loss as well.

**Table 3.** Accuracy and loss values used as performance measures for evaluating the model.

Models	Tauc (%)	TLoss	Vauc (%)	VLoss
Baseline CNN	92	0.0314	95.4	0.0713%
VGG16	96	0.0032	99.2	0.0384

Our model has been assessed in two ways: first, by the precision of a test set, and second, by providing an image as input of brain tumor and evaluating whether our system can identify it or not. Below are the graphs of our bespoke deep learning model's accuracy and loss across 50 iterations (Fig. 7).

Accuracy and loss of the VGG16 model against 50 epochs (Fig. 8).

The Fully convolutional model achieves high accuracy with fewer epochs than the bespoke CNN model. However, the bespoke CNN model is faster to train, taking only 14 s compared to VGG16's 66 s per epoch. Thus, the custom model is trained much faster than the VGG16 model.

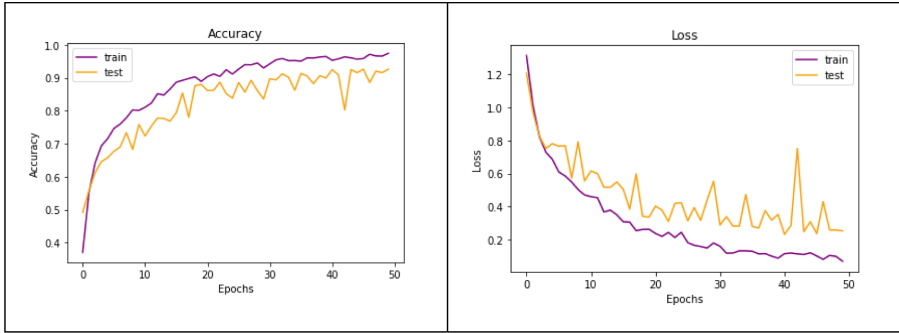


Fig. 7. The modified model evaluated for its accuracy and loss.

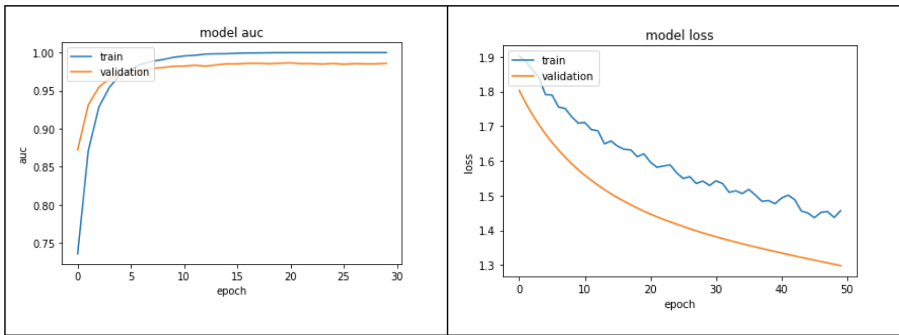


Fig. 8. Accuracy and loss of baseline CNN VGG16

### 5 Conclusion

This technique may be applied for industrial applications as we were successfully achieving an accuracy of 99.5%, which represents a great result. The efficiency was 91.28%, 91.43%, 90.89%, and 94.2%, respectively, when compared to prior multi-classification-classification studies as SVM and KNN in [10], CNN from [11]. Our efficiency using their dataset was 97.8%, 96.7%, 96.01%, and 96.3%, respectively. However, we discovered a considerably high accuracy of 99.8%. The system still has to be taught to recognize brain cancers early on. When it comes to a person’s health, identifying is without a doubt crucial. The fact that this model is rather sluggish due to its large number of parameters, which may be resolved by applying a variety of other effective current models, is one of its major disadvantages. Finding the location of the tumor will be made easier by the model’s ability to be trained to act with 3D images. It may be trained to spot brain tumors early in the process.

### References

1. DeAngelis, L.M.: Brain tumors. N. Engl. J. Med. **344**, 114–123 (2001)

2. Sharif, M.I., Khan, M.A., Alhoussein, M., Aurangzeb, K., Raza, M.: A decision support system for multimodal brain tumor classification using deep learning. *Complex Intell. Syst.* **8**, 3007–3020 (2021)
3. Day, M.R., Lacarra, E.D., Sacks, J.M., Sadraei, A.: Theranostics of glioblastoma multi-forme: In Vitro Characterization of Targeted Nanoemulsions and Creation of a 3D Statistical Heatmap to Visualize Nanoemulsion Uptake (2014). <https://web.wpi.edu/Pubs/E-project/Available/E-project-043015-050001/>. Accessed 13 July 2022
4. Işin, A., Direko, C., Şah, M.: Review of MRI-based brain tumor image segmentation using deep learning methods. *Procedia Comput. Sci.* 2016, 102, 317–324
5. Litjens, G., et al.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)
6. Mohan, G., Subashini, M.M.: MRI based medical image analysis: survey on brain tumor grade classification. *Biomed. Signal Process. Control* **39**, 139–161 (2018)
7. Raza, A., Ayub, H., Khan, J.A., Ahmad, I., Salama, A.S., Daradkeh, Y.I., Javeed, D., Ur Rehman, A., Hamam, H.: A hybrid deep learning-based approach for brain tumor classification. *Electronics* **11**, 1146 (2022)
8. Alanazi, M.F., et al.: Brain tumor/mass classification framework using magnetic-resonance-imaging-based isolated and developed transfer deep-learning model. *Sensors* **22**, 372 (2022)
9. Coudray, N., et al.: Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat. Med.* **24**, 1559–1567 (2018)
10. Othman, M.F., Basri, M.A.M.: Probabilistic Neural Network for brain tumor classification. In: *Proceedings of the 2011 Second International Conference on Intelligent Systems, Modelling and Simulation*, pp. 136–138. Phnom Penh, Cambodia (2011)
11. Deepak, S., Ameer, P.M.: Brain tumor classification using deep CNN features via transfer learning. *Comput. Biol. Med.* **111**, 103345 (2019)
12. Li, J.P., Kumar, R., Ali, Z., Khan, I., Uddin, M.I., Agbley, B.L.Y.: MCNN: a multi-level CNN model for the classification of brain tumors in IoT-healthcare system. *J. Ambient Intell. Humaniz. Comput.* **1**, 1–12 (2022)
13. Alanazi, M.F., et al.: Brain tumor/mass classification framework using magnetic-resonance-imaging-based isolated and developed transfer deep-learning model. *Sensors* **22**, 372 (2022)
14. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv* 2014, [arXiv:1409.1556](https://arxiv.org/abs/1409.1556). <http://arxiv.org/abs/1409.1556>. Accessed 13 July 2022
15. Saha, R., Debi T., Arefin, M.S.: Developing a framework for vehicle detection, tracking and classification in traffic video surveillance. In: Vasant, P., Zelinka, I., Weber, G.W. (eds.) *Intelligent Computing and Optimization. ICO 2020. Advances in Intelligent Systems and Computing*, vol. 1324. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-68154-8\\_31](https://doi.org/10.1007/978-3-030-68154-8_31)
16. Fatema, K., Ahmed, M.R., Arefin, M.S.: Developing a system for automatic detection of books. In: Chen, J.I.Z., Tavares, J.M.R.S., Iliyasu, A.M., Du, K.L. (eds.) *Second International Conference on Image Processing and Capsule Networks. ICIPCN 2021. Lecture Notes in Networks and Systems*, vol. 300. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-84760-9\\_27](https://doi.org/10.1007/978-3-030-84760-9_27)
17. Rahman, M., Laskar, M., Asif, S., Imam, O.T., Reza, A.W., Arefin, M.S.: Flower recognition using VGG16. In: Chen, J.I.Z., Tavares, J.M.R.S., Shi, F. (eds.) *Third International Conference on Image Processing and Capsule Networks. ICIPCN 2022. Lecture Notes in Networks and Systems*, vol. 514. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-12413-6\\_59](https://doi.org/10.1007/978-3-031-12413-6_59)
18. Yeasmin, S., Afrin, N., Saif, K., Imam, O.T., Reza, A.W., Arefin, M.S.: Image classification for identifying social gathering types. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture*

- Notes in Networks and Systems, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_10](https://doi.org/10.1007/978-3-031-19958-5_10)
19. Ahmed, F., et al.: Developing a classification CNN model to classify different types of fish. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_50](https://doi.org/10.1007/978-3-031-19958-5_50)
  20. Brain Tumor Dataset | Kaggle Brain Tumor Dataset. <https://www.kaggle.com/datasets/mniissahaann/braintumors>. Accessed 5 Oct 2022
  21. Brain Tumor Dataset | Kaggle Brain Tumor Dataset. [https://figshare.com/articles/dataset/brain\\_tumor\\_dataset/1512427?fbclid=IwAR3VaHmuktRYQKmRLMYpDVD2DUxFWQqMlm4fqe5voJuNidukZD6WzC-kf0U](https://figshare.com/articles/dataset/brain_tumor_dataset/1512427?fbclid=IwAR3VaHmuktRYQKmRLMYpDVD2DUxFWQqMlm4fqe5voJuNidukZD6WzC-kf0U). Accessed 5 Oct 2022



# Deciphering Handwritten Text: A Convolutional Neural Network Framework for Handwritten Character Recognition

Md Jakir Hossain<sup>1</sup>, Sarah Samiha Zaman<sup>1</sup>, Fardin Rahman Akash<sup>1</sup>, Farhana Alam<sup>1</sup>, Ahmed Wasif Reza<sup>1</sup> (✉), and Mohammad Shamsul Arefin<sup>2,3</sup> (✉)

<sup>1</sup> Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh  
wasif@ewubd.edu

<sup>2</sup> Department of Computer Science and Engineering, Daffodil International University,  
Dhaka 1341, Bangladesh  
sarefin@cuet.ac.bd

<sup>3</sup> Department of Computer Science and Engineering, Chittagong University of Engineering and  
Technology, Chattogram 4349, Bangladesh

**Abstract.** HCR (Handwritten Character Recognition) is considered one of the most challenging research areas, given the vast array of potential applications. Character recognition has been the focus of research since the beginning of Artificial Intelligence. Numerous studies, including HCR, have been conducted in this sector. A typical procedure requires two steps: feature extraction and Classification. Many forms of neural networks have been used in this cause over the years, with notable results. CNN has altered the scenario in recent years. It has had remarkable success in this industry due to its cutting-edge extraction of features and Classification. To produce recognized characters, CNN uses images for input and sends them through a sequence of layers, including a convolutional layer, a nonlinear function, a pooling layer, and interconnected layers. We utilized a dataset containing 372,450 handwritten character images covering the entire alphabet in English. We created a model using the CNN model and achieved 99% test accuracy. CNN is an efficient and powerful approach for HCR. Our model's high accuracy suggests that it has the potential to be applied in various practical scenarios such as postal address reading, digital libraries, and traffic sign detection.

**Keywords:** Handwritten character recognition · CNN · Layers · AI · Image

## 1 Introduction

(HCR) Handwritten Character Recognition is complex due to the variability in individual writing styles and the inherent noise in the writing process. However, with the advancement of ML techniques and algorithms, the performance of handwritten character recognition systems has dramatically improved. These systems are widely used in various applications such as signature verification, document analysis, and postal automation. The demand for handwritten character recognition systems is increasing in multiple industries, such as finance, healthcare, and education.

CNN has shown promising results in this area. A burst of work has been done in this area using CNN. Although multiple neural network methods are utilized for HDR, training and testing these models takes a long time due to many hidden layers. CNN has provided a wide range of alternatives for improving performance [1]. CNN is rapidly gaining popularity among machine learning and deep learning models because of its excellent classifier. CNN has recently gained popularity for detecting handwritten letters and photos since it employs fewer hidden layers than other neural network techniques [2].

Handwritten Character Recognition (HCR) takes input from paper documents or images, detects each character, and interprets the document as output. HCR is widely used in numerous research areas, such as bank check processing, digitizing handwritten documents to understand history, understanding handwritten journals, recognizing number plates, and checking postal codes [3]. Improved document processing: HCR can automatically digitize handwritten documents, such as historical archives or medical records, improving access to information and streamlining document processing. Enhanced accessibility: HCR can improve accessibility for people with disabilities, such as those with visual impairments, who may have difficulty reading handwritten text.

HCR is a challenging task requiring complex machine learning algorithms, which can lead to advancements in artificial intelligence and computer vision. The impact of research on HCR can be significant, as it has the potential to improve efficiency and accessibility in a variety of industries. For example, in the finance industry, HCR can automatically process checks and other financial documents, while in healthcare, it can improve the accuracy and speed of medical record keeping. Furthermore, advancements in HCR can have a broader impact on society by improving access to historical archives and other important documents.

## 2 Literature Review

Research on BHC (Bangla Handwritten Character Recognition) has been done by some researchers from KUET, Bangladesh. In this paper, the researchers have described the steps of handwritten character recognition. According to them, there are two significant steps to HCR. The first is to extract features from the character set and the second is to differentiate each character using the learning tools. From this paper, we understand that the standard features of different characters, overlapping, interconnections of Bangla characters, etc., make this problem more challenging. From the article, we can see that the accuracy of the test set of their HCR tool is around 84.20–85.36%. If the training part had been done more efficiently, the results of the test set might have improved [4].

Five researchers from Daffodil International University, Bangladesh, have developed a model called 'BornoNet,' which can recognize BHC using CNN. To build this model, they have used a mixture of three datasets: 'ISI,' 'CMATERdb,' and 'BanglaLekha-Isolated'. In this model, a 13-layer CNN has been used. Researchers claim that the proposed model has shown promising results, such as around 95.71–98% accuracy on test datasets. But from the report, it is noticeable that the accuracy of the result from the training set is also very high, which endangers the model from being over-fitted. This model's main attraction is the use of cross-validation among three datasets. Even

with a satisfactory accuracy level, this model has many more areas to improve, such as overlapping character recognition [5].

“HCR from Images using CNN-ECOC,” a work on handwritten character recognition, has been published in an international conference, International Conference. In this article, the main attraction is the combination of CNN with ECOC. The feature extraction part has been done with CNN, and the classification part has been covered using ECOC. The training and test Datasets consist of 17,500 and 1500 images of Bangla text. Using this method, other researchers have carried out handwritten character recognition in other languages. This model has recognized 50 individual Bangla characters with an 85.36% accuracy rate. In this article, it is said that the results of different experiments support the proposal of the accuracy and reliability to be improved using this hybridization of CNN with ECOC. All CNN architecture does not give the perfect accuracy with the ECOC approach. From the chart shown in the article, it is seen that only AlexNet and ZfNet shows high accuracy on the test result, but it is noticeable that even without the ECOC approach, these two CNN architectures offer similar kind of high accuracy [6].

In [7], the paper discusses using CNN for handwritten character recognition (HCR). The purpose of this research is to look into CNN’s capacity to detect characters from a picture collection, as well as the accuracy of recognition through training and testing. The dataset used for experimentation is NIST and the results show an accuracy of 92.91% on 200 images with a training set of 1000 images from NIST.

Rahaman et al. [8] suggested a model that CNN used to recognize Bangla Handwritten Characters and Digits. They use two datasets: CMATERdb and the BanglaLekha Isolated dataset, which together contain 118,698 photos (21000 images). Their suggested model has a mean classification accuracy of 97.43%.

Bora et al. [6] have proposed a system for BHCR using CNN with the method of inception module. The datasets are  $32 \times 32$  dimension images, and the train and test datasets consist of 50,000 and 5000 photos of Bangla text. This model uses one fully connected layer in the inception module. The accuracy of this model is 96.7%.

In [9], this research focuses on developing an Android application for character recognition using Optical Character Recognition (OCR) to translate handwritten documents into machine-encoded text. The proposed system uses preprocessing, segmentation, feature extraction, and post-processing to recognize characters from handwriting styles. The results show high accuracy, Precision, and recall, comparable to feature extraction-based schemes for HCR. In [10–14] authors applied different image processing techniques for performing different tasks.

### 3 Methodology

Figure 1 shows the Proposed System for HCR Framework Using Convolutional Neural Network. A proposed system for English HCR using a CNN would likely involve several key components:

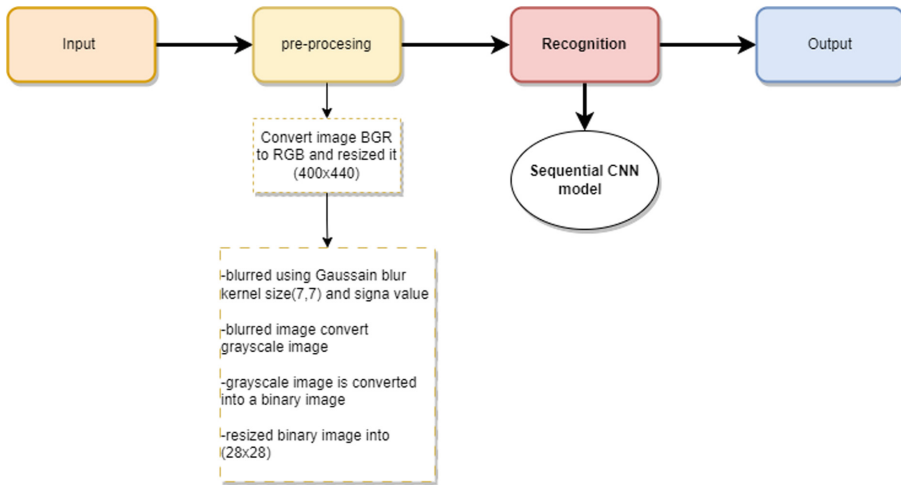
**Data collection.** A dataset of handwritten English characters would need to be collected and labeled for use in training the CNN.

**Preprocessing.** The data in the dataset would need to be preprocessed to standardize their size and format and make them appropriate for input to a CNN.

**CNN architecture.** A CNN architecture would be designed and implemented to recognize the handwritten characters. This could include multiple convolutional and pooling layers, and Classification requires ultimately linked layers.

**Training.** To decrease the error rate on the labeled data, CNN would be trained on the dataset using a supervised learning technique.

It's crucial to note that this is a broad concept, and the specifics of this system will change based on the quantity of the dataset and the complexity of the CNN architecture utilized.



**Fig. 1.** Proposed system for English HCR framework using CNN

### 3.1 Dataset Description

We have worked on a dataset that contains 26 folders and has 372,450 images of handwritten characters, the English alphabet from ‘A’ to ‘Z.’ Initially, the images were presented as 784 columns of pixel data in the CSV file. So, we convert it to  $28 \times 28$  pixels. Every alphabet in the dataset is center fitted to a  $20 \times 20$ -pixel box. We have collected the dataset from a website called “Kaggle” [15]. After collecting the dataset, we categorized it. In this dataset, the model is trained to recognize whether a given image contains a specific character. For example, a binary HCR dataset could include images of one handwritten letter or digit [16]. The model is trained to classify each image as either containing the letter/digit or not. For the training dataset, we have used 20% samples. Figure 2 shows the example of data used for analysis.



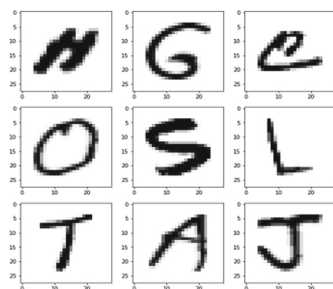


Fig. 2. Examples of data used for analysis/training.

## 4 Implementation

This section thoroughly explains the research’s implementation approach, experimental setting, and model performance evaluation. It describes the stages and methods utilized to perform the study and assesses its efficacy.

### 4.1 Device Set-Up

**Device.** Experiments for the project were carried out using a Windows-based machine outfitted with an Intel Core i5-8250U CPU, a clock speed of 1.6–1.8 GHz, 8 GB RAM, and a 64-bit x64-based architecture.

### 4.2 Implementation

In this research, we have used CNN models for English HCR. We have used the Sequential CNN model to evaluate the performance of these models and compare their accuracy.

**Sequential CNN Model.** Multiple convolutional layers interconnected build the CNN, equivalent to the complicated many hidden layers of other neural network techniques. CNN receives pictures as input, processes them using convolutional layers, nonlinear functions, pooling, and connected layers, and outputs a recognized character. Convolution layers are often composed of max-pool layers that minimize the number of retrieved features. Then—the output of the max-pool and convolution layers are flattened into a single-dimensional vector and are given as an input to the fully connected network layer. Figure 3 shows the architecture of our Sequential CNN model implemented on the dataset. We used hyperparameters in this CNN model (Learning rate, number of epochs, kernel size). The learning rate controls the step size of the optimizer during training, and a value of 0.001 is a common starting point. The kernel size of (3, 3) specifies a  $3 \times 3$  filter for the convolutional layers in the CNN, which is a common choice for small image inputs. The ReLU activation function is also commonly used in CNNs and effectively improves training performance. Finally, the number of epochs specifies the number of times the training data will be iterated during training, and five epochs is a relatively small number, which may lead to underfitting if the task is complex or the dataset is

large. However, this is just a general observation, and the optimal number of epochs can vary depending on the job and dataset.

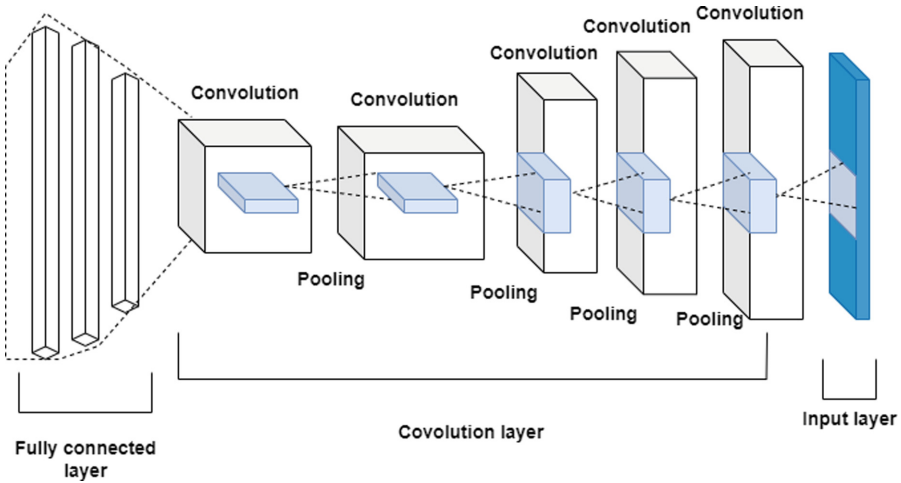


Fig. 3. Sequential CNN model

## 5 Experimental Result

Table 1 shows the evaluation metrics of a Sequential CNN model for handwritten character recognition (HCR). The model is evaluated on both training and validation datasets.

Table 1. Evaluation and interpretation of experimental outcomes

Models	Accuracy of training (%)	Loss of training (%)	Accuracy of validation (%)	Loss of validation
Sequential CNN	<b>99</b>	<b>0.025</b>	99	2.00e – 04

**Accuracy of Training.** For the training dataset, training accuracy assesses how effectively the model predicts the proper output. Accuracy is the proportion of correctly classified instances out of the total number of cases in the dataset, shown in Eq. (1) [17]. The model achieved an accuracy of 99% on the training dataset, indicating that it performed very well on the training data.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

**Table 2.** Evaluation metrics precision, recall, and F1 score.

Precision (%)	Recall (%)	F1 score (%)
99.27	99.27	99.27

Table 2 shows the Evaluation metrics using Precision, recall, and F1 score. The Precision, recall, and F1 score of 99.27% indicates that the model correctly classified many instances as positive or negative, with fewer false positives and false negatives. The precision metric represents the ratio of true positives to the total number of cases that were predicted as positive. The recall metric represents the ratio of true positives to the total number of positive samples. The F1 score is the harmonic mean of Precision and recall. In this case, the high scores for all three metrics indicate that the model is performing well and achieving a good balance between Precision and memory, resulting in an increased overall accuracy score.

**Loss of Training.** The loss of training measures how well the model fits the training data. A lower loss indicates a better fit. In this case, the training loss is 0.025, which is relatively low, meaning that the model fits the training data well.

**Accuracy of Validation.** The accuracy of validation is a measure of how well the model generalizes to new, unseen data [18]. The model achieved an accuracy of 99% on the validation dataset, which is a good sign that the model is performing well on new data.

**Loss of Validation.** The loss of validation metric determines how effectively the model generalizes to new data. Validation loss measures the difference between predicted and actual values for the validation set during machine learning model training. It is calculated using a loss function, which measures the difference between the predicted and actual values for each instance in the validation set [19]. The most used loss functions for CNNs are cross-entropy loss and mean squared error (MSE) loss. The loss function is defined in Eq. (2):

$$L(y, \hat{y}) = - \sum_i y_i \log(\hat{y}_i) \quad (2)$$

In Eq. (2),  $y$  is the actual label,  $\hat{y}$  is the predicted label, and  $I$  range over the number of classes. The validation loss is then calculated as the average of the loss function over all instances in the validation set.

A lower loss indicates a better fit. In this case, the validation loss is  $2.00 \times 10^4$ , which is very low, meaning that the model is generalizing well to new data.

In Fig. 4, we have shown training accuracy, validation accuracy, Training loss, and Validation Loss versus epoch number. Overall, these evaluation metrics suggest that the Sequential CNN model performs very well on the task of HCR, both in terms of fitting the training data and generalizing it to new data. However, it is essential to note that the model's performance may vary on different datasets and tasks, and further evaluation and testing would be required to confirm the model's effectiveness.

Table 3 compares all the models used in other research mentioned in this paper.



**Fig. 4.** Epoch versus accuracy and loss

**Table 3.** Comparative analysis with previously used techniques

S. No.	Author	Model	Accuracy (%)
<b>1</b>	<b>Proposed framework</b>	<b>Sequential CNN</b>	<b>99</b>
2	Rahman [3]	CNN	84.20
3	Azad Rabby [4]	CNN	95.71
4	Bora [5]	CNN	85.36

Even though the model used in [4] has shown 95.71% accuracy, the other two papers have also used CNN models, but the accuracy they have achieved is 84.20% and 85.36%, respectively. The proposed framework, in this case, achieved the highest reported accuracy of 99% on both the training and validation datasets, which suggests that the model can learn the features of the handwritten characters in the dataset well and generalize to new data. The other three models reported lower accuracies, ranging from 84.20 to 95.71%, suggesting that these models may have needed help to learn the dataset’s features or generalize to new data to some extent. However, it is essential to note that the comparison of models should not be based solely on reported accuracy, as different models may have been trained and evaluated on different datasets or under different conditions. Therefore, it is essential to consider other factors, such as the dataset’s size,

the task's complexity, the computational resources required to train the model, and more, to make a fair and comprehensive comparison of different models for HCR.

## 6 Conclusion

The variety of writing styles and inherent noise in the writing process make recognizing handwritten characters challenging. However, the effectiveness of handwritten character recognition systems has significantly increased because of the development of ML techniques and DL algorithms. We have studied different research papers regarding handwriting recognition and found that many articles have used CNN and have significant results. However, we have proposed a model using structural CNN that has resulted in 99% test accuracy, which is the most effective until now. We plan to work in this field more and offer more features. This project uses convolutional neural networks to recognize handwritten characters. Its potential impact in the industry is improving handwritten text recognition efficiency and accuracy. The future of such projects looks promising as technology advances and more applications arise for HCR in various fields, such as banking, healthcare, and education.

## References

1. Ahlawat, S., Choudhary, A., Nayyar, A., et al.: Improved handwritten digit recognition using convolutional neural networks (CNN). **20**, 3344 (2020)
2. Ali, S., Shaukat, Z., Azeem, M., Sakhawat, Z., Mahmood, T., Rehman, K.: An efficient and improved scheme for handwritten digit recognition based on convolutional neural network. *SN Appl. Sci.* **1**, (9) (2019). <https://doi.org/10.1007/s42452-019-1161-5>
3. Cruz, R.M.O., Cavalcanti, G.D.C., Ren, T.I.: An ensemble classifier for offline cursive character recognition using multiple features extraction technique. *IJCNN*, 1–8 (2010)
4. Rahman, M.M., Akhand, M.A.H., Islam, S., Chandra Shill, P., Hafizur Rahman, M.M.: Bangla handwritten character recognition using convolutional neural network. *Int. J. Image Graph. Signal Process.* **7**(8), 42–49 (2015). <https://doi.org/10.5815/ijigsp.2015.08.05>
5. Azad Rabby, A.K.M.S., Haque, S., Islam, M.S., Abujar, S., Hossain, S.A.: Bornonet: Bangla handwritten characters recognition using convolutional neural network. *Procedia Comput. Sci.* **143**, 528–535 (2018). <https://doi.org/10.1016/j.procs.2018.10.426>
6. Bora, M.B., Daimary, D., Amitab, K., Kandar, D.: Handwritten character recognition from images using CNN-ECOC. *Procedia Comput. Sci.* **167**, 2403–2409 (2020). <https://doi.org/10.1016/j.procs.2020.03.293>
7. Khandokar, I., Munirul, H.M., Ernawan, F., Islam, M., Kabir, M.N.: Handwritten character recognition using convolutional neural network. *J. Phys. Conf. Ser.* **1918**, 042152 (2021). <https://doi.org/10.1088/1742-6596/1918/4/042152>
8. Rahaman, M.A., Mahin, M., Ali, M.H., Hasanuzzaman, M.: BHCDR: real-time bangla handwritten characters and digits recognition using adopted convolutional neural network. In: 1st International Conference on Advances in Science, Engineering and Robotics Technology 2019, ICASERT 2019, May 2019. <https://doi.org/10.1109/ICASERT.2019.8934476>
9. Patel, D.K., Som, T., Singh, M.K.: Improving the recognition of handwritten characters using neural network through multiresolution technique and euclidean distance metric. *Int. J. Comput. Appl.* **45**(6) (2012)

10. Saha, R., Debi T., Arefin, M.S.: Developing a framework for vehicle detection, tracking and classification in traffic video surveillance. In: Vasant, P., Zelinka, I., Weber, G.W. (eds.) *Intelligent Computing and Optimization. ICO 2020. Advances in Intelligent Systems and Computing*, vol. 1324. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-68154-8\\_31](https://doi.org/10.1007/978-3-030-68154-8_31)
11. Fatema, K., Ahmed, M.R., Arefin, M.S.: Developing a system for automatic detection of books. In: Chen, J.I.Z., Tavares, J.M.R.S., Iliyasu, A.M., Du, K.L. (eds.) *Second International Conference on Image Processing and Capsule Networks. ICIPCN 2021. Lecture Notes in Networks and Systems*, vol. 300. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-84760-9\\_27](https://doi.org/10.1007/978-3-030-84760-9_27)
12. Rahman, M., Laskar, M., Asif, S., Imam, O.T., Reza, A.W., Arefin, M.S.: Flower recognition using VGG16. In: Chen, J.I.Z., Tavares, J.M.R.S., Shi, F. (eds.) *Third International Conference on Image Processing and Capsule Networks. ICIPCN 2022. Lecture Notes in Networks and Systems*, vol. 514. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-12413-6\\_59](https://doi.org/10.1007/978-3-031-12413-6_59)
13. Yeasmin, S., Afrin, N., Saif, K., Imam, O.T., Reza, A.W., Arefin, M.S.: Image classification for identifying social gathering types. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_10](https://doi.org/10.1007/978-3-031-19958-5_10)
14. Ahmed, F., et al.: Developing a classification CNN model to classify different types of fish. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_50](https://doi.org/10.1007/978-3-031-19958-5_50)
15. Grother, P.J.: A-Z Handwritten Alphabets in .csv format. Kaggle. <https://doi.org/10.18434/T4H01C>
16. Chauhan, R., Ghanshala, K.K., Joshi, R.C.: Convolutional neural network (CNN) for image detection and recognition. In: *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, Jalandhar, India, 2018, pp. 278–282. <https://doi.org/10.1109/ICSCCC.2018.8703316>
17. Kelleher, J.D., Tierney, B., Tierney, B.: *Data Science: An Introduction*. CRC Press (2018)
18. Choudhary, A., Rishi, R., Ahlawat, S.: *Off-Line Handwritten Character Recognition using Features Extracted from Binarization Technique*. 2212-6716 © 2013. American Applied Science Research Institute. <https://doi.org/10.1016/j.aasri.2013.10.045>
19. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. MIT Press (2016)



# Applying Machine Learning Techniques to Forecast Demand in a South African Fast-Moving Consumer Goods Company

Martin Chanza<sup>1,2</sup>, Louise De Koker<sup>1,2</sup>, Sasha Boucher<sup>1,2</sup>, Elias Munapo<sup>1,2</sup>✉, and Gugulethu Mabuza<sup>1,2</sup>

<sup>1</sup> Department of Business Statistics and Operations Research, North-West University, Potchefstroom, South Africa

Martin.chanza@nwu.ac.za, {Louise.DeKoker, sasha.boucher}@mandela.ac.za, emunapo@gmail.com, gugulethu.mabuza@abbott.com

<sup>2</sup> Business School, Nelson Mandela University, Gqeberha, South Africa

**Abstract.** Inventory planning is a critical function in FMCG companies, and forecasting is important to determine future demand accurately. Demand forecasting helps FMCG companies anticipate short- and medium-term risks. However, the real-time benefits of machine learning techniques within demand forecasting within South African FMCG companies are under-explored. The main goal of this study was to compare the forecasting ability of statistical forecasting methods and machine learning models in demand forecasting in the FMCG sector. A case study approach was followed, using sales data from a specific category in a selected FMCG company for 2014–2019. Moving Average (MA), Seasonal Autoregressive Integrated Moving Average (SARIMA) and Artificial Neural Network (ANN) models were used in the data analysis. The findings revealed that the ANN model is more accurate in predicting demand than MA and SARIMA models. The Supplements category has consistently grown in the number of units sold and sales for the six-year period. The sales data for supplements showed a steady rising trend in accordance with price inflation. There is an increasing demand in baby products in the supplements category. Machine learning models are superior in predicting demand in the FMCG sector. For further research, we recommend using the Auto-Regressive Integrated Moving Average (ARIMAX) model for modeling demand when multivariate data is present.

**Keywords:** Machine learning · SARIMA · Demand forecasting · FMCG

## 1 Introduction

Organizations that produce fast-moving consumer goods (FMCG) in South Africa find themselves in an ever-increasing competitive environment [1]. This requires FMCG organizations to effectively perform inventory planning. Inventory planning is an important part of accurate forecasting in a fast-moving consumer goods (FMCG) industry. Effective inventory planning requires future demand to be forecast appropriately [1].

This study is focused on an FMCG company in South Africa which, for the purposes of this study, was referred to as Company C. Company C is listed in the top 40 JSE companies and is one of the largest manufacturers of non-perishable foods and complementary medicines in South Africa. Company C has forecasting challenges such as high inventory storage costs and frequent stock shortages in its medicinal portfolio. The study was focused on the complementary medicine range in the baby category. Complementary medicine was chosen because the products in this department have high-profit margins with low volumes and contribute 16% to the baby category's income. Hence, the importance of accurate forecasting in this department.

The FMCG industry presents a number of issues for supply chain managers, including real-time inventory visibility, unforeseen market risks, and a lack of forecast accuracy [2]. Studies show that in emerging markets, such as South Africa, demand forecasting is a relatively new task in pharmaceutical companies, and this has a direct impact on the Complementary and Alternative Medicines demand forecasting accuracy [2]. Demand forecasting is an important activity in FMCG organizations in managing their business.

The main goal of this study was to compare the forecasting ability of statistical forecasting methods and machine learning models in demand forecasting in the FMCG sector. This will assist the sector in having the appropriate methodologies ready for implementation once the IT infrastructure has been set up and the necessary technology and software have been acquired. This paper discusses the data analysis and findings from the data collected from Company C, between 2014 and 2019.

## 2 Literature Review

Demand planning, an operational supply chain management technique, is a part of demand forecasting. Good demand planning helps ensure that inventory levels are correctly aligned with expected demand by using historical data and acts as a strategic resource [3]. This results in an organization meeting customer expectations, reducing lead times, and managing limited resources insofar as it avoids overproduction and stock redundancy [1].

Despite the benefits of demand forecasting, there are demand-influencing factors that need to be considered by decision-makers [4]. In particular, well-coordinated marketing and supply chain functions, as well as environmental factors, such as competitors, strikes, economic turns, and external factors, should be considered as it influences control actions [5].

Against this background, an organization's supply chain does not operate in isolation. Organizations in the upstream of the supply chain struggle with demand information asymmetry, which influences supply chain efficiency [6]. Therefore, sophisticated demand forecasting methods are useful to the entire supply chain because it enhances the organizational deliverables and all the stakeholders in the supply chain [1, 7]. This means that selecting the most efficient demand forecasting method is a key decision area for managers and a way to reduce the supply-demand discrepancy. In the context of this study, the interest lies in quantitative forecasting methods.

Quantitative forecasting predicts future demand based on historical data and uses numerical data collected over time (i.e., time series) whilst focusing on cyclical patterns



and data trends. Time series data can be based on a single observation with one variable or multiple observations with more than one variable [15]. In this line, Chase [8] explains that machine learning (ML) demand forecasting, compared to traditional methods, has revealed supply-demand efficiency. This means applying ML for demand forecasting can potentially promote supply chain efficiency [9, 10].

However, in the context of this study, we are interested in supervised learning and algorithms associated with demand forecasting [8]. Supervised learning, the most common form of ML, involves a machine that is given inputs of data with labeled data and thereby learns pattern recognition.

In particular, linear regression algorithms are commonly used for sales and demand forecasting: Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA). On the other hand, regression through Artificial Neural Networks (ANN) is modeled on neurons to simulate the human brain, which performs non-linear processes based on inputs toward forecasting [11].

### 3 Methodology and Data

The methodology consists of four main phases: exploratory data analysis (EDA), the modeling process, which comprises (demand modeling, moving averages, SARIMA and artificial neural networks), model accuracy and lastly forecasting, as shown in Fig. 1.

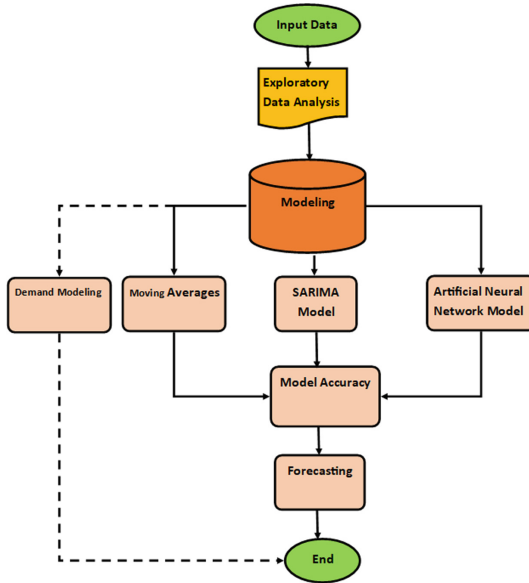
This study employed a quantitative modeling approach to analyze the demand patterns of complementary medicine data using statistical and machine learning models to forecast demand. The overall monthly sales time series data spanning the period 2014–2019 was used. The company has four categories in the complementary medicinal department. Each type represents the different brands of medicine used to treat minor diseases in the baby category. The categories of complementary medicine are: Minor Ailments, Supplements, Stomach and Teething types. The data were analyzed using the R software.

#### 3.1 Exploratory Data Analysis

Exploratory data analysis is the crucial process of doing preliminary analyses on data to discover patterns, and the distributions in the data, find relationships, detect outliers and draw insights using summary statistics and graphical representations [12]. Visual plots such as time series plots, seasonal plots, histograms, density plots, boxplots and QQ plots were used to explore the data. The data was tested for normality and transformed using the logarithmic transformation.

#### 3.2 Modeling

This section describes the statistical and machine learning models used in demand modeling.



**Fig. 1.** The proposed methodology for demand Forecasting

**3.2.1 Moving Averages**

A statistical method called the Simple Moving Average is used to prepare a smoothed version of the original dataset to forecast [11]. One of the most widely used methods for pre-processing time series is the use of moving averages. Creating a new series whose values are made up of the mean of the raw data in the initial time series is necessary to calculate a moving average [13]. The formula can represent a moving average [14]:

$$\hat{T}_t = \frac{1}{m} \sum_{j=-k}^k y_{t-j}, \text{ where } m = 2k + 1$$

**3.2.2 Seasonal Autoregressive Integrated Moving Average (SARIMA) Process**

SARIMA models are ARIMA models with a seasonal component. A seasonal ARIMA or SARIMA model can be represented as:

$$SARIMA(p, d, q) \times (P, D, Q)_S$$

with p = non-seasonal AR order, d = non-seasonal differencing, q = the non-seasonal MA order, P = seasonal AR order, D = seasonal differencing, Q = is the seasonal MA order and S is the seasonal length.

Mathematically, the SARIMA model can be represented as:

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_S^D y_t = \delta + \theta_q(B)\Theta_Q(B^S)\varepsilon_t$$

where:

$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  is the non-seasonal AR component.

$\Phi_P(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots$  is the seasonal AR component.  
 $\Phi_P B^{PS}$

$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$  is the non-seasonal MA component.

$\Theta_Q(B^S) = 1 + \Theta_1 B^S + \Theta_2 B^{2S} + \dots + \Theta_Q B^{QS}$  is the seasonal MA component.

$\nabla^d = (1 - B)^d$  is the non-seasonal difference operator.

$\nabla_S^D = (1 - B^S)^D$  is the seasonal difference operator.

The iterative model-building strategy known as the Box-Jenkins methodology [15] is used to build SARIMA models. The Box-Jenkins methodology consists of four steps namely model identification, estimation, diagnostic checking, and forecasting.

### 3.2.3 Neural Network Forecasting

A feed-forward neural network is a type of artificial neural network in which there is no cycle in the connections between the nodes. Just the data moves forward from the input nodes, via the hidden nodes, and to the output nodes. There are no loops in the network. Each layer of nodes receives inputs from the previous layers [14]. The neural network is built using the function `nnetar` in the R package. The fitted model is of the form  $\text{NNAR}(p,k)$  where  $p$  stands for number of lagged inputs and  $k$  the number of nodes in the hidden layer.

## 3.3 Accuracy Measures

Regarding the accuracy metrics for the evaluation of the prediction performance of the two models, four metrics were used, namely: Mean Error (ME), Mean Absolute Error (MAE), Root Mean Squared Error (Root Mean Squared Error) RMSE) and Mean Absolute Percentage Error (MAPE). They can be represented using the following formulas:

- i. Mean Error

$$ME = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)$$

- ii. Mean Absolute Error

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- iii. iii. Root Mean Squared Error (RMSE): the square root of MSE.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

iv. Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{y_i} \times 100$$

$$n$$

where  $y_i$  is the actual and  $\hat{y}_i$  is the predicted and  $n$  the sample size. The model with the smallest accuracy metric is chosen as the best model.

### 4 Empirical Results

This section discusses the data analysis and the results from the study. Firstly the, exploratory data analysis (EDA) is conducted. The next step is demand modeling, followed by moving averages, SARIMA model and modeling using Artificial Neural networks. The last section concludes by comparing the accuracy of models and using the best model to forecast demand. The R statistical package was used for the data analysis.

#### 4.1 Exploratory Data Analysis

The exploratory data analysis provides an understanding of the patterns, trends and distribution of the CAM data over the period 2014–2019. Figure 2 shows the plot of overall sales in ZAR and the seasonal plot of the sales data.

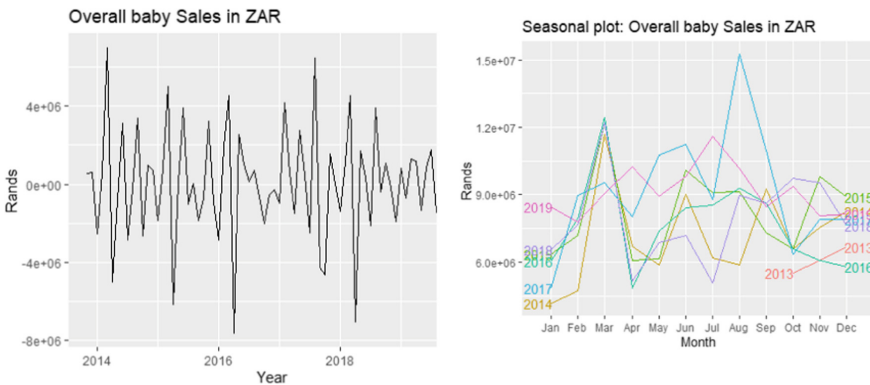


Fig. 2. Overall sales and seasonal plot of overall sales

The plot of the actual data shows high volatility of the sales within each year and a strong seasonal pattern from year to year. However, the seasonal plot displays the seasonal variations within each year. The months of March and August recorded the highest sales. Normality tests showed that the sales data is non-normal. The data is skewed to the right.

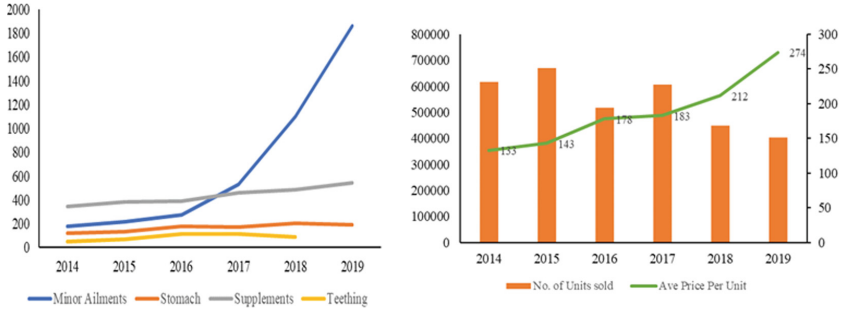


Fig. 3. Category prices, units sold, and average prices

## 4.2 Demand Modeling

Sales data for the four categories show a great demand for stomach medications. Planning and forecasting for demand require careful consideration of pricing. Figure 3 displays the category pricing, units sold, and average prices over time.

The minor ailments category had a significant increase compared to the other categories. Sales contribute to the gross profit and can easily indicate the business performance better than the other two measures. The number of units sold decreased between 2017 and 2019. The decrease in the units within this department was because of the changes in the medicinal basket of items over the period [17]. The changes in the medicine's basket were due to discontinuing one of the categories within this department. Factors such as out-of-stock items for prolonged periods contributed to the decrease in the overall number of units sold in this category.

## 4.3 Moving Averages

A 13-month moving average was calculated for the individual years which showed an increasing trend. The true underlying behavior of the series can be seen more clearly with the use of a moving average. Among the most widely used methods for time series pre-processing are moving averages. Creating a new series whose values are made up of the average of the original time series' raw observations is necessary to calculate a moving average [13]. The 13-month moving average curve showed fewer fluctuations than the original data.

## 4.4 SARIMA Model Results

In the identification stage, tentative modes were identified using the ACF and PACF to identify competing models. Models with both non-seasonal and seasonal differencing were tested. The AIC criterion was used in selecting the best model. The model with the smallest AIC was selected as the best model. The best model is  $ARIMA(2, 1, 0) \times (2, 0, 0)_{12}$  with an AIC value of 2359.623. The fitted model is given by the formula:

$$(1 - \Phi_1 B^s - \Phi_2 B^{2s})(1 - \phi_1 B - \phi_2 B^2)(1 - B)^d x_t = \varepsilon_t, \text{ where } \varepsilon_t \sim WN(0, \sigma^2).$$

The model can be represented as:

$$(1 - 0.2104B^{12} - 0.3125B^{24})(1 + 0.5718B + 0.5169B^2)(1 - B)^d x_t = \varepsilon_t.$$

The next step was to conduct the diagnostic checking of the residuals. Figure 4 displays the analysis of residuals. The residual autocorrelations are very small and are generally within significance bounds. The histogram of the residuals of the overall sales model, shows that the residuals are normally distributed. This means the fitted model is suitable for forecasting. Finally, it is evident that our fitted model  $ARIMA(2, 1, 0) \times (2, 0, 0)_{12}$  given all of the diagnostic-checking results is the best fitted model to forecasting Overall demand.

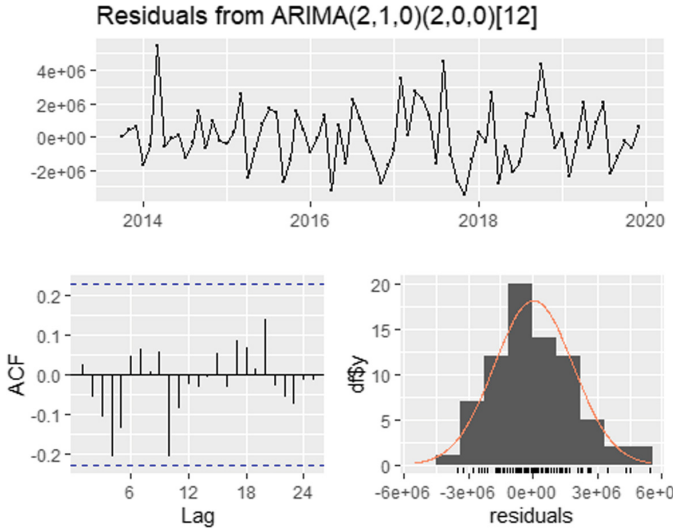


Fig. 4. Diagnostic checking

### 4.5 Forecasting Using ANN Model

We now discuss the results of the artificial neural network. A Feed-forward neural network with a single hidden layer and lagged inputs are used for forecasting overall sales. Table 1 displays the ANN model results. The fitted model is the  $NNAR(4, 1, 3)_{12}$  model as shown in Table 1.

### 4.6 Model Accuracy

Accuracy results for seasonal ARIMA or SARIMA model and ANN model are presented in Table 2. The mean error (ME), mean absolute error (MAE), the root mean square error (RMSE) and the mean absolute percentage error (MAPE) are presented. All the accuracy metrics for ANN are smaller than those for SARIMA, hence ANN is the best model in forecasting demand of baby products.

**Table 1.** ANN model results

```

Series: datafts
Model: NNAR(4,1,3)[12]
Call: nnetar(y = datafts)

Average of 20 networks, each of which is
a 5-3-1 network with 22 weights
options were - linear output units

sigma^2 estimated as 8.645e+11

```

**Table 2.** Accuracy measures

Model	ME	MAE	RMSE	MAPE
SARIMA	39357	1423226	1819954	17.87627
ANN	678	726782	920343	9.299668

## 5 Conclusions

The ANN model  $NNAR(4, 1, 3)_{12}$  shows the best performance in demand forecasting in the FMCG sector. The Supplements category has consistently grown in the number of units sold and sales for the six-year period. In demand planning and forecasting the analysis of product sales and price could provide insights into positioning products in a business. For further research, we recommend the use of Auto-Regressive Integrated Moving Average ARIMAX model for modeling demand when multivariate data is present. ARIMAX model is very flexible as it works with stationary and non-stationary data and with any pattern. Economic variables, marketing variables and product sales could be considered in demand forecasting in the FMCG sector.

## References

1. Basson, L.M., Kilbourn, P.J., Walters, J.: Forecast accuracy in demand planning: a fast-moving consumer goods case study. *J. Transp. Supply Chain Manag.* **13** (2019). <https://doi.org/10.4102/jtscm.v13i0.427>
2. Altay, N., Litteral, L.A., Rudisill, F.: Effects of correlation on intermittent demand forecasting and stock control. *Int. J. Prod. Econ.* **135**(1), 275–283 (2012). <https://doi.org/10.1016/j.ijpe.2011.08.002>
3. Wang, X., Petropoulos, F.: To select or to combine? The inventory performance of model and expert forecasts. *Int. J. Prod. Res.* **54**(17), 5271–5282 (2016). <https://doi.org/10.1080/00207543.2016.1167983>
4. Hofmann, E., Rutschmann, E.: Big data analytics and demand forecasting in supply chains: a conceptual analysis. *Int. J. Logist. Manag.* **29**(2), 739–766 (2018). <https://doi.org/10.1108/IJLM-04-2017-0088>

5. Deshmukh, A.K., Mohan, A.: Demand chain management: the marketing and supply chain interface redefined. *IUP J. Supply Chain Manag.* **13**(1), 20–36 (2016)
6. Feizabadi, J.: Machine learning demand forecasting and supply chain performance. *Int. J. Logist. Res. Appl.* **25**(2), 119–142 (2022). <https://doi.org/10.1080/13675567.2020.1803246>
7. Hu, M., Li, H., Song, H., Li, X., Law, R.: Tourism demand forecasting using tourist-generated online review data. *Tour. Manag.* **90**, 104490 (2022). <https://doi.org/10.1016/j.tourman.2022.104490>
8. Chase Charles, W. :Machine learning is changing demand forecasting. *J. Bus. Forecast.* **35**(4), 43–45 (2017)
9. Aamer, A., Eka Yani, L., Alan Priyatna, I.: Data analytics in the supply chain management: review of machine learning applications in demand forecasting. *Oper. Supply Chain Manag. Int. J.* **14**(1), 1–13 (2020). <https://doi.org/10.31387/oscm0440281>
10. Dikshit, A., Pradhan, B., Santosh, M.: Artificial neural networks in drought prediction in the 21st century—a scientometric analysis. *Appl. Soft Comput.* **114**, 108080 (2022). <https://doi.org/10.1016/j.asoc.2021.108080>
11. Sajid, M.J.: Machine learned artificial neural networks vs linear regression: a case of chinese carbon emissions. *IOP Conf. Ser. Earth Environ. Sci.* **495**(1), 012044 (2020). <https://doi.org/10.1088/1755-1315/495/1/012044>
12. Chatfield, C.: Exploratory data analysis. *Eur. J. Oper. Res.* **23**(1), 5–13 (1986). [https://doi.org/10.1016/0377-2217\(86\)90209-2](https://doi.org/10.1016/0377-2217(86)90209-2)
13. Ivanovski, Z., Milenkovski, A., Narasnov, Z.: Time series forecasting using a moving average model for extrapolation of number of tourist. *UTMS J. Econ.* **9**(2), 121–132 (2018)
14. Hyndman, R.J., Athanasopoulos, G.: *Forecasting: principles and practice*. OTexts (2018)
15. Box, G.E.P., Jenkins, G.M.: *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco (1976)
16. Quiza, R., Davim, J.: *Computational methods and optimization*. In: *Machining of Hard Materials*. Springer, London, UK (2011). Accessed March 03, 2023 [online]. Available [https://doi.org/10.1007/978-1-84996-450-0\\_6](https://doi.org/10.1007/978-1-84996-450-0_6)
17. Gjika Dharmo, E., Puka, L., Zaçaj, O.: Forecasting consumer price index (CPI) using time series models and multi regression models (Albania case study). In: *10th International Scientific Conference “Business and Management 2018”*, Vilnius Gediminas Technical University, Lithuania, Sept 2018. <https://doi.org/10.3846/bm.2018.51>





# A Review on Machine Learning Algorithms for Cost Estimation in Construction Projects

Vijay Kumar<sup>1</sup>, Sandeep Singla<sup>1</sup>(✉), and Aarti Bansal<sup>2</sup>

<sup>1</sup> Department of Civil Engineering, RIMT University, Mandi Gobindgarh, Punjab, India  
dr.sandeepsinglaz@gmail.com

<sup>2</sup> Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India  
aartibansal@chitkara.edu.in

**Abstract.** Any construction project's success hinges on an accurate estimate of the project's construction costs. A challenging task in the early stages of construction projects is estimating the costs due to the lack of detailed design and documentation. For cost estimation in the literature, machine learning techniques have proven successful. In this paper, we have studied and analyzed the various machine learning algorithms deployed for cost estimation in construction projects. Out of various machine learning algorithms, artificial neural network (ANN) is the most preferred algorithm for it. As a result of our research and analysis, we have outlined the open research issues.

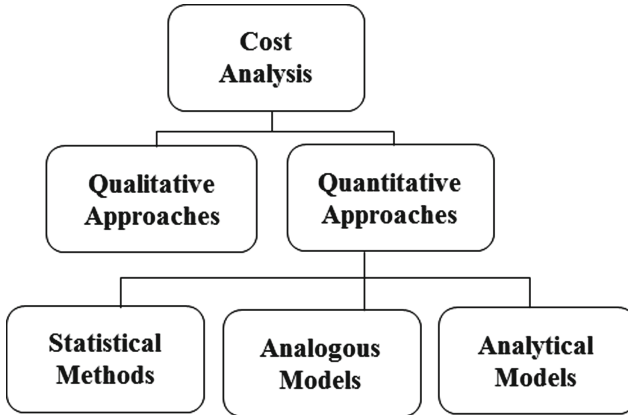
**Keywords:** Artificial intelligence · Cost estimation · Construction project · Data augmentation · Machine learning

## 1 Introduction

To be successful, a construction project must adhere to three interdependent constraints: budget, schedule, and quality. During the construction phase, a project's quality can be evaluated and improved, but the budget and schedule must adhere to their contractually outlined estimates. With these estimates Wang et al. [1], the contractors as well as stakeholders are deeply involved. Pre-feasibility studies can be more logical if stakeholders and decision-makers have accurate cost estimates. They can also monitor and control cash flows even during the construction phase of projects if they have accurate cost estimates. Stakeholders and/or contractors often lose money when the cost of a project is overestimated. Many methods for effectively estimating construction costs have already been tried and tested to help projects stay on budget and make a profit, both in practice and in the literature.

### 1.1 Cost Estimation Modelling Techniques

Qualitative as well as quantitative approaches to cost estimation will be discussed in more detail later on in this article. Figure 1 depicts a comprehensive view of preliminary cost modelling methodologies [2].



**Fig. 1** Comprehensive view of preliminary cost modeling methodologies [2]

**Qualitative Analysis:** The estimator's familiarity with the project's parameters and other factors informs qualitative procedures like expert judgement and heuristic rules. Expertise is based on the results of decisions made in the past, and these decisions can have good or bad results. To verify estimation accuracy, experts and peers use the expert judgement technique, which is outlined in Rad [3]. Experiential professionals who can professionally guarantee the dependability of estimates can use this technique, which is based on intuition and depends heavily on unwritten but poorly documented extrapolation procedures. When it comes to cost estimation, heuristic rules come from instinctive judgements and serve as a general rule to speed up the estimation process.

**Quantitative Analysis:** When it comes to project cost estimation, quantitative methods rely on data collection and analysis as well as the use of quantitative models, techniques, including tools. Statistical, similar, and analytical techniques to quantitative cost estimates fall into three broad types.

- **Statistical Methods in Cost Estimation:** If you're looking for an explanation for the association between the final cost and its relevant attributes, statistical methods use formulas or even other possible alternatives to do so. For example, mass, volume, and price are all considered in parametric cost estimate methods [4]. In fact, the project's cost is assessed by determining the nature of the relationship between these factors, with the result being a mathematical operation on the variables involved. When starting a new project, this tactic is useful because there is usually not much information available [5], but it lacks the basic justification of results required [4]. According to Duverlie and Castelain [6], there are three forms of parametric cost estimation:

1. **The method of scales:** This approach can be used with currently available technologies to create simple items of various sizes. This method's pre-requisite is an assessment of the most influential technical parameters. It is then compared to the results of completed projects, resulting in a hybrid approach that utilises both analogous and parametric methodologies. This technique implies a linear relationship

between cost and characteristics of interest, which Duverlie and Castelain [6] say is its fundamental flaw.

2. Statistical models: Mathematical formulas can be built using this strategy since the actions are broken down into key categories. There are three basic forms of data in this model [6] technical characteristics, connections between data or final variables, and constants.
  3. Cost estimation formulae (CEF): Using a restricted set of technical factors, the CEF calculates the final cost [6]. Physical values are the main parameter categories. Function and dimensional values are used to determine the functional description. In accordance with the information provided in the solution, regression and optimization approaches are the most likely parametric methods. The following are some of the disadvantages of parametric cost estimation methods. When using these approaches, various findings are the only issues, with no understanding of where they came from. But a lack of crucial parameters in the early phases leads to confusion about the results. Additionally, the developer ought to be aware of the effects each parameter has on the total cost. CEFs, in particular, are unable to deal with unique situations. In some circumstances, it is necessary to conduct regression analysis on four or five previous instances to get the most accurate cost estimate. Despite these drawbacks, they are nonetheless considered useful cost estimating techniques due to their speed of execution [6].
- Analogous models: When a similar historical case is utilised as the basis for a new model, it is called an analogous model. The similarity between the cost structures is attributable to functional and geometrical homogeneity [5] which claims that comparable methods are the most straightforward way to estimate via. Using a database of previously completed projects as a guide, the cost of a given project can be projected. So, project managers must take into account as many variables as possible in their estimation process. This procedure is simple and accurate, although it is an approximate estimate [3].
  - Analytical models: An alternative technique is to use analytical models, which estimate costs by precisely describing the cost associated with each processing phase feature and then utilising a bottom-up strategy to aggregate the project overall cost [7].

## 1.2 Machine Learning for Cost Estimation

Early cost estimating is challenging due to insufficient designs and documentation. Machine learning can solve this problem. Machine learning outperforms human-made rules for data-driven projects in accuracy, automation, speed, customization, and scalability, according to Yoonseok Shin [8].

### 1.3 Main Contribution

This research analyzes construction cost estimation machine learning algorithms. Following that, define the open research challenges based on the study to assist the other authors in contributing their work in this field. We found that key points.

- The weight and bias parameter values play an important role in the machine learning (ML) algorithms. Thus, finding the optimal values of these parameter enhances the performance of the machine learning algorithm for cost estimation.
- Some of the ML algorithms performance is superior for small dataset but degrades when dataset is huge. Therefore, need to explore machine learning algorithms that provides superior performance for huge dataset because in the construction projects, the dataset is large.
- Most machine learning algorithms utilize the same dataset for training and testing, but to verify model resilience, the dataset should be random.
- The hybridization of machine learning algorithm for cost estimation model enhances the performance of the model.

### 1.4 Paper Organization

The paper organization is as follows. Section 2 defines the related work of the cost estimation model is based on machine learning algorithm. Based on the existing study, the open research challenges are defined in Sect. 3. Finally, conclusion is defined in Sect. 4.

## 2 Related Work

In the present cost estimation model of the construction projects, machine learning algorithms are deployed. In this section, we have studied and analyzed the existing methods.

Wang et al. [1], This study used the model and data from 98 public school projects in the Hong Kong Special Administrative Region to test out Shapley Additive Explanations (SHAP). In addition, a comparison of many common machine learning algorithms in building cost estimates is used to verify the findings. Economics has an essential role in the reduction of construction cost estimating mistakes and is even more significant than the features of the projects. The results will enable stakeholders in the area of management and construction engineering to make the right choices and reveal the true extent of the impacts of other important elements on the estimated cost of construction. Shin [8], This work investigates if the boosting strategy can be used to solve the regression issue of early building cost estimate using a BRT (boost regression tree). For the BRT used in this work, Friedman presented the stochastic gradient boosting tree module. It is an entirely new method of data mining. It uses a stochastic gradient boosting technique that expands and enhances the regression tree. Thus, it has the benefits of both a boosting technique and a regression tree, such as high conceptual simplicity, interpretability, computing efficiency, etc. The boosting strategy may be particularly useful when paired with other data mining approaches, such as a neural network, a support vector machine (SVM),

and a decision tree. According to the newest developments in fusion of computational intelligence approaches, this characteristic is an excellent fit. Elhegazy et al. [9], Their technique suggests that when there is a significant quantity of data available for the purpose of model training, their ANN approach is able to provide cost estimates relatively quickly and correctly. In addition, after the model has been trained, applying it to real-world problems does not call for an in-depth understanding of the ANN mechanism that lies behind the surface. In situations like these, all the cost estimator has to do to provide an effective outcome of the cost to assemble the structural components is to input the fundamental attributes of the structure and the associated unit cost of raw materials. During the phase of building projects referred to as value engineering, it is anticipated that these speedy and precise cost projections would be tremendously advantageous. As a result, estimators are given the opportunity to analyze the numerous design options and find the optimal balance between the cost, efficiency, and dependability of the engineering structures. Matel et al. [10], the purpose of this study is to provide an artificial neural network method for the task of cost estimating engineering services. In order to build the model, they first identified the relevant elements affecting engineering service prices. After that, a model is built using the information obtained from 132 different initiatives. After that, a heuristic strategy was devised to enhance and fine-tune the model's performance. With minimal samples, artificial neural networks (ANNs) are eventually shown to be able to get a somewhat accurate price estimate, even with minimal samples. In addition, the model presented in this research outperformed models from comparable studies. When taking MAPE into account, the accuracy of the model created in this research improved by 14.5%. Sanni-Anibire et al. [11], This study shows how machine learning can be used to develop a model for estimating major construction project costs. MLRA, ANN, SVMs, KNN, and MCS. Twelve models were compared using the same measures. The best multi-classifier model with KNN as the combined classifier had R2 of 0.81, MAPE of 80.95 percent, and RMSE of 6.09. Using new digital technologies like machine learning, this study demonstrated the potential of the construction sector to address challenges. The method proposed in this paper is very useful in the construction of early cost estimating models for both research and practice. The model that was made could help people make decisions about big building projects when they are still figuring out how much they will cost. Mir et al. [12], Artificial neural network (ANN)-based uncertainty estimation is proposed as a solution for cost estimation in this study. In order to teach an ANN to create intervals on its own, researchers use the optimum lower upper bound estimation (also known as LUBE). Asphalt and steel prices in the United States are predicted using the suggested technique. In order to accurately anticipate the price of raw materials, single-point estimates derived from standard regression analysis and artificial neural network models are shown to be inadequate. On the other hand, using prediction intervals may help you avoid project failure due to inaccurate preliminary cost estimates by giving them a more accurate view of what materials will cost. The accuracy of the model is tested by comparing the findings of three alternative cost functions to the suggested optimum LUBE cost function. Results demonstrate that the suggested LUBE cost function provides the highest accurate prediction intervals. In this work, a stacking approach is used to keep tabs on the training and validation processes. For project managers, the interval forecasting technique is an innovative approach to cost prediction

studies that will provide them with fresh insights into the potential pitfalls of project budgets. Al-Tawal et al. [13], This research investigated artificial neural network (ANN) approaches in the early stages of the building design process. More than 100 projects built in Jordan over the last five years have been utilised to train and evaluate neural network models. 53 design elements were initially used to build the first ANN model at the architectural design stage before being lowered to 41 for the second model at this level. The third ANN model made use of all 27 design parameters that were accessible during the concept design stage. The models achieved an average cost estimation accuracy of 98%, 98%, and 97% during the detailed, schematic, and concept design phases, respectively. In this study, they implement a simple linear regression model, the Ordinary Least Square (OLS) method, as proposed by Sphurti and Dharwadkar [14]. It is possible to utilise the Ordinary Least Square approach to get the optimal answer, and it works well with tiny datasets. For the purpose of testing the model's accuracy, data from the Pune area of India for the last 12 years was used. It has been shown that the suggested model can accurately forecast 91–97% of the time. Mohammed Arafa and Mamoun Alqedra [15], The goal of this study was to find a way to accurately and efficiently estimate construction project costs in the planning stages using artificial neural networks. The construction industry in the Gaza Strip has contributed 71 projects to a database of ongoing construction initiatives. During the pre-design phase of the project, there are many key factors that may be acquired from accessible engineering drawings and data. These values were chosen in order to calculate a good estimate for the project's basic framework. The input layer of the ANN model includes the square footage of the ground floor, the number of stories, the average floor area, the number of columns, the number of elevators, the kind of foundation, and the number of rooms. The artificial neural network model that was constructed had one hidden layer with 7 neurons. The output layer of the ANN model was comprised of a single neuron that was meant to represent an early cost estimate of construction. The findings that were acquired from the trained models suggested that neural networks had a reasonable amount of success in forecasting the initial stage estimation of building costs by making use of fundamental information regarding the projects and doing so without the requirement of a more detailed design. An early estimate of construction costs may be influenced by the ground floor space, how many floors there are, and how many elevators there are in a building, according to a sensitivity study.

### 3 Open Research Challenges

In this section, we have defined the open research challenges that are determined from the existing study.

- In the literature, artificial neural network (ANN) is the most preferred machine learning algorithm for cost estimation [9–13, 15–17]. Weight and bias values play an important role in ANN. The determination of the optimal value of these parameters enhances the network performance. The optimal parameter values can be determined using swarm intelligence algorithms or fuzzy logic. In the market, a huge number of swarm intelligence algorithms are available. These algorithms are differentiated from each other based on the number of parameters, exploration, and exploitation

rate. So, we must deploy those swarm intelligence algorithm that requires minimum parameter and provides better exploration and exploitation rate.

- In the literature, Sanni-Anibire et al. [11] deployed the support vector machine (SVM) algorithm for cost estimation. The SVM algorithm is superior for small databases, but for large databases its performance degrades. In the construction project, the database is large. Thus, SVM is not a suitable algorithm for construction projects.
- In the machine learning algorithm, the same dataset is split into training and testing data. After training the machine learning model, the same dataset is tested. Therefore, a number of authors achieved superior results, but in the real scenario, random data should be tested to check the robustness of the model. Data augmentation methods can be deployed to achieve this goal. In the data augmentation, the new dataset can be generated using numerous approaches such as addition, multiplication, averaging, or subtracting.
- In the existing cost estimation models, single machine learning algorithm is taken under consideration. However, the hybridization of the machine learning algorithms enhances the performance of the cost estimation models.

## 4 Conclusion

Cost estimation plays an important role in construction projects. However, in a real scenario, cost estimation is a difficult task in the early stages due to inefficient data. In the literature, both qualitative and quantitative methods are employed for cost estimation. In terms of quantitative methods, machine learning is the most preferred field. As a result, we have investigated and analyzed the many machine learning techniques now used to predict construction project costs. The other writers can then use our stated open research challenges to improve their own machine learning models. Besides that, we have defined how these challenges can be resolve in the existing models.

## References

1. Wang, R., Asghari, V., Cheung, C.M., Hsu, S.C., Lee, C.J.: Assessing effects of economic factors on construction cost estimation using deep neural networks. *Autom. Constr.* **134**, 104080 (2022)
2. Tayefeh Hashemi, S., Ebadati, O.M., Kaur, H.: Cost estimation and prediction in construction projects: a systematic review on machine learning techniques. *SN Appl. Sci.* **2**(10), 1–27 (2020). <https://doi.org/10.1007/s42452-020-03497-1>
3. Rad, P.F.: *Project Estimating and Cost Management*. Berrett-Koehler Publishers (2001)
4. Qian, L., Ben-Arieh, D.: Parametric cost estimation based on activity-based costing: a case study for design and development of rotational parts. *Int. J. Prod. Econ.* **113**(2), 805–818 (2008)
5. Hegazy, T., Ayed, A.: Neural network model for parametric cost estimation of highway projects. *J. Constr. Eng. Manag.* **124**(3), 210–218 (1998)
6. Duverlie, P., Castelain, J.M.: Cost estimation during design step: parametric method versus case based reasoning method. *Int. J. Adv. Manuf. Technol.* **15**(12), 895–906 (1999)
7. Caputo, A.C., Pelagagge, P.M.: Parametric and neural methods for cost estimation of process vessels. *Int. J. Prod. Econ.* **112**(2), 934–954 (2008)

8. Shin, Y.: Application of boosting regression trees to preliminary cost estimation in building construction projects. *Comput. Intell. Neurosci.* (2015)
9. Elhegazy, H., et al.: Artificial intelligence for developing accurate preliminary cost estimates for composite flooring systems of multi-storey buildings. *J. Asian Archit. Build. Eng.* **21**(1), 120–132 (2022)
10. Matel, E., Vahdatikhaki, F., Hosseinyalamdary, S., Evers, T., Voordijk, H.: An artificial neural network approach for cost estimation of engineering services. *Int. J. Constr. Manag.* **22**(7), 1274–1287 (2022)
11. Sanni-Anibire, M.O., Mohamad Zin, R., Olatunji, S.O.: Developing a preliminary cost estimation model for tall buildings based on machine learning. *Int. J. Manage. Sci. Eng. Manage.* **16**(2), 134–142 (2021)
12. Mir, M., Kabir, H.D., Nasirzadeh, F., Khosravi, A.: Neural network-based interval forecasting of construction material prices. *J. Build. Eng.* **39**, 102288 (2021)
13. Al-Tawal, D.R., Arafah, M., Sweis, G.J.: A model utilizing the artificial neural network in cost estimation of construction projects in Jordan. *Eng. Constr. Archit. Manage.* (2020)
14. Arage, S.S., Dharwadkar, N.V.: Cost estimation of civil construction projects using machine learning paradigm. In: 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 594–599 (2017)
15. Arafa, M., Alqedra, M.: Early stage cost estimation of buildings construction projects using artificial neural networks. *J. Artif. Intell.* **4**(1) (2011)
16. Dubey, A., Gupta, U., Jain, S.: Medical data clustering and classification using TLBO and machine learning algorithms. *Comput. Mater. Continua* **70**(3), 4523–4543 (2021)
17. Singh, S., Ramkumar, K.R., Kukkar, A.: Machine learning techniques and implementation of different ML algorithms. In: 2021 2nd Global Conference for Advancement in Technology (GCAT), Bangalore, India, 1–6 (2021)





# A Computer Assisted Detection Framework of Kidney Diseases Based on CNN Model

Tanjina Akter Ripa<sup>1</sup>, Nafis Faiyaz<sup>1</sup>, Mahmud Hassan<sup>1</sup>, Rehnuma Naher Sumona<sup>1</sup>,  
Mohammed Sharafullah Anem<sup>1</sup>, Ahmed Wasif Reza<sup>1</sup>(✉),  
and Mohammad Shamsul Arefin<sup>2</sup>(✉)

<sup>1</sup> Department of Computer Science and Engineering, East West University, Dhaka 1212, Bangladesh

wasif@ewubd.edu

<sup>2</sup> Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chattogram 4349, Bangladesh

sarefin@cuet.ac.bd

**Abstract.** Kidney Tumors (KT) are the tenth most prevalent tumor in both men and women globally. Using image processing techniques, we provide several approaches and models in this research to identify renal illnesses from the dataset. In building our model, we employed deep learning CNN (Convolutional Neural Network), Keras, VGG16, SVM, U-NET, and Water-Shed. As an illustration, we found 100% accuracy in the Keras model and 50% accuracy in the VGG16 model. To develop these models, we initially only employed four classes: normal, cyst, stone, and tumor. Evaluating our model against the input photos gives accurate results to determine what stage of kidney illness it is in. We have tried our model multiple times to ensure it can reliably identify the condition from a single photograph. If our model is refined, it may be used to help the medical community diagnose kidney illnesses and administer the appropriate care.

**Keywords:** Kidney disease · CNN · Keras · VGG16 · Deep learning · Image processing

## 1 Introduction

Each aspect of human life is being improved by technology as time passes. One of the fields where innovation has demonstrated its most significant promise is the medical field. Any disease must first be well understood in order to be treated. Any disease must be fully understood in its current state. Certain diseases may cause in the human body without appropriate treatment, which cannot be given without definitive evaluation. In order to examine one or both kidneys and to identify potentially dangerous disorders such as tumors, stones, cysts, congenital disabilities, hepatorenal syndrome, fluid buildup around the kidneys, and the presence of abscesses, CT scans of the kidneys are helpful. These kidney illnesses can be identified by CT scan, which facilitates accurate diagnosis and effective treatment.

This study will use ML and DIP ideas to identify certain kidney diseases. The following dataset provides four signifiers: a healthy kidney, a kidney with a stone diagnosis, and kidneys that have been found to have tumors and otherwise cysts. The purpose of this paper is to clarify the signs mentioned above. Each of these issues will be displayed, and these issues will finally be found. Each component will be treated independently and broken down into several segments. The collected CT scan images will act as the inputs. One of the following criteria will be present in the outputs. The detection techniques and difficulties found will be detailed in our study with appropriate visual demonstrations. The issue must first be identified. There are numerous algorithms for detecting objects like CNN, ANN, SVM, and others.

This paper uses CNN (Convolutional Neural Network) because it is very good at identifying patterns and objects. Numerous methods are used by CNN, including Keras, VGG16, AlexNet, ResNet50, etc. VGG16 and Keras have been used to categorize the substance and determine whether it is a stone, cyst, or tumor. U-Net and SVM have both been applied to image classification. The images will be split into many sections using U-Net. In order to acquire a close-up and better picture of the issue, a grayscale image of image pixels is obtained using the Watershed technique.

This project aims to identify whether the kidney is affected by any disease or not based on CT-Radiography. Identification of renal illnesses is one of the paper's goals. Describe the renal condition in detail by using images correctly pinpointing the issue. To obtain a suitable image, eliminate noise and poor contrast. To make visuals more straightforward and more precise, utilize filters. The disorders can be recognized from a good photograph.

The rest of the paper is organized as follows: Sect. 1 comprises the introduction of this research paper and provides the aim and objective, and Sect. 2 briefly summarizes the works' background. Section 3 briefly reviews existing or previous works related to detecting kidney disease. Section 4 explains the planned methodology, dataset, methodological approach, proposed models, and pseudo code. Section 5 has the result, and performance evaluation of this research paper, and finally, in Sect. 6, our research paper concludes with limitations and suggestions for future work.

## 2 Background

This section focuses on the background context of disease identification through dataset images.

This paper sheds insight into the development of deep learning and how it has been applied to image recognition and analysis. Work on deep learning applications has expanded significantly, specifically regarding the auto-diagnosis of radiological scans and diagnostics. EfficientNet, ResNet, Inception, and exception systems have been popular in this categorization task, which also employs a method for learning to migrate across time. The post-model is the first step for the designated purposes in the transfer learning method, which is a deep learning strategy. It refers to applying a learned knowledge model to address a new issue. Recently, transformer concepts—which are widely used for language comprehension—have been applied to visual identification tasks, where they beat rival models in classification techniques.

The categorization of renal illness makes extensive use of deep learning techniques. Renal ultra-ultrasound pictures, as is standard practice, are enlarged using procedures like average, Gaussian, and morphological analysis. Following this, image attributes are recovered using Primary Component Analysis (PCA) and sometimes the K-nearest neighbor (KNN) classifier. Only CT scan pictures were employed in a few methods for dual-class categorization. We created a database of kidney stones, cysts, and tumor CT pictures in light of the dearth of available information and the conclusions of the study publications. The “CT KIDNEY DATASET: Regular, Cyst, Tumor, and Stone” dataset was created and annotated. We then designed six models and examined each one to decide which would work best for real-time usage.

Medical diagnostics cannot be completed without the analysis of radiological or tissue samples, as well as medical pictures, like chronic diseases detection images, but a considerable improvement in the diagnostic process might come from the categorization and fragmentation of images using deep learning techniques, primarily convolutional neural networks [1]. Convolutional neural networks (CNN) are frequently used to extract picture features and recognize various objects and detect kidney diseases. Object identification has been employed in a wide area of medical imaging since its specific effect on finding illnesses of all types [2]. Since end-stage renal illness and patient mortality are the results of inappropriate diagnostics and treatment of chronic diseases, machine learning (ML) methods have taken on a significant contribution to disease prediction and have become an effective instrument in the sector of medical science [3].

### 3 Related Work

In this part, we concentrate on the current and related work of renal disease identification from the dataset photographs. According to numerous scientific researchers, various classification and detection approaches can be utilized in identifying kidney disease.

In [4], the authors intended to convey that kidney stone disease can be a serious, globally prevalent health concern. Paleolithic period diseases go undetected, but they harm the kidneys as they progress. By utilizing image classification algorithms, areas of concern and relevant features are identified. The method of choice for finding kidney damage is contrast-enhanced CT scans. The segmentation of the kidneys via CT images has been proposed utilizing various automated techniques and provided a commonly misunderstood kidney identification that fits into this category. The level-set concept in this study incorporates limits on ellipsoid shape. This method was developed based on CT images without being aware of the contrasting phases beforehand.

In [5], the authors sought to argue that the most enlightening research, predicated on 109 CT scans using algorithms for machine learning, including HOG, MMD, LBP, and SURF, had the most significant prediction performance of 95%. Intelligent analysis procedures attempt to work more effectively in less time and lighten the burden of the radiologist. The lack of data constituted one of the difficulties researchers encountered; typically, few data points are available for medical imaging, which raises the danger of overexertion. The research limits are also impacted by the fact that several studies have been conducted utilizing the same data set.

In [6], S. Kalannagari Viswanath and Ramalingam Gunasundari said that the intended prediction had been that system would be tested using various kidney pictures from the

database and that the findings would correctly classify the various types of stones with an effectiveness of 98.8%. With relatively little resource usage, the device's responsiveness is very impressive. In a further study, medical and biological detectors will be placed in the abdominal area to capture kidney parts, and the system will be created for fast and accurate deployment.

In [7], the authors advised using complex convolutional neural networks using YOLO architectural designs, which divide input images into grids to categorize patients. A linked vector for architectural designs indicates if any requested objects are present in the grids. The architecture's operating principle is to keep the entity with the best prediction performance out of all the objects identified in that area while discarding the others (Gothane 2007).

In [8], the authors attempted to determine how reliability, susceptibility or recalls, specificity, or PPV metrics are used to generate the quantitative assessment of each of the six models. The reliability, specificity, and susceptibility are determined using the true positive (TP), false positive (FP), true negative (TN), as well as false negative (FN) samples.

In [9], the authors said renal cysts, kidney stones, and hydronephrosis are the three most typical kidney abnormalities. They utilize consolidated all descriptions and concentrate on the disease of the kidney, which in this data set includes a variety of diseases but is caused by a modest quantity of renal stones and hydronephrosis (lesion). To recognize those main aims, they applied various data enhancement techniques, such as flipping, reducing, scaling, and trimming the kidney in photographs.

In [10], the authors sought to conclude that a methodology based on deep learning techniques and ROIs supplied by radiologists demonstrated impressive outcomes in identifying invasive carcinoma subtypes. They are hopeful that it will aid in further study of the topic and enable us to work with radiologists to identify the classification of a patient's disease in actual clinical settings.

In [11], the authors proposed creating a CNN model to facilitate the multi-class classification identification of computerized renal tissue slices stained with periodic acid-Schiff (PAS). They discovered that glomeruli, tubules, and interstitial fluid could be accurately identified using CNN. In calculating the proportion of estimated glomerulosclerosis, the study showed that CNNs performed better than pathologists.

In [12], Gilbert Moeckel and Benjamin Wu asserted that the paper was found. An effective way to make slide scanning more accurate and quick requires an upfront investment in retraining the AI and giving data. While algorithms like neural networks and decision trees are examples of deep classification, conventional ML techniques like clustering algorithms and K-means clustering are categorized as unsupervised learning.

In [13], the authors stated in the article that we found that Up to 48 h before the onset of KDIGO phase two or phase three, a CNN can forecast AKI, with AUROC performance outperforming XGBoost classifiers and the SOFA score system. By using frequently recorded characteristics in the HER, it is possible to automatically screen a large patient's needs for potential AKI without specialist testing.

In [14], Vasanthselvakumar R., Balasubramanian M., and Palanivel S. noted in their work that automatic identification and classification of CKDs has been carried out utilizing both traditional and deep learning techniques. Convolutional neural network technology is employed for the categorization of the diseases. To achieve higher detection rates in the future, traditional detection schemes may replace current ones.

In [15], the authors claimed that they could find the presented method employs a standard CKD dataset from the University of California, Irvine (UCI), as well as an ensemble DL classifier for the detection of CKD. In an expansion of SMOTE, outliers are removed using ADASYN. For detecting preterm deliveries and evaluating visual health, ADASYN has been demonstrated to be helpful in applications involving medical imaging. In contrast to simple training situations, it generates additional synthetic instances for data from minority classes.

In [16], the authors reported that the article they found Taiwan's National Health Insurance started the Integrated Care of CKD (ICC) initiative in 2002, while China Medical University Hospital (CMUH) joined this program in 2003. Whenever 12 weeks or over occasionally, blood ammonium nitrogen and serum chlorine levels—two molecular indicators of kidney injury—were assessed. From the two GE models with two sizes—960,720 for the LOG IQ E9 and 820 614 for the LOGIQ P3—37,696 photos were chosen.

In [17], the authors asserted that they could find biomedical technology and clinical procedures have been greatly influenced by recent advancements in machine automation. Kidney stone identification utilizing CT scans and automated kidney partition has received little attention. The inner kidney's anatomy will eventually be annotated and segmented using improved CT imaging.

In [18], Dong-Hyun Kim and Soo-Young Ye explained that chronic kidney disease (CKD) is treatable if discovered early. However, renal replacement therapy, such as dialysis or transplantation, becomes essential as the condition worsens. Diagnoses for kidney cancer, autoimmune conditions, cystic diseases, chronic renal disease, etc., can be made using ultrasound technology. They utilized details such as kidney size and interior echo properties to assess the level of disease.

The works in [19–25] also focused on image analysis for performing different important tasks. The previous research papers were analyzed to create the suggested methodology. The papers mentioned above provide the most crucial information on CT scans utilized to pinpoint regions of concern and pertinent characteristics in kidney stone disease (CKD), an issue of worldwide health significance. Newer studies have employed machine learning algorithms like CNN(Keras), VGG16, etc. to predict CT scans with 95% accuracy. Our study employs algorithms like CNN(Keras), VGG16, etc. This study suggests a method for precisely identifying kidney diseases from visuals. However, we considered four different kidney image variants within the kidney diseases dataset on Kaggle.

## 4 Materials and Methods

### 4.1 Dataset

In our dataset, we have four types of data on the kidney. These are Cysts, Normal, Stone, and Tumors. We also have a numeric dataset named KidneyData. We used CNN (Keras), CNN (Vgg-16), Support Vector Machine (SVM), Watershed-Segmentation, and U-Net models for our image datasets. In order to avoid using the data augmentation approach for this investigation, we searched for a rather large dataset. We also need a dataset that includes normal kidney pictures, cyst, and their various kinds. We used two freely accessible Kaggle datasets [26] to solve this issue. Finally, a dataset of 12,446 CT pictures was acquired. There are 12,446 photos in total. There are 3709, 5077, 1377, and 2283 data points for Cyst, Normal, Stone, and Tumor, respectively. Due to the lack of severe issues with measurement errors in the dataset, we did not need to utilize different procedures to deal with these minor errors. Figure 1 shows some sample images from the dataset.

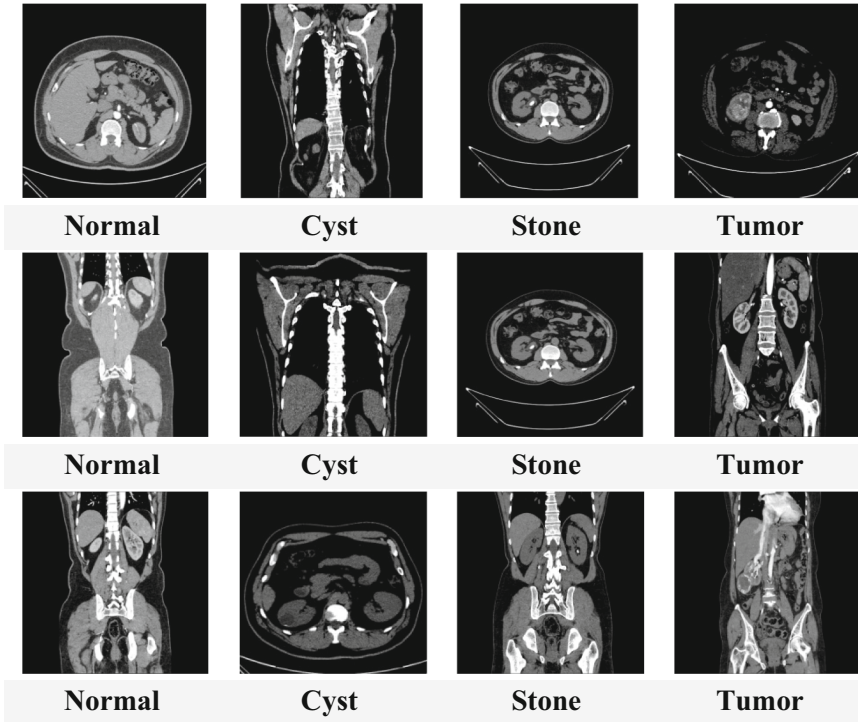


Fig. 1. Different samples from the dataset of kidney diseases (normal, cyst, stone, and tumor)

## 4.2 Methodological Approach and Proposed Models

We divided our whole dataset into two datasets as our dataset is vast. One of them is the Kidney folder which contains normal and stone files. On the other hand, another folder is the dataset which contains cyst and tumor files. We used these two datasets individually for CNN (Keras), CNN (VGG-16), and SVM.

Computer vision tasks, including object identification, picture recognition, and image classification, are well-suited for CNN. Recurrent neural networks (RNN), long short-term memory (LSTM), artificial neural networks (ANN), and others are also neural networks used for comparable tasks.

We demonstrated how to use VGG16 for image classification in this CNN model. Using transfer learning (VGG16), we will build and fine-tune a model. A pre-trained CNN model called VGG16 can increase image processing accuracy. The pre-trained weight will distinguish between normal and stone datasets and cyst and tumor datasets. We took different samples from our datasets by using the SVM model (Fig. 2). We performed data augmentation and applied the ReLU activation function. Also, we used dense layers and applied tanh and sigmoid activation functions, respectively, and got the model summary.

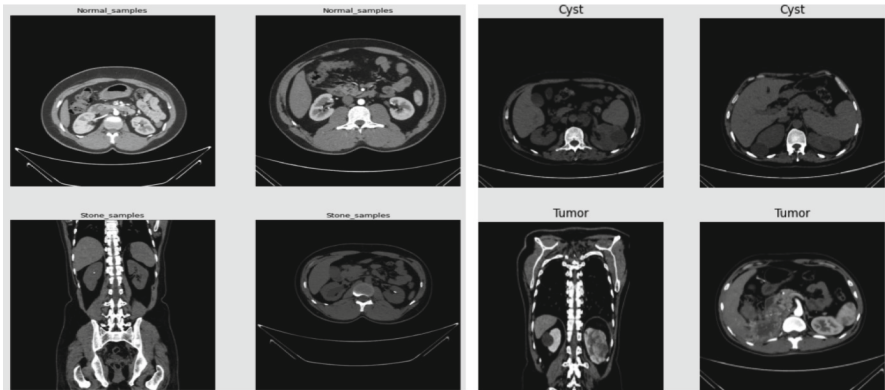


Fig. 2. Different samples from datasets by using the SVM model.

In this study, we performed U-Net and Water-Shed Segmentation on our whole dataset, including cyst, normal, stone, and tumor. Image segmentation is breaking an image into groups of pixels represented by a mask or labeled image. We can process only the crucial portions of a picture by segmenting it instead of processing the complete image using water-shed segmentation. U-net is also another segmentation algorithm that works better than CNN. The task of finding significant characteristics may be performed by CNNs without human supervision. They are particularly good at classifying and recognizing images. Using several images as an example, it may teach itself the essential characteristics of each class. The fact that the VGG16 network is a large network and requires more time to train its parameters is one of its major drawbacks. Based on these reasons we got good accuracy on CNN rather than VGG16.

### 4.3 Pseudo Code

<p><b>CNN(Keras)</b></p> <ol style="list-style-type: none"> <li>1. Load libraries</li> <li>2. Get categories</li> <li>3. Import tensorflow</li> <li>4. Model.add(MaxPooling())</li> <li>5. Model.add(Flatten())</li> <li>6. Model.add(Dense())</li> <li>7. Model.fit()</li> </ol> <p><b>CNN(VGG-16)</b></p> <ol style="list-style-type: none"> <li>1. Load libraries</li> <li>2. Get categories</li> <li>3. Train_datagen = tf.keras.preprocessing.image.Image(Data Generator)</li> <li>4. Subset = 'training' &amp; subset = 'validation'</li> <li>5. Vgg = VGG16()</li> <li>6. Folders = glob()</li> <li>7. Prediction = Dense()</li> </ol>	<p><b>SVM</b></p> <ol style="list-style-type: none"> <li>1. Path to train directory</li> <li>2. Get the list of all images</li> <li>3. Take all the normal cases and label it as zero</li> <li>4. Train_data = pd.DataFrame()</li> <li>5. Shuffle the data</li> <li>6. Cases_count = train data[]</li> <li>7. Sns.barplot()</li> <li>8. Get few samples for both cases</li> <li>9. Data augmentation(layer.Cnv2D() &amp; layers.Maxpooling2D())</li> <li>10. Model.data(layers.Dense())</li> </ol>
---	--

---

### U-Net and Water-Shed Segmentation

---

1. Load dataset
  2. Unique\_classes.append(path)
  3. Print indexes of classes
  4. Load images and mask path
  5. Preprocessing (Resize mask → resizeimage → normalize image → inverse of preprocessig → replace mask value with class 0 & 6 → load batch of data)
  7. Perform U-net
  7. Sure background area
  8. Sure foreground area
  9. Finding unknown region
  10. Change Figsize
  11. Plt.title(images[index])
  12. Plt.figure()
- 

## 5 Experimental Result

### 5.1 Result

In this paper, CNN was used for object detection, and the results on cysts and tumors were impressive. 100% accuracy was achieved using Keras. VGG16 was also used, but the results could have been better than the CNN results. Then SVM was used to detect Stones and Normal healthy kidneys. The results were also very satisfying. Case with



cysts was found a lot more than tumor cases. However, successes with multiple images have yet to be achieved. This works only on single images. We used U-Net and Watershed algorithms for image segmentation. U-Nets architecture was modified and extended to work with fewer training images and to yield more precise segmentations. Also, the Watershed algorithm was used for object segmentation purposes, such as separating different objects in an image. This allowed us to count the objects for further analysis of the separated objects. The outputs were generated with different views and angles. Furthermore, the detection results were very impressive.

## 5.2 Performance Evaluation

We can conclude from the previous arguments that CNN (Keras) makes it more effective in our system. In the validation phase, CNN (Keras) provides an accuracy of 100.0%, whereas the customized CNN model provides an accuracy of 99.9%. Due to validation loss, CNN (Keras) also provides us with less value. The customized CNN model offers excellent accuracy and fewer losses.

Following are the graphs of our customized CNN (Keras) model's accuracy and loss across 25 iterations (Fig. 3).

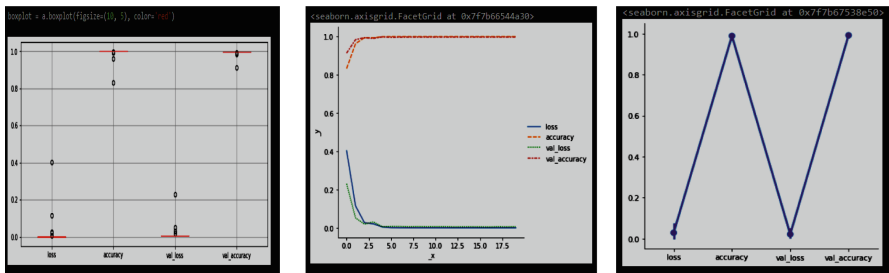


Fig. 3. Model accuracy and loss of CNN (Keras) model

We divided our data set into two parts and then used CNN (Keras) to detect cysts and tumors. It is a very successful object detection tool, and we got 100.0% accuracy and nearly 0.0% loss in detecting cysts and tumors. We did the same for rest two parts also.

Following are the graphs of our customized CNN (VGG16) model's accuracy and loss across 25 iterations (Fig. 4).

On just a set of identification-based, SVM is trained. The fundamental benefit of SVM is that it can be applied (Fig. 5) to problems involving regression analysis and classification. With the division of our dataset, we also obtained four case levels for SVM, including cases for tumors, cysts, stones, and normal instances.

For image segmentation, we used U-Net and Watershed algorithms, as shown above in Figs. 6 and 7. We obtained grayscale images and edge detection outputs of the input images. This helped us with further analysis, and the results were 100% accurate with Cysts, Tumor, Stones, and Normal cases. Table 1 shows the training and validation loss and accuracy of the models used.

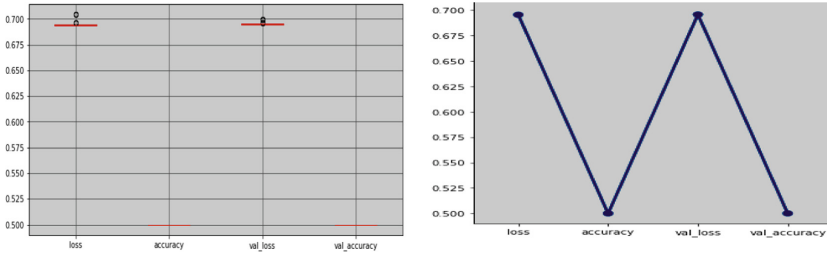


Fig. 4. Model accuracy and loss of CNN (VGG16) model

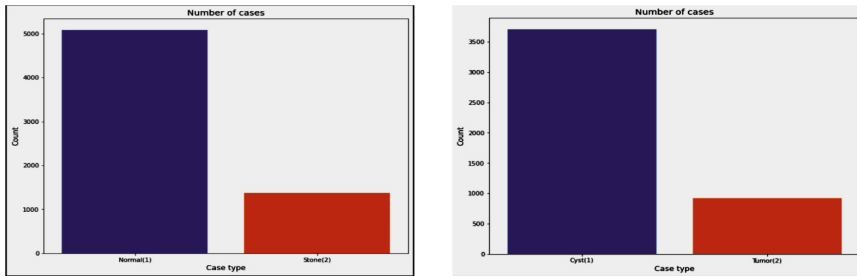


Fig. 5. SVM apply to the whole dataset

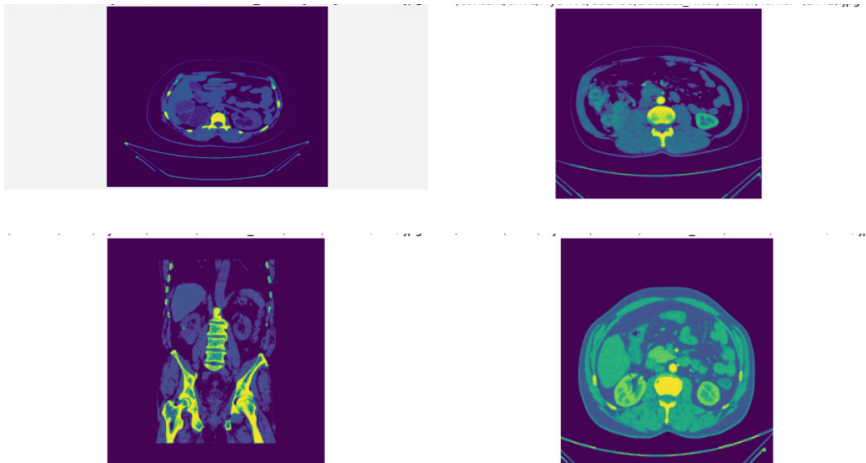
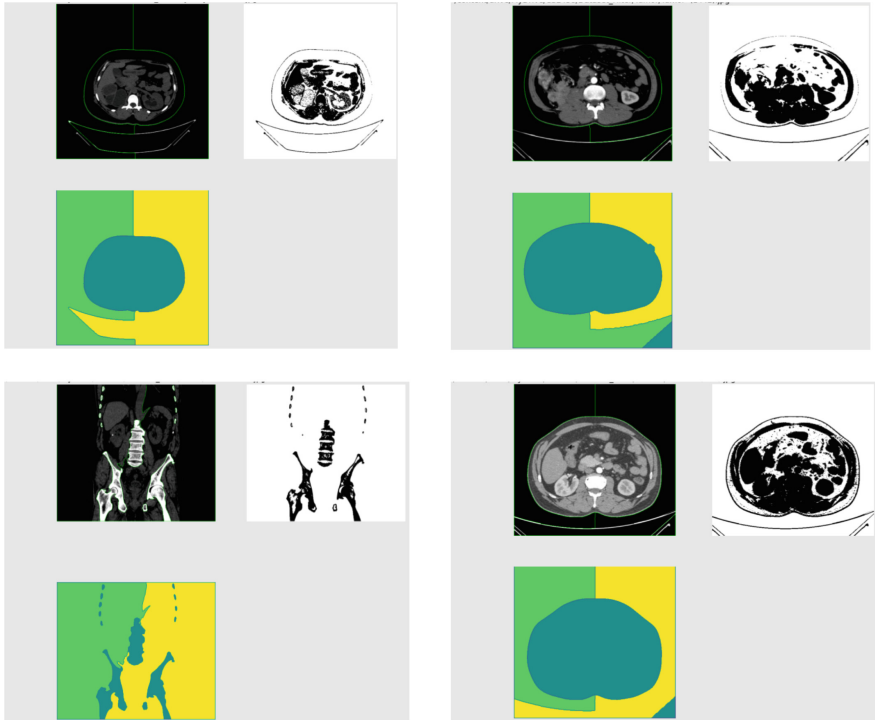


Fig. 6. U-net Segmentation to detect cyst, tumor, stone and normal case

We used CNN and VGG16 as detection algorithms. As we can see from the above table, better results were acquired using CNN. We can see that the CNN results are almost twice as better as VGG16. Recurrent neural networks (RNN), long short-term memory (LSTM), artificial neural networks (ANN), and many other detection algorithms can also be used. We used U-Net for segmentation and acquired good results. Whereas



**Fig. 7.** Water-shed segmentation to detect cysts, tumors, stone and normal case

**Table 1.** Training and validation loss and accuracy

Model	Training loss (%)	Training accuracy (%)	Validation loss (%)	Validation accuracy (%)
CNN (Keras)	0	100	0	100
VGG16	68.7	50	69	50

others use segmentation methods such as V-Net, ResNet50, and ReLU, their accuracy is between 90 and 99%. For detection and classification, detection algorithms like DAG-SGN and InceptionV3 can also be used, but CNN worked very satisfyingly for our study. We obtained this much accuracy because our data was properly preprocessed. Properly preprocessed means the data was cleaned, the missing values were handled and the data was normalized. Besides, the data was augmented to increase the diversity of the training data and prevent overfitting. That's for the dataset. Also, we used proper training algorithms in our case which was CNN (Keras). It helped us to prevent overfitting and improve model performance. That's how we got 100% accuracy in detecting the problems.

## 6 Conclusion

In this study, we used deep learning models to study photos and find signs of renal illness. The CNN (Keras) model has been successfully tested on pictures of renal problems and was found to be exceptionally accurate in identifying disorders (up to 100%). This model's limitation is that it was created only to identify a single condition. Therefore, it cannot categorize several circumstances in a single image or state. It cannot discover many conditions simultaneously with different images. In the next, we plan to develop image detection methods that may be used to classify several items from a single image concurrently. We suggested CNN works better at detecting diseases than illness detection, according to validation studies on CT images of something like the kidney. Although CNN (Keras) provided an appropriate reading in our scenario, VGG-16 is a significantly improved division for correctly classifying the photos. To appropriately identify the conditions in our dataset photos for this article, we use U-Net and Watershed to design the edges. In the future, we have to work to create a model that can accurately and with accuracy identify every disease from a single image, potentially advancing medical research and therapy. We split our dataset in two to appropriately run all of the algorithms because we needed to determine kidney diseases. The Kidney Diseases categorized photographs were the final result that was looked for after evaluating and applying all the techniques to use. The given data set has all the disease images separately, so we get the result quickly to apply the operation on that dataset easily and get the proper result.

## References

1. Zhang, H., Botler, M., Kooman, J.P.: Deep learning for image analysis in kidney care. *ScienceDirect J. Adv. Kidney Dis. Health* **30**(1), 25–32 (2023). <https://doi.org/10.1053/j.akdh.2022.11.003>
2. Alzu'bi, D., Abdullah, M., Hmeidi, I., AlAzab, R., Gharaibeh, M., El-Heis, M., Almotairi, K.H., Forestiero, A., Hussein, A.M., Abualigah, L.: Kidney tumor detection and classification based on deep learning approaches: a new dataset in CT scans. *Hindawi J. Healthcare Eng.* **2022**, 22. <https://doi.org/10.1155/2022/3861161>
3. Dritsas, E., Trigka, M.: Machine learning techniques for chronic kidney disease risk prediction. *Big Data Cogn. Comput.* **6**(3), 98. Received 29 June 2022. Revised 25 Aug 2022. Accepted 8 Sept 2022. Published 14 Sept 2022. <https://doi.org/10.3390/bdcc6030098>
4. Farzaneh, N., Reza Soroushmehr, S.M., Patel, H., Wood, A., Gryak, J., Fessell, D., Najarian, K.: Automated kidney segmentation for traumatic injured patients through ensemble learning and active contour modeling. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2018**, 3418–3421. <https://doi.org/10.1109/EMBC.2018.8512967>
5. Gharaibeh, S.M., Alzu'bi, D., Abdullah, M., Hmeidi, I., Nasar, M.R.A., Abualigah, L., Gandomi, A.H.: Radiology imaging scans for early diagnosis of kidney tumors: a review of data analytics-based machine learning and deep learning approaches. *Big Data Cogn. Comput.* **6**, 29. <https://doi.org/10.3390/bdcc6010029>, Received 8 Feb 2022. Accepted 4 Mar 2022. Published 8 Mar 2022. <https://www.mdpi.com/journal/bdcc>
6. Kalannagari Viswanath, S., Gunasundari, R.: Analysis and implementation of kidney stone detection by reaction diffusion level set segmentation using xilinx system generator on FPGA. *Hindawi Publ. Corp. VLSI Des.* **2015**, 10. Received 20 Oct 2014; Revised 15 Ap 2015; Accepted 20 Apr 2015. <https://doi.org/10.1155/2015/581961>

7. Bayram, A.F., Gurkan, C., Budak, A., Karatas, H.: A detection and prediction model based on deep learning assisted by explainable artificial intelligence for kidney diseases. *Euro. J. Sci. Technol.* **40**, 67–74. Sept 2022, © 2022 EJOSAT. 1st International Conference on Innovative Academic Studies ICIAS 2022, 10–13 Sept 2022. <https://doi.org/10.31590/ejosat.1171777>
8. Islam, M.N., Hasan, M., Hossain, M.K., Alam, M.G.R., Uddin, M.Z., Soylu, A.: Vision transformer and explainable transfer learning models for auto detection of kidney cyst, stone and tumor from CT-radiography. *Nat. Portfolio Sci. Rep.* **12**, 11440 (2022). <https://doi.org/10.1038/s41598-022-15634-4>, [www.nature.com/scientificreports](http://www.nature.com/scientificreports)
9. Zhang, H., Chen, Y., Song, Y., Xiong, Z., Yang, Y., Jonathan Wu, Q.M.: Automatic kidney lesion detection for CT images using morphological cascade convolutional neural networks. *IEEE Access* **99**, 1 (2019). <https://doi.org/10.1109/ACCESS.2019.2924207>, LicenseCC BY 4.0
10. Han, S., Hwang, S.I., Lee, H.J.: The classification of renal cancer in 3-phase CT images using a deep learning method. *J. Dig. Imag.* **32**, 638–643. <https://doi.org/10.1007/s10278-019-00230-2>. Published online 16 May 2019
11. Yao, L., Zhang, H., Zhang, M., Chen, X., Zhang, J., Huang, J., Zhang, L.: Application of artificial intelligence in renal disease. *ScienceDirect Clin. eHealth J.* [www.keaipublishing.com/CEH](http://www.keaipublishing.com/CEH), <https://doi.org/10.1016/j.ceh.2021.11.003>. Received 11 Oct 2021. Available online 15 Nov 2021
12. Wu, B., Moeckel, G.: Application of digital pathology and machine learning in the liver, kidney and lung diseases. *J. Pathol. Inf.* **PII**, S2153-3539(22)00784-2. <https://doi.org/10.1016/j.jpi.2022.100184>. Reference JPI100184. Received date 21 Sept 2022. Revised date 28 Nov 2022. Accepted date 28 Dec
13. Le, S., Allen, A., Calvert, J., Palevsky, P.M., Braden, G., Patel, S., Pellegrini, E., Green-Saxena, A., Hoffman, J., Das, R.: Convolutional neural network model for intensive care unit acute kidney injury prediction. *Clin. Res.* 1289–1298. <https://doi.org/10.1016/j.ekir.2021.02.031>
14. Vasanthselvakumar, R., Balasubramanian, M., Palanivel, S.: Detection and classification of kidney disorders using deep learning method **14**(2), 258–270. ISSN (online): 2454-7190. <https://doi.org/10.26782/jmcms.2019.04.00021>
15. Alsuhibany, S.A., Abdel-Khalek, S., Algarni, A., Fayomi, A., Gupta, D., Kumar, V., Mansour, R.F.: Ensemble of deep learning based clinical decision support system for chronic kidney disease diagnosis in medical internet of things environment. *Hindawi Comput. Intell. Neurosci.* **2021**, 13. <https://doi.org/10.1155/2021/4931450>. Received 15 Nov 2021. Revised 9 Dec 2021. Accepted 16 Dec 2021. Published 27 Dec 2021
16. Kuo, C.-C., Chang, C.-M., Liu, K.-T., Lin, W.-K., Chiang, H.-Y., Chung, C.-W., Ho, M.-R., Sun, P.-R., Yang, R.-L., Chen, K.-T.: Automation of the kidney function prediction and classification through ultrasound-based kidney imaging using deep learning. *NPJ Digital Medicine*, 29 Scripps Research Translational Institute. [www.nature.com/npjdigitalmed](http://www.nature.com/npjdigitalmed)
17. Li, D., Xiao, C., Liu, Y., Chen, Z., Hassan, H., Su, L., Liu, J., Li, H., Xie, W., Zhong, W., Huang, B.: Deep segmentation networks for segmenting kidneys and detecting kidney stones in unenhanced abdominal CT images. *Diagnostics* **12**, 1788 (2022). <https://doi.org/10.3390/diagnostics12081788>, <https://www.mdpi.com/journal/diagnostics>. Received 6 June 2022. Accepted 20 July 2022. Published 23 July 2022
18. Kim, D.-H., Ye, S.-Y.: Classification of chronic kidney disease in sonography using the GLCM and artificial neural network. *Diagnostics* **11**, 864. <https://doi.org/10.3390/diagnostics11050864>. <https://www.mdpi.com/journal/diagnostics>. Received 22 Apr 2021. Accepted 10 May 2021. Published 11 May 2021
19. Saha, R., Debi, T., Arefin, M.S.: Developing a framework for vehicle detection, tracking and classification in traffic video surveillance. In: Vasant, P., Zelinka, I., Weber, G.W. (eds.) *Intelligent Computing and Optimization. ICO 2020. Advances in Intelligent Systems*

- and Computing, vol. 1324. Springer, Cham (2021). [https://doi.org/10.1007/978-3-030-68154-8\\_31](https://doi.org/10.1007/978-3-030-68154-8_31)
20. Fatema, K., Ahmed, M.R., Arefin, M.S.: Developing a system for automatic detection of books. In: Chen, J.I.Z., Tavares, J.M.R.S., Iliyasu, A.M., Du, K.L. (eds.) Second International Conference on Image Processing and Capsule Networks. ICIPCN 2021. Lecture Notes in Networks and Systems, vol. 300. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-84760-9\\_27](https://doi.org/10.1007/978-3-030-84760-9_27)
  21. Rahman, M., Laskar, M., Asif, S., Imam, O.T., Reza, A.W., Arefin, M.S.: Flower recognition using VGG16. In: Chen, J.I.Z., Tavares, J.M.R.S., Shi, F. (eds.) Third International Conference on Image Processing and Capsule Networks. ICIPCN 2022. Lecture Notes in Networks and Systems, vol. 514. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-12413-6\\_59](https://doi.org/10.1007/978-3-031-12413-6_59)
  22. Yeasmin, S., Afrin, N., Saif, K., Imam, O.T., Reza, A.W., Arefin, M.S.: Image classification for identifying social gathering types. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_10](https://doi.org/10.1007/978-3-031-19958-5_10)
  23. Ahmed, F. et al.: Developing a classification CNN model to classify different types of fish. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_50](https://doi.org/10.1007/978-3-031-19958-5_50)
  24. Mukto, M.M., Al Mahmud, M.M., Haque, I., Imam, O.T., Reza, A.W., Arefin, M.S.: Developing a tool to classify lethal weapons by analyzing images. In: Chen, J.I.Z., Tavares, J.M.R.S., Shi, F. (eds.) Third International Conference on Image Processing and Capsule Networks. ICIPCN 2022. Lecture Notes in Networks and Systems, vol. 514. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-12413-6\\_18](https://doi.org/10.1007/978-3-031-12413-6_18)
  25. Meharaj-Ul-Mahmud, Ahmed, M.A., Alam, S.M., Imam, O.T., Reza, A.W., Arefin, M.S.: Human posture estimation: in aspect of the agriculture industry. In: Chen, J.I.Z., Tavares, J.M.R.S., Shi, F. (eds.) Third International Conference on Image Processing and Capsule Networks. ICIPCN 2022. Lecture Notes in Networks and Systems, vol. 514. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-12413-6\\_38](https://doi.org/10.1007/978-3-031-12413-6_38)
  26. Islam, M.N., Mehedi, M.H.K.: CT kidney dataset: normal-cyst-tumor and stone. In: Kaggle, Dataset to detect auto Kidney Disease Analysis. <https://www.kaggle.com/datasets/nazmul0087/ct-kidney-dataset-normal-cyst-tumor-and-stone>



# Rice Blast Disease Detection Using CNN Models and DCGAN

Abdullah Al Munem<sup>1</sup>(✉), Lamyea Tasneem Maha<sup>1</sup>, Rafid Mahmud Haque<sup>1</sup>, Noor Fabi Shah Safa<sup>1</sup>, Mozammel H. A. Khan<sup>1</sup>, and Mohammad Ashik Iqbal Khan<sup>2</sup>

<sup>1</sup> Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh  
{2019-3-60-052, 2019-1-60-055, 2019-1-60-085, 2019-1-60-060}@std.ewubd.edu, mhakhan@ewubd.edu

<sup>2</sup> Bangladesh Rice Research Institute, Gazipur, Bangladesh

**Abstract.** Blast is one of the prominent rice diseases. Automatic disease detection can assist farmers to take action timely and accurately. In this experiment, the CNN model has been employed to classify rice disease, especially Blast and Brown Spots. Several pre-trained models have been trained and evaluated on a combined dataset. The dataset was created by capturing images from the field and collecting them from multiple sources. DCGAN has been used to create new synthetic images to enhance the dataset. The obtained result before and after using DCGAN has been compared for determining the improvement of the model performance. The experimental result shows that InceptionResNet V2 outperforms the other pre-trained model after augmenting the images using DCGAN. The InceptionResNet V2 achieved the highest accuracy of 0.911, precision of 0.922, and recall of 0.900 on the combined dataset after the augmentation using DCGAN.

**Keywords:** Rice disease · Leaf blast · CNN · Transfer learning · Image processing · DCGAN

## 1 Introduction

Almost 50% of the world's population depends on rice as a staple meal, and it is cultivated in over 100 nations. [1]. For many developing countries, where a large number of people suffer from chronic malnutrition, rice is strongly linked to ensuring political stability and food security [2]. Rice consumption is reportedly increasing in different regions and for over 200 million families, rice is also one of the major income sources [2]. Due to the increasing importance of rice worldwide, the management of rice diseases has become a matter of utmost importance. Typical rice illnesses include leaf blast, brown spot, bacterial sheath rot, seedling blight, sheath rot, bacterial leaf blight, false smut, bakanae, and bacterial wilt. [3]. Among these, the rice blast disease is reported to have the most devastating effects [4]. The fungus named, *Magnaporthe grisea*, causes rice blast disease resulting in lesions on the leaves, nodes, and necks which are commonly

known as the leaf, node, and neck blast, respectively [5, 6]. Since rice blast disease causes devastating effects, rice farmers must be able to detect these diseases so that they can take the next step as soon as possible. However, another disease called the rice brown spot disease shows similar symptoms as rice blast disease. Brown spot disease shows gray or white centers with brown borders. Hence it is important to consider brown spots while identifying the blast disease.

This is where machine learning comes in which uses different algorithms to analyze a given dataset [7]. The machine learning model that this experiment will use is the Convolutional Neural Network (CNN). The model will be capable of identifying rice blast disease when an image of the infected rice plant is given as input. To build a CNN model, a pre-labeled training data set is required [8]. A dataset including pictures of rice plants with blast and brown spot infections as well as pictures of healthy rice plants will be used to train this model. Most of the previous work on crop diseases has used CNN models. In fact, The authors of the paper [9], showed that CNN models can identify crop diseases efficiently.

To accurately predict rice blast disease, CNN models which have been pre-trained have been used in this experiment. The dataset was preprocessed and augmented using several methods to improve the model performance. This study performs multiple experiments with 11 CNN models on the same dataset. This allowed all these models to be compared thoroughly.

Since these models were compared, the experiment was able to find the best model in terms of classwise evaluation results. The classwise evaluation results showed that InceptionResNetV2 outperformed other models. In the following sections existing works related to rice disease (Sect. 2), the workflow and methodology of this experiment (Sect. 3), interpretation of experimental findings and discussion (Sect. 4), and the future research possibilities (Sect. 5) are described elaborately.

## 2 Related Work

Similar works have been done related to this field in the past, mainly focusing on the detection and determination of rice plant diseases using different architectures of CNN and comparison between the accuracies of different models to find the best model for detecting rice plant diseases.

In study [10], three architectures of CNN have been used for classifying diseases which are DenseNet12, ResNet50 and VGG16. DenseNet121 showed high validation accuracies which are 90.91%, 90.63% and 94.64% for different datasets, GitHub dataset, novel dataset and the UCI Machine Learning Repository dataset respectively. A combination of the 3 datasets has given 82.03% validation accuracy using DenseNet121.

In another study [11], there are architectures like VGG16, InceptionV3 and a two-stage small CNN architecture named Simple CNN have been used. They achieved an accuracy below 50% for SqueezeNet, and below 80% for the other 3 models. Fine-tuned VGG16 has given the highest accuracy of 97.12%. The Simple CNN architecture has achieved an accuracy of 94.33%.



A study in [3] mentioned a new rice disease detection method based on deep convolutional neural networks (CNN) techniques. Compared to max-pooling and mean-pooling, stochastic pooling gained the best outcomes with 95.48% recognition accuracy which is higher than Particle Swarm Optimization (PSO) (88%), Support Vector Machine (SVM) (91%), and BP algorithm (92%).

Another study [12] proposed two models based on GoogleNet and AlexNet CNN to detect soybean plant diseases. GoogleNet and AlexNet CNN achieved an accuracy of 96% and 98%, respectively.

Another study [13] developed a LINE Bot System which is able to take the input of an image of suspected infected rice plants and detect the rice disease using YOLOv3, an object detection algorithm. Performance improvement from 91.1 to 95.6% in terms of Average True Positive Points for all classes was shown after refining the dataset in terms of size and symptoms. However, in the real deployed version of the system, the performance in terms of Average True Positive Point was only 78.86%.

In another research [14], Artificial Neural Network (ANN) was used as the classification algorithm to classify whether a rice plant was blast disease infected or healthy. A collection of 300 pictures of healthy rice plants and rice plants with blast disease were compiled into a dataset. The training accuracy for blast-infected and healthy plants was 99% and 100% respectively. However, the testing accuracy was not as promising as the training accuracy and was found to be 90% and 86% for blast-infected and healthy plants, respectively.

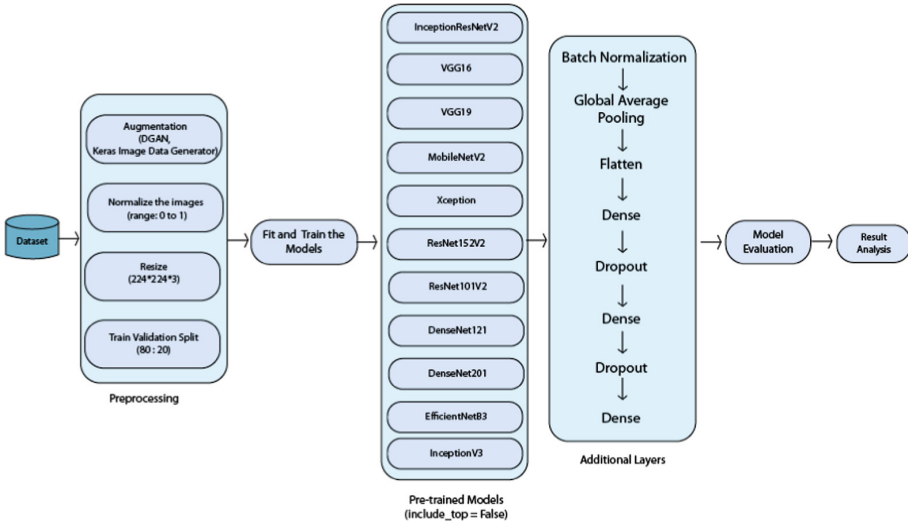
So, to conclude, rice disease has become a threat and it is important to ensure that rice disease is being correctly detected by the farmers. Hence, different machine learning algorithms are being used to identify diseases. So, this experiment is focused on developing a CNN model to predict rice blast disease effectively and efficiently.

### 3 Methodology

In this experiment, several pre-trained CNN models have been employed for rice disease classification. The models are selected based on the literature review. A combined dataset has been created by collecting RGB images from multiple sources to train and validate the models. The images are converted into grayscale and segmented images and fed into a single model (Inception ResNet V2) to verify which type of image gives the best result. The images have been preprocessed and augmented before being fed into the model. After determining the type of image that can be considered for conducting this experiment, all the selected models are trained and validated on those images. To validate the performance of each model, Precision, F1-score, Recall, and Accuracy have been used. The working flow of this experiment is shown in Fig. 1.

#### 3.1 Dataset Preparation

The dataset that has been used in this experiment contains four classes (Healthy Leaf, Leaf Blast, Brown Spot, and Invalid). The images for the dataset are collected from

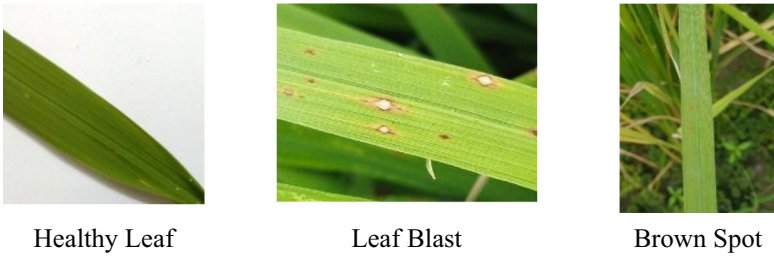


**Fig. 1.** Working flow of the method of this experiment

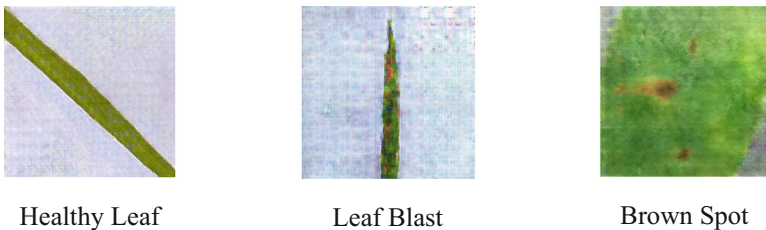
Kaggle and captured from Bangladesh Rice Research Institute (BRRI) rice field. The images are captured from the field and carefully selected for the dataset. The dataset maybe provided on request. Some of the images required background removal to include in the dataset. Hence, the background of those images is removed manually using Adobe Photoshop. There are a significant number of datasets in Kaggle that contain images of different rice diseases. Multiple open-source datasets are downloaded from Kaggle and the relavant and proper images for the experiment are manually selected. Then, the images of the BRRI field and Kaggle are merged into a single dataset. After that, DCGAN has been used for image augmentation. The DCGAN creates new images for Leaf Blast, Healthy Leaf, and Brown Spot classes from the merged dataset. Then the synthetic images for each class have been merged with the final dataset. Table 1 shows the statistical summary of the final dataset. Figures 2 and 3 shows the sample images for each class and the sample images generated using DCGAN, respectively.

**Table 1.** Statistical summary of the dataset

Classes	Number of images (without DCGAN)	Number of images (with DCGAN)
Healthy leaf	2200	5500
Leaf blast	2100	5300
Brown spot	2200	5500
Invalid	1500	4700



**Fig. 2.** Sample images of the dataset for the four classes (without preprocessing)



**Fig. 3.** Sample images for the four classes generated by DCGAN

### 3.2 Data Pre-processing

To verify which type of images can give a better model performance, the original RGB images are converted into grayscale and segmented images. The converted images are separated into two more datasets, one for grayscale and the other one for segmented images. For segmentation purposes, adaptive thresholding techniques have been used. Then the three datasets of RGB, greyscale, and segmented images are trained and evaluated using a single model. It is not necessary to train all the models using all three datasets because if one type of image can give better results than the other, it is very likely that this type of image will give better performance for the rest of the model. The images and the class labels are preprocessed using the algorithm 1.

#### Algorithm 1: Pre-processing

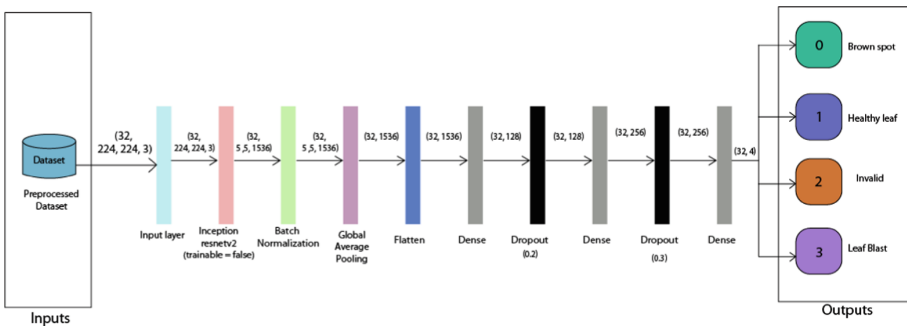
- **Begin**
- Load dataset
- Create new images using DCGAN for augmentation
- Merge new images with original dataset
- **For each image in dataset**, the following has been applied using ImageDataGenerator class
  - Normalize and rescale
  - Resize images into  $224 * 224 * 3$
  - Data augmentation (Horizontal flip, Vertical flip, Horizontal shift, Vertical shift, Random zoom)

- Create one-hot encoded 2D array for class labels
- **End for**
- **End**

### 3.3 Network Architecture

#### 3.3.1 Convolutional Neural Network (CNN)

Several pre-trained CNN models have been employed to conduct this experiment. Figure 4 shows the network architecture for a single model (Inception ResNet V2). The rest of the model has the same additional layers. Transfer learning has been used in this experiment. To train an efficient generalized model, it is necessary to feed lots of data. Pre-trained models are trained on a large dataset of multiclass images. The models have a lot of trainable parameters, and it takes a significant amount of time to train. Integrating those pre-trained models can generalize any domain-specific model with fewer data and less number of trainable parameters. Since a large number of parameters already have pre-trained weights, it takes less time to train and gives better performance in a generalized manner.



**Fig. 4.** Network architecture of modified inception ResNet V2 model

The dataset is fed into the input layer of the models. The input layer passes the images to the pre-trained layers for each model. The output vector of the pre-trained model is passed through batch normalization, global average pooling, and flatten layer. The batch normalization layer can speed up the training process, decrease the importance of initial weights, and slightly regularize the model. By decreasing the range of initial weights, batch normalization solves a major issue called internal covariate shift. Global Average Pooling reduces computational time and resources by averaging all features within a feature map. The pre-trained layers are used for extracting feature vectors from the images and the fully connected dense layers are used for classifying the rice disease.

### 3.3.2 Deep Convolutional Generative Adversarial Network (DCGAN)

In this experiment, DCGAN has been used for image augmentation. The DCGAN architecture consists of two distinct models, one is the generator and another one is the discriminator. The generator creates synthetic images from random noise, and the discriminator takes the images and classifies the images as real or fake. During training, the discriminator is trained on real images and updates the weights, then takes images from the generator and classifies whether the generated image is fake or real. After classifying the generated image, based on the generator loss the generator updates the weights and generates a new set of fake images until the discriminator classifies the generated fake image as a real image. Meanwhile, the discriminator also updates the weights based on discriminator loss. In this process, the generator learns to create real-looking synthetic images.

The generator model takes 1D random noise and pass-through a dense layer. Then the output vector of the dense layer is reshaped and fed into multiple batch normalization and transposed convolution layers. Transposed convolution layers are used to upsample the input vectors (feature maps) to desired feature maps. The discriminator takes the input images and pass-through multiple convolution layers. The convolution layers extract and downsample the feature vectors. Then the feature vectors are fed into fully connected dense layers for classification.

Table 2 shows the selected values of the hyperparameters for this experiment.

## 4 Experimental Result and Discussion

The models are assessed on the basis of accuracy, precision, and recall. At first, the models are trained and evaluated on the final dataset which does not include the generated images by DCGAN, to find out the best suitable model for the rice disease classification. Then the best model was trained and evaluated on the merged dataset including the generated images by DCGAN and compared with the previous result to determine whether augmenting the dataset using the DCGAN can improve the model performance or not. To find out which type of image can give the better result on the validation dataset, the InceptionResNet V2 model was trained and evaluated on RGB, Grayscale, and Segmented image dataset.

Table 3 shows that the InceptionResNetV2 model gives the highest Accuracy (0.8717), Precision (0.8807), and Recall (0.8587) for the RGB images on the validation dataset. Although the training result of Grayscale images is close to or the same as RGB images, the validation result of Grayscale images is lower than RGB images. Thus, the dataset of RGB images is considered for further experiments. Total of ten models have been trained and evaluated on RGB images to find out the best model for classifying rice disease. Table 4 shows the experimental result of this step.

**Table 2.** The selected values for the hyperparameter tuning

Hyperparameters	Value(s)	
Image size	224 × 224 × 3	
Validation split	20%	
Pooling	Max-pooling	
Activation	Hidden layers (CNN)	ReLu
	Output layer (CNN)	Softmax
	Hidden layers and output layer (generator, DCGAN)	Leaky ReLU
	Hidden layers (discriminator, DCGAN)	Leaky ReLU
	Output layer (discriminator, DCGAN)	Sigmoid
Optimizer	CNN	Adam
	DCGAN	Adam
Learning rate (CNN)	0.0001	
Loss	CNN	Categorical cross entropy
	DCGAN	Binary cross entropy
Metrics	CNN	Accuracy, precision, recall, F1-score
	DCGAN	Accuracy
Epoch	CNN	50
	DCGAN	700
Batch size	CNN	32
	DCGAN	32

**Table 3.** InceptionResNetV2’s evaluation result on RGB, grayscale and segmented images during training (T) and validation (V)

Image Type	Loss (T)	Accuracy (T)	Precision (T)	Recall (T)	Loss (V)	Accuracy (V)	Precision (V)	Recall (V)
RGB	0.321	<b>0.870</b>	<b>0.883</b>	0.856	<b>0.352</b>	<b>0.871</b>	<b>0.880</b>	<b>0.858</b>
Grayscale	<b>0.320</b>	0.860	0.882	<b>0.859</b>	0.332	0.850	0.864	0.834
Segmented	0.557	0.748	0.801	0.675	0.560	0.762	0.809	0.683

From Table 4, DenseNet201 achieved the highest Accuracy (0.8994, 0.8815), Precision (0.9087, 0.8920), and Recall (0.8899, 0.8707), and EfficientNetB3 achieved the lowest Accuracy (0.3220, 0.2750), Precision (0.5887, 0.0000), and Recall (0.0109, 0.0000)

**Table 4.** Experimental results on train (T) and validation (V) dataset

Model name	Loss (T, V)	Accuracy (T, V)	Precision (T, V)	Recall (T, V)
InceptionResNetV2	0.321, 0.352	0.870, 0.871	0.883, 0.880	0.856, 0.858
VGG16	0.490, 0.522	0.789, 0.776	0.827, 0.811	0.742, 0.741
VGG19	0.533, 0.537	0.764, 0.771	0.818, 0.812	0.705, 0.707
MobileNetV2	0.353, 0.410	0.852, 0.830	0.870, 0.854	0.831, 0.807
Xception	0.342, 0.372	0.860, 0.841	0.875, 0.853	0.841, 0.826
ResNet152V2	0.300, 0.352	0.883, 0.863	0.894, 0.881	0.871, 0.845
ResNet101V2	0.290, 0.352	0.883, 0.856	0.894, 0.862	0.871, 0.843
DenseNet121	0.301, <b>0.325</b>	0.879, 0.876	0.891, 0.889	0.866, 0.856
DenseNet201	<b>0.263</b> , 0.326	<b>0.899</b> , <b>0.881</b>	<b>0.908</b> , <b>0.892</b>	<b>0.889</b> , <b>0.870</b>
EfficientNetB3	1.323, 1.407	0.322, 0.275	0.588, 0.000	0.019, 0.000
InceptionV3	0.318, 0.357	0.874, 0.858	0.885, 0.880	0.861, 0.842

on both training and validation dataset. The accuracy score of InceptionResNetV2, MobileNetV2, Xception, ResNet152V2, ResNet101V2, DenseNet201, DenseNet121 and InceptionV3 is more than 80%. To find out the best model among those, the class-wise comparison for each model is shown in Table 5.

From Table 5, we can see InceptionResNetV2 reached the maximum level of F1-score (0.82), and precision (0.83), and DenseNet201 achieved the highest recall (0.87) for the class Leaf Blast. InceptionResNetV2 reached the maximum precision (0.96) and ResNet152V2 reached the highest F1-score (0.93) and recall (0.92) for Brown Spot. InceptionResNetV2 reached the maximum recall (0.92), and F1-score (0.87), and DenseNet201 achieved the highest precision (0.87) for Healthy Leaf. DenseNet121 also achieved the highest F1-score (0.87) for the Healthy Leaf. For the Invalid class, the results are not significant since almost every model achieved 1.00 or 0.99 of F1-score, precision, and recall. So, this class is not mentioned in Table 6. As InceptionResNetV2 gives comparatively efficient results for each class, it can be concluded that InceptionResNetV2 is the most suitable pre-trained model to classify rice disease.

**Table 5.** Evaluation result for leaf blast (LB), brown spot (BS), and healthy leaf (HL) class on validation dataset

Model name	Precision (LB, BS, HL)	Recall (LB, BS, HL)	F1-score (LB, BS, HL)
InceptionResNetV2	<b>0.83, 0.96</b> , 0.84	0.81, 0.87, <b>0.92</b>	<b>0.82</b> , 0.91, <b>0.87</b>
MobileNetV2	0.68, 0.94, 0.81	0.81, 0.80, 0.79	0.74, 0.86, 0.80
Xception	0.74, 0.93, 0.84	0.82, 0.84, 0.83	0.78, 0.88, 0.84
ResNet152V2	0.81, 0.93, 0.85	0.83, <b>0.92</b> , 0.85	0.82, <b>0.93</b> , 0.85
ResNet101V2	0.76, 0.94, 0.84	0.84, 0.88, 0.82	0.79, 0.91, 0.83
DenseNet121	0.81, 0.95, 0.84	0.82, 0.87, 0.89	0.81, 0.91, <b>0.87</b>
DenseNet201	0.75, 0.95, <b>0.87</b>	<b>0.87</b> , 0.86, 0.82	0.80, 0.91, 0.84
InceptionV3	0.78, 0.93, 0.85	0.82, 0.89, 0.85	0.80, 0.91, 0.85

**Table 6.** InceptionResNetV2's evaluation result before and after the image augmentation using DCGAN during training (T) and validation (E)

Dataset	Loss (T)	Accuracy (T)	Precision (T)	Recall (T)	Loss (V)	Accuracy (V)	Precision (V)	Recall (V)
Before	0.321	0.870	0.883	0.856	0.352	0.871	0.880	0.858
After	<b>0.144</b>	<b>0.941</b>	<b>0.949</b>	<b>0.935</b>	<b>0.264</b>	<b>0.911</b>	<b>0.922</b>	<b>0.900</b>

From Table 6, accuracy, precision, and recall for both training and validation datasets are increased and the loss is decreased after the image augmentation using DCGAN. Thus, it can be concluded that image augmentation using the DCGAN can improve the model performance for rice disease classification. Although it improves the model performance, it also increases the training time and resource utilization because images in the dataset are increased through augmentation.

## 5 Conclusion

Since rice is the main source of carbs and is consumed and farmed by a significant number of people worldwide, it is necessary to identify rice disease autonomously without human intervention. It will speed up the process of dealing with the disease. In this experiment, several pre-trained CNN models have been trained and evaluated to do this task. By analyzing the experimental result, it has been concluded that InceptionResNetV2 is the best suitable model to classify rice disease based on the particular dataset that was used in this experiment. It has also been concluded that image augmentation using DCGAN can improve the model performance. Another finding of this experiment is that RGB image is the best suitable type of image to train the model.



Although InceptionResNetV2 gives the best performance among all the pre-trained models that have been employed in this experiment, ResNet152V2, DenseNet121, and DenseNet201 also give good performance. Depending on the dataset and type of disease, all of those models can give precise results. From the class-wise comparison of Sect. 4, it is shown that some of the models can give better results for a specific type of disease. For example, it seems that ResNet152V2 can identify Brown Spot more precisely compared to other models since ResNet152V2 obtained the maximum f1-score (0.93) and recall (0.92). But InceptionResNetV2 gives better performance for every class, thus it is considered the best suitable model for rice disease classification. To utilize the different characteristics of those models and improve the result for each class, the ensemble learning techniques can be experimented with in the future.

## References

1. Fukagawa, N.K., Ziska, L.H.: Rice: importance for global nutrition. *J. Nutr. Sci. Vitaminol.* (Tokyo) **65**, S2–S3 (2019). <https://doi.org/10.3177/jnsv.65.S2>
2. Muthayya, S., Sugimoto, J.D., Montgomery, S., Maberly, G.F.: An overview of global rice production, supply, trade, and consumption. *Ann. N. Y. Acad. Sci.* **1324**(1), 7–14 (2014). <https://doi.org/10.1111/nyas.12540>
3. Lu, Y., Yi, S., Zeng, N., Liu, Y., Zhang, Y.: Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* **267**, 378–384 (2017). <https://doi.org/10.1016/j.neucom.2017.06.023>
4. Asibi, A.E., Chai, Q., Coulter, J.A.: Rice blast: a disease with implications for global food security. *Agronomy* **9**(8), 1–14 (2019). <https://doi.org/10.3390/agronomy9080451>
5. Yang, Y.H., Yang, D.S., Lei, H.M., Li, C.Y., Li, G.H., Zhao, P.J.: Griseaketides A-D, new aromatic polyketides from the pathogenic fungus *Magnaporthe grisea*. *Molecules* **25**(1), 1–8 (2020). <https://doi.org/10.3390/molecules25010072>
6. Neupane, N., Bhusal, K.: A review of blast disease of rice in Nepal. *Artic. J. Plant Pathol. Microbiol.* **12**(1), 528 (2021). <https://doi.org/10.35248/2157-7471.20.12.528>
7. Sarker, I.H.: Machine learning: algorithms, real-world applications and research directions. *SN Comput. Sci.* **2**(3), 1–21 (2021). <https://doi.org/10.1007/s42979-021-00592-x>
8. O’Shea, K., Nash, R.: An Introduction to Convolutional Neural Networks, pp. 1–11 (2015)
9. Moin, N.B., Islam, N., Sultana, S., Chhoa, L.A., Ruhul Kabir Howlader, S.M., Ripon, S.H.: Disease detection of Bangladeshi crops using image processing and deep learning—a comparative analysis. In: 2022 2nd International Conference on Intelligent Technologies (CONIT). <https://doi.org/10.1109/conit55038.2022.9847715>
10. Islam, A., Islam, R., Haque, S.M.R., Islam, S.M.M., Khan, M.A.I.: Rice leaf disease recognition using local threshold based segmentation and deep CNN. *Int. J. Intell. Syst. Appl.* **13**(5), 35–45 (2021). <https://doi.org/10.5815/ijisa.2021.05.04>
11. Rahman, C.R., et al.: Identification and recognition of rice diseases and pests using convolutional neural networks. *Biosyst. Eng.* **194**, 112–120. <https://doi.org/10.1016/j.biosystemseng.2020.03.020>
12. Jadhav, S.B., Udupi, V.R., Patil, S.B.: Identification of plant diseases using convolutional neural networks. *Int. J. Inf. Technol.* **13**(6), 2461–2470 (2020). <https://doi.org/10.1007/s41870-020-00437-5>

13. Temniranrat, P., Kiratiratanapruk, K., Kitvimonrat, A., Sinthupinyo, W., Patarapuwadol, S.: A system for automatic rice disease detection from rice paddy images serviced via a Chatbot. *Comput. Electron. Agric.* **185**, 1–19 (2021). <https://doi.org/10.1016/j.compag.2021.106156>
14. Ramesh, S., Vydeki, D.: Rice blast disease detection and classification using machine learning algorithm. In: *Proceedings of 2nd International Conference on Micro-Electronics Telecommunication Engineering. ICMETE 2018*, pp. 255–259. <https://doi.org/10.1109/ICMETE.2018.00063>



# Evaluation of Performance of Different Machine Learning Techniques for Structural Models

Melda Yücel, Gebrail Bekdaş, and Sinan Melih Nigdeli<sup>(✉)</sup>

Department of Civil Engineering, Istanbul University-Cerrahpaşa, 34320 Avcılar, Istanbul, Turkey  
{bekdas,melihnig}@iuc.edu.tr

**Abstract.** In the field of structural engineering, generating the optimum design is extremely important as well as that to adjust effort and, minimizing the operation time is also a significant issue. With this respect, while the optimum designs for structural models are tried to generate, several methodologies can be useful in terms of decreasing the working time and providing saving time. In this meaning, for the current study, different machine learning methodologies are beneficial to determine the optimum parameters for two different structural models including a 3-bar truss together with a simply supported reinforced concrete (RC) beam. In this process, three machine learning methods are investigated and compared with the previous studies where artificial neural networks (ANNs) were observed to predict the optimum parameter results. Also, from the mentioned methodologies, the best one is selected, and the prediction performance with error rates according to actual data is investigated in terms of a different test dataset for both structural models. By this means, with respect to variable design alternatives for structural models, the most effective, and successful methodology to predict the optimal parameters, and target functions can be determined directly in a short time.

**Keywords:** Structures · Machine learning · Prediction · Bagging · Support vector regression · k-nearest neighbor · Artificial neural networks

## 1 Introduction

In modern days, it is a requirement to make an optimum design. While most of the engineering calculations fit into mathematical optimization methods, several complex optimum design problems fall into non-linear types of research areas due to the existence of design constraints related to safety, usability and feasibility. In that case, it is needed to use metaheuristic methods that are used in the optimum design of structural systems. The design involves the optimum design of several systems [1–5] including reinforced concrete (RC) structures [6] and structural control systems [7–10].

In the metaheuristic methods, the optimization process is iterative and it is coded with all problem calculations including the definition of all design constants. Whereas artificial intelligence (AI) can be also used in this area to find rapid solutions to different cases. As a human who can learn, think and decide, AI can predict solutions. Via machine learning, several prediction models have been proposed for engineering problems [11].

In the present study, two structural models were investigated. The optimum data of these models were used in machine learning. The problems are the 3-bar truss problem and the simply-supported RC beam. Prediction models were generated via three machine learning methods such as support vector regression (SVR), k-nearest neighbor (k-NN) and bagging.

## 2 The Prediction Methodologies

During the generation of machine learning models, three different methodologies benefited including both support vector regression (SVR), which is one of the numerical prediction techniques, and k-nearest neighbor as a classification technique together with an ensemble model known as Bagging. The details of them in terms of working principle for prediction processes were presented below.

### 2.1 Support Vector Regression (SVR)

The support vector machine (SVM) method converts to the support vector regression (SVR) algorithms in the case that they can be utilized for the solution of nonlinear regression problems. In this technique as SVR, for each regression training data, it is provided that a kind of function is developed, which will provide to arise of the amount of minimum estimation error by producing the closest and correct output value to this data. In this regard, the least squares estimation method, which is widely used in regression problems, cannot be so effective and useful in the case that extreme/outlier values exist in the problem dataset. In these circumstances, it is consistent that the necessity of usage of a loss function, which can be adapted and is indicated as  $\varepsilon$  toward any negative case [12, 13]. This  $\varepsilon$  loss function can be formalized via Eq. (1):

$$\varepsilon(x, y, f) = |y - f(x)|_{\varepsilon} = \max(0, |y - f(x)| - \varepsilon) \quad (1)$$

Here,  $f$  is a real-valued function in  $x$ , and  $f(x)$  is an estimator function, which produces the predicted value for the targeted/real  $y$  output.

### 2.2 K-Nearest Neighbor (k-NN)

K-nearest neighbor namely k-NN, which is essentially a classification method, is also can be utilized for regression problems. k-nearest neighbor method, which was proposed at the beginning of the 1950s, and then, extended by T. M. Cover and P. E. Hart with respect to classification, is based on learning via mimicking, namely comparison of a specific test group with the other training groups similar to this group. In this meaning, this method provides to realize predicting of value or defining the value of this training group via classification of other samples, which are the closest/most similar to the mentioned training group. In this direction, when an unknown group is given, the k-nearest neighbor classifier searches a pattern space for the k-training groups, which are the closest ones to it. Anymore, these k-training groups became “k-nearest neighbors” for unknown groups [14–16].

### 2.3 Bagging

Bagging (bootstrap aggregating), which is one of the ensemble methods, was developed by Leo Breiman in 1996. In the process of prediction of a numerical result via bagging methodology, which was generated via a combination of more sub-predictive models, an average value is determined for the results provided by all sub-models, whereas a majority rate is benefited during the classification of a label. In this regard, each of the mentioned sub-models is generated by using different subsets of training data, and a higher success is aimed to provide than a single predictive model for the main combined model, which consists via combining of these sets. Also, training samples, which are used by each sub-predictive model, are selected from the main dataset randomly, and each sample data can be taken place in multiple models [17].

## 3 Numerical Examples

### 3.1 Structural Models

In the present study, two different structural models were handled to evaluate and compare the prediction performances of the mentioned machine learning methodologies. In this respect, the first model was determined as a 3-bar truss design [18] together with the second one is handled as simply-supported reinforced concrete (RC) beam [19] from the previous studies. By this means, in Figs. 1 and 2, the structural details of both models can be investigated in terms of loading conditions, geometric section properties, etc., respectively.

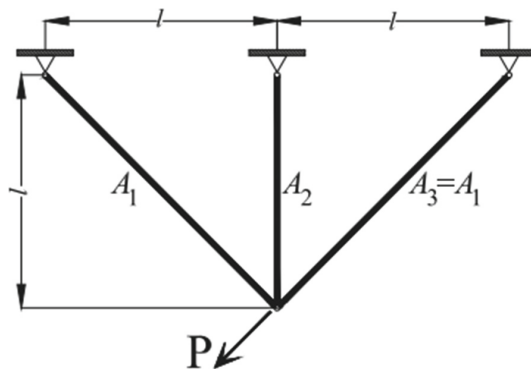


Fig. 1. Structural model and details for 3-bar truss design [18]

On the other side, to generate the prediction models, the main dataset from the previous studies, which were mentioned above section, was investigated. By this means, seven separate prediction models were reproduced by using each current machine learning methodologies corresponding to all of the input parameters.

With this respect, in Tables 1 and 2, all of the details of data for the handled structural models can be observed in terms of input, and output parameters besides numerical information, respectively.

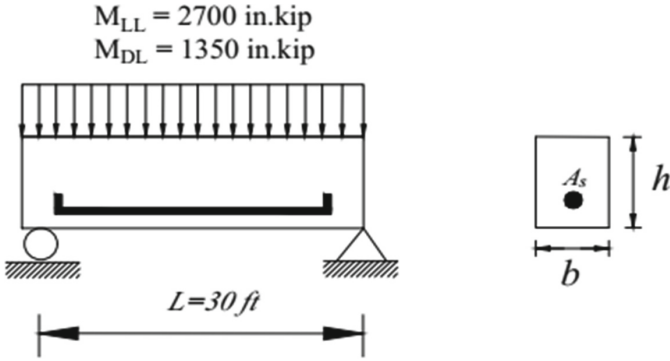


Fig. 2. Structural model and details for simply-supported RC beam design [19]

Table 1. Details of prediction models for 3-bar truss and RC simply-supported beam.

Model	Inputs	Symbolization	Outputs	Symbolization
3-bar truss	Load	P	Section area of the 1st and 3rd bars	$A_1 = A_3$
			Section area of the 2nd bars	$A_2$
			Minimum volume	$Min f (v)$
Simply-supported RC beam	Length of beam	$L$	Height of beam	$h$
	Dead load	$M_{DL}$	Breadth of beam	$b$
	Live load	$M_{LL}$	Section area of reinforcement	$A_s$
			Minimum cost	$Min f (c)$

Table 2. Some numerical information for both structural models.

Model	Inputs	Range of data	Increment	Unit
3-bar truss	P	0.1–2.8	0.1	kN
Simply-supported RC beam	$L$	15–20	0.5	Feet
	$M_{DL}$	1000–3000	200	In-kip
	$M_{LL}$			In-kip

### 3.2 Investigation of Prediction Models

In this section, the prediction success and efficiency of the generated seven separate prediction models were evaluated in comparison to artificial neural networks (ANNs), which have been utilized in previous studies [18, 19]. With this respect, all of the output

parameters were tried to predict with the help of the observed machine learning methodologies. In this meaning, the main dataset for both structures was investigated, and a test dataset was evaluated (including different structural model couples) with the best one among the mentioned methodologies in terms of prediction convergence rate, and smallness of errors. Here, In Tables 3 and 4, the observations for main models can be seen corresponding to predicted values, and error measurements for output parameters. On the other side, the same information was presented in Tables 5 and 6 for both structural models with respect to the test dataset, respectively.

**Table 3.** Prediction values and evaluation metrics for 3-bar truss structure provided by the investigated machine learning methodologies.

<i>Parameter</i>	<i>Metric</i>	SVR	k-NN	Bagging	ANNs [18]
$A_1 = A_3$	Mean absolute error (MAE)	0.0074	0.0289	0.0488	0.0034
	Root mean squared error (RMSE)	0.0232	0.0334	0.0571	0.0049
	Mean squared error (MSE)	0.0005	0.0011	0.0033	0.000024
$A_2$	Mean absolute error (MAE)	0.0244	0.0303	0.0476	0.0097
	Root mean squared error (RMSE)	0.0818	0.0523	0.0957	0.0138
	Mean squared error (MSE)	0.0067	0.0027	0.0092	0.0002
Min $f(v)$	Mean absolute error (MAE)	0.5374	10.7618	17.2208	0.0086
	Root mean squared error (RMSE)	1.5856	12.1602	21.6349	0.0190
	Mean squared error (MSE)	2.5141	147.8705	468.0689	0.0004

## 4 Results

### 4.1 Prediction Performance and Error Measurements for Training Models

In terms of the main training model, the best one among the handled methodologies is determined as a support vector machine (SVR) in general meaning. Also, Bagging may be considered a relatively effective method in comparison to k-NN in terms of RC simply-supported beam. The reason for this case is based on the smallness of error rates and in this way, the highness of the prediction performance of the selected methodology. If all of the error metrics are investigated, it can be seen that especially MAE and RMSE values provided by SVR are extremely low for the 3-bar truss structure in terms of prediction of all output parameters (MAE are 0.0074, 0.0244, 0.5374, and RMSE are 0.0232, 0.0818, 1.5816 for  $A_1 = A_3$ ,  $A_2$ , and  $\min f(v)$ , respectively). However, it is clear that the best methodology is ANNs because the smallest errors namely differences/deviations observed between actual and predicted values for outputs are ensured in comparison to the other methodologies. Although the best methodologies are determined as SVR, and Bagging for predicting  $h$ ,  $b$ , and  $A_s$  parameters successfully in terms of RC simply-supported beam, the most effective is ANNs to determine  $\min f(c)$  with the smallest error

**Table 4.** Prediction values and evaluation metrics for simply-supported RC beam structure provided by the investigated machine learning methodologies.

<i>Parameter</i>	<i>Metric</i>	SVR	k-NN	Bagging	ANNs [19]
<i>h</i>	Mean absolute error (MAE)	0.5885	0.4450	0.0780	0.4480
	Root mean squared error (RMSE)	0.7334	0.5936	0.1583	0.5543
	Mean squared error (MSE)	0.5379	0.3524	0.0251	0.3072
<i>b</i>	Mean absolute error (MAE)	0.1508	0.1217	0.0186	0.1254
	Root mean squared error (RMSE)	0.1893	0.0196	0.0356	0.1545
	Mean squared error (MSE)	0.0358	0.0004	0.0013	0.0239
$A_s$	Mean absolute error (MAE)	0.0822	0.1076	0.0131	0.1648
	Root mean squared error (RMSE)	0.1148	0.1845	0.0504	0.2048
	Mean squared error (MSE)	0.0132	0.0340	0.0025	0.0419
Min $f(c)$	Mean absolute error (MAE)	2.2084	2.7411	3.5971	0.4219
	Root mean squared error (RMSE)	3.1697	3.5337	4.4745	0.5228
	Mean squared error (MSE)	10.0470	12.4870	20.0212	0.2733

**Table 5.** Comparison of the best-selected among the utilized methodologies, and previous study in terms of 3-bar truss.

<i>Parameter</i>	<i>Metric</i>	SVR	ANNs [18]
$A_1 = A_3$	Mean absolute error (MAE)	0.0221	0.0049
	Root mean squared error (RMSE)	0.0478	0.0056
	Mean squared error (MSE)	0.0022	0.0000
$A_2$	Mean absolute error (MAE)	0.0787	0.0131
	Root mean squared error (RMSE)	0.1733	0.0148
	Mean squared error (MSE)	0.0300	0.0002
Min $f(v)$	Mean absolute error (MAE)	1.7267	0.0918
	Root mean squared error (RMSE)	3.5475	0.1335
	Mean squared error (MSE)	12.5851	0.0178

rates (MAE, MSE and RMSE are 0.4219, 0.2733 and 0.5228, respectively). Nonetheless, due to that error rates of ANNs are quite low in terms of other parameters. For this reason, it can be accepted that the most successful, and effective methodology is determined as ANNs for the generation of the optimal structural designs by detecting the desired parameter values directly.



**Table 6.** Comparison of the best-selected among the utilized methodologies, and previous study in terms of RC simply-supported beam.

<i>Parameter</i>	<i>Metric</i>	SVR	Bagging	ANNs [19]
<i>h</i>	Mean absolute error (MAE)	0.7072	0.6256	0.4798
	Root mean squared error (RMSE)	0.8004	0.7889	0.6036
	Mean squared error (MSE)	0.6407	0.6224	0.3643
<i>b</i>	Mean absolute error (MAE)	0.1708	0.1357	0.1138
	Root mean squared error (RMSE)	0.1977	0.1728	0.1362
	Mean squared error (MSE)	0.0391	0.0299	0.0185
$A_s$	Mean absolute error (MAE)	0.1918	0.1404	0.1667
	Root mean squared error (RMSE)	0.2495	0.1827	0.1886
	Mean squared error (MSE)	0.0622	0.0334	0.0355
Min $f(c)$	Mean absolute error (MAE)	1.9299	3.6928	0.3659
	Root mean squared error (RMSE)	2.3776	4.9435	0.5147
	Mean squared error (MSE)	5.6534	24.4386	0.2649

## 4.2 Prediction Performance and Error Measurements for Test Models

Corresponding to the test models, the different datasets utilized in the previous studies [18, 19] were evaluated via only the selected best methodologies in terms of both structural models. For this respect, SVR besides SVR with Bagging was investigated for 3-bar truss and simply-supported RC beam, respectively. When the numerical results and calculations for predictions together with error metrics are observed and compared with ANNs, it can be noticed that ANNs are the most effective methodology in terms of convergence to actual data with minimum prediction errors for all outputs, especially for minimum volume ( $\min f(v)$ ) (MAE, MSE and RMSE are 0.0049, 0.0000, 0.0056; 0.0131, 0.0002, 0.0148; 0.0918, 0.0178, 0.1335 for  $A_1 = A_3$ ,  $A_2$  and  $\min f(v)$ , respectively). Also, there is a great difference between SVR and ANNs in terms of the MSE at the rate of 12.5653. In the consideration of these metrics for RC beam in terms of all output parameters, it can be seen that ANNs are more successful than other methodologies similar to the 3-bar truss structure. Although Bagging and SVR are relatively effective according to ANNs with respect to design parameter as  $A_s$ , ANNs are powerful to converge to the optimum values of other design parameters together with the target function as  $\min f(c)$ , especially. The main reason for this situation is related to low deviation namely error metrics to predict the actual values of optimum output parameters.

## 5 Conclusion

In summary, it can be understood from the observations for predicting the structural parameters, and target functions, ANNs are the most successful and effective algorithm in comparison to the other machine learning methodologies for both 3-bar truss, and

simply-supported RC beam. The cause of this is the smallness of error values between the predicted values and actual data for outputs. Also, while the target functions as the minimum volume, for a 3-bar truss, and minimum cost for the RC beam can be determined similarly to actual data via ANNs. These investigations reflect that ANNs are relatively effective in terms of convergence to actual data with low error/deviation rates, and reliable to detect almost equal optimum namely actual outputs for design parameters besides target functions.

## References

1. Toklu, Y.C., Bekdas, G., Nigdeli, S.M.: *Metaheuristics for Structural Design and Analysis*. John Wiley & Sons (2021)
2. Carbas, S., Toktas, A., Ustun, D.: *Nature-Inspired Metaheuristic Algorithms for Engineering Optimization Applications*. Springer Nature. <https://doi.org/10.1007/978-981-33-6773-9>
3. Kaveh, A., Eslamlou, A.D., Khodadadi, N.: Dynamic water strider algorithm for optimal design of skeletal structures. *Periodica Polytechnica Civ. Eng.* **64**(3), 904–916 (2020)
4. Kayabekir, A.E., Bekdaş, G., Yücel, M., Nigdeli, S.M., & Geem, Z.W.: Harmony search algorithm for structural engineering problems. In: *Nature-Inspired Metaheuristic Algorithms for Engineering Optimization Applications*, pp. 13–47. Springer, Singapore (2021)
5. Azizi, M., Shishehgarkhaneh, M.B., Basiri, M.: Optimum design of truss structures by material generation algorithm with discrete variables. *Dec. Anal. J.* **3**, 100043 (2022)
6. Nigdeli, S.M., Bekdaş, G., Kayabekir, A.E., Yucel, M.: *Advances in Structural Engineering—Optimization: Emerging Trends in Structural Optimization* (2020)
7. Katebi, J., Shoaee-parchin, M., Shariati, M., Trung, N.T., Khorami, M.: Developed comparative analysis of metaheuristic optimization algorithms for optimal active control of structures. *Eng. Comput.* **36**, 1539–1558 (2020)
8. Ulusoy, S., Kayabekir, A.E., Nigdeli, S.M., Bekdaş, G.: Metaheuristic-based structural control methods and comparison of applications. In: *Nature-Inspired Metaheuristic Algorithms for Engineering Optimization Applications*, pp. 251–276. Springer, Singapore (2021)
9. Roozbahan, M., Jahani, E.: Optimal design of elastic and elastoplastic tuned mass dampers using the mouth brooding fish algorithm for linear and nonlinear structures. In: *Structures*, vol. 43, pp. 1084–1090. Elsevier (2022)
10. Ocak, A., Bekdaş, G., Nigdeli, S.M., Kim, S., Geem, Z.W.: Optimization of tuned liquid damper including different liquids for lateral displacement control of single and multi-story structures. *Buildings* **12**(3), 377 (2022)
11. Bekdaş, G., Nigdeli, S.M., Yücel, M.: *Artificial Intelligence and Machine Learning Applications in Civil, Mechanical, and Industrial Engineering*. IGI Global. ISBN: 9781799803010 (2020)
12. Wu, X., Kumar, V.: *The Top Ten Algorithms in Data Mining*. CRC Press, Florida, USA. ISBN: 978-1-4200-8964-6 (2009)
13. Witten, I.H., Frank, E., Hall, M.A.: *Data Mining Practical Machine Learning Tools and Techniques Third Edition*. Morgan Kaufmann Publishers, Massachusetts, USA. ISBN: 978-0-12-374856-0 (2011)
14. Bramer, M.: Principles of data mining. In: Mackie, I. (ed.) *Undergraduate Topics in Computer Science*, 3rd ed., vol. 180. Springer, London, UK, ISBN: 978-1-4471-7306-9 (2016)
15. Han, J., Kamber, M.: *Data Mining: Concepts and Techniques*, 2nd edn. Morgan Kaufmann, California, USA (2006)

16. Taneja, S., Gupta, C., Goyal, K., Gureja, D.: An enhanced k-nearest neighbor algorithm using information gain and clustering. In: 2014 Fourth International Conference on Advanced Computing & Communication Technologies, pp. 325–329. 8–9 February 2014, Rohtak-India, IEEE (2014)
17. Breiman, L.: Bagging predictors. *Mach. Learn.* **24**(2), 123–140 (1996)
18. Yücel, M., Bekdaş, G., Nigdeli, S.M.: Prediction of optimum 3-bar truss model parameters with an ANN model. In: International Conference on Harmony Search Algorithm, pp. 317–324. Springer, Singapore (2020)
19. Bekdaş, G., Yücel, M., Nigdeli, S.M.: Estimation of optimum design of structural systems via machine learning. *Front. Struct. Civ. Eng.* 1–12 (2021). <https://doi.org/10.1007/s11709-021-0774-0>



# Age Estimation from Human Facial Expression Using Deep Neural Network

Md. Ashiqur Rahman, Shuhena Salam Aonty, and Kaushik Deb<sup>(✉)</sup>

Department of Computer Science and Engineering, Chittagong University of Engineering and Technology (CUET), Chattogram 4349, Bangladesh  
u1704026@student.cuet.ac.bd  
{shuhena,debkaushik99}@cuet.ac.bd

**Abstract.** Human age estimation from a single image has been significant for researchers, yet challenging tasks in computer vision. Age estimation is becoming more complicated as the vast diversity of individuals' faces includes those with diverse races, genders, and geographical locations. As a result, the accuracy of age determination depends on a large image dataset. Therefore, an attention module can be used to give extra attention to important patches in the image. Whereas a conventional Convolutional Neural Network model treats every pixel of an image equally. This paper proposes a CNN model with a self-attention module, and two different datasets are merged for evaluation. However, an attention mechanism is regarded as a dynamic weight adjustment process. In this process, scaled dot-product is used as an attention-scoring function for the query and key vector element-wise multiplication. It adjusts weight dynamically through back-propagation while training the model. Moreover, transfer learning is utilized to get optimal performance faster. However, the proposed model achieves state-of-the-art results, outperforming other models by attaining the lowest mean absolute error, 3.515.

**Keywords:** Age classification · CNN model · Attention module · Transfer learning

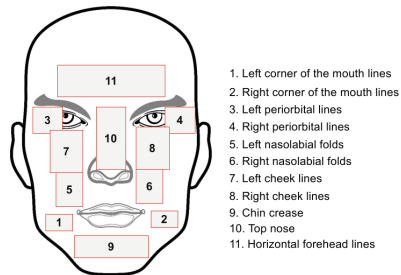
## 1 Introduction

The emergence of various patterns in facial appearance is thought to be a primary cause of human age, which is a significant personal characteristic. Age is a notion that describes a person's age at a specific period and is a crucial biological component. The aging process is ongoing and always progresses higher. It cannot be undone.

The aging process varies globally due to geography, race, gender, environment, weather, habits, food, and living situations. These are the aging process's global effects [1]. These are the aging process's global effects. Under the same

general circumstances, the aging process might also differ from person to person. Moreover, human aging processes and growth rates are caused by internal biological changes such as hormonal shifts and catastrophic diseases like AIDS and cancer. The human aging process is a natural phenomenon that can not be stopped via artificial means. Because of this, the human age plays a positive role in various fields, including human-computer interactions, identification, precision advertising, and biosecurity [1, 11].

A person's lifespan can be roughly divided into the childhood or formative stage and the aging or adulthood stage. During the first stage of human life, from birth to adulthood, the shape of the face changes due to craniofacial growth. Craniofacial investigations show that the form of human faces changes from circular to oval [2], which results in the shift of fiducial landmarks on the face. The most obvious change in the second stage of human life, from maturity to old age, is skin aging (texture change); however, a minor change in shape also occurs. In [2], the aging process is represented by eleven facial features closely related to aging. Figure 1 lists some of these characteristics, including the corners of the mouth, nose, and eyes. According to biology, as people age, their skin becomes rougher, darker, thinner, and less elastic due to gravity and collagen insufficiency, forming spots and wrinkles [3]. These changes in texture and shape due to aging must be extracted to utilize the human age estimation system.



**Fig. 1.** Eleven skin areas for aging process.

For more than 20 years, age estimation from face images has been explored. Nevertheless, it is an incredibly challenging field for scholars. Numerous researchers added new approaches to the current methods to address the inadequate dataset and the similarity in photos for nearby ages. They established new methods like conditional probability neural network (CPNN) and IIS-Learning from Label Distribution (IIS-LLD) [9]. Attention modules like Convolutional Block Attention Module (CBAM), Squeeze-and-Excitation Networks (SENet) [4], and several state-of-the-art deep learning algorithms like DenseNet [5] and MobileNetV2 [6] have all been applied to image classification in recent years. In this work, we attempted to classify human age from faces using a variety of existing CNN architectures using an attention module, which helps to extract more significant features from the human face.

This is a representation of our major contributions to this work:

- (i) Build a scratch CNN model with attention module to estimate single human age from their facial expression.
- (ii) Create a combined dataset to increase the number of images for every age, which helps the model to obtain greater generalization ability.
- (iii) Evaluate various deep learning models based on transfer learning on our merged dataset.

## 2 Related Research

The assessment of human facial age has been the subject of extensive research over the last few decades. The oldest research in this area is found in [7], based on an investigation of skin wrinkles and the theory of craniofacial growth. They divide faces into three age groups: babies, teens, and senior adults. The Active Appearance Model (AAM) [8], a statistical method for modeling face images, is frequently used to depict facial appearance. AAM approach, which outperforms anthropometric models, takes textural and geometric models into account to deal with any age.

The authors of [9] proposed label distribution for age estimation and considered each picture connected with the label distribution, namely the Gaussian and triangular distribution, to increase label image for age estimation. They proposed two algorithms, CPNN and IIS-LLD achieved a better outcome for estimating the age of a human face. Learning algorithms require a significant amount of data. However, numerous images are on social media sites and labeled weekly. In [10], the authors suggested an age-difference learning method to use these weekly labeled photos through deep convolutional neural networks using label distribution.

Recently, deep learning-based algorithms gained much attention from researchers as they are very powerful in automatically extracting more accurate features. Still, a huge amount of data are needed to feed deep learning-based algorithms. In [11], Cross-Dataset CNN (CDCNN) was proposed to utilize low-quality images using a cross-dataset learning method with high-quality images. They used VGG-16 architectures pre-trained on ImageNet for training images. In this paper, we utilize this method to overcome insufficient data. In [12], the attention mechanism played an important role in natural language processing (NLP), now widely used in computer vision, showing a prominent result in computer vision. Different attention module is already used in computer vision for image classification, recognition, and segmentation [4]. We utilize the self-attention mechanism in our research through a deep convolution neural network.

## 3 Materials and Proposed Methodology

In this section, we discussed the detail of our proposed methodology. Figure 2 illustrates its structure, which consists of multiple convolution layers, average pooling layers, an attention module, two dense layers, and a dense output layer.

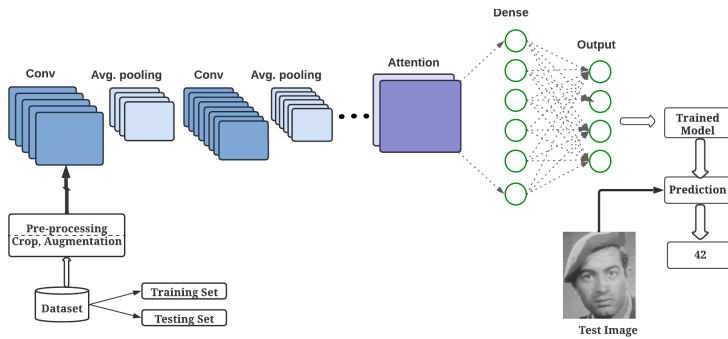


Fig. 2. Proposed age classification model.

### 3.1 CNN Architecture

**Transfer Learning.** Deep learning algorithms require enormous amounts of data for outstanding performance. Transfer learning is helpful for deep learning algorithms to generalize patterns from large benchmark datasets like ImageNet and COCO. This method is especially beneficial when the dataset is relatively small. In experiments, models such as DenseNet201 and MobileNetV2 trained on ImageNet were utilized.

**DenseNet.** DenseNet [5] uses skip connections to reduce vanishing gradients and improve performance. Every layer communicates with every other layer, giving its feature map to layers above it and receiving input from layers below. This condensed network makes training more uncomplicated and efficient, and it has fewer parameters than ResNet because features are combined.

**MobileNetV2.** MobileNetV2 [6] is a lightweight convolutional neural network designed to train with fewer parameters, making it highly beneficial for devices with limited capacity. It is more effective than MobileNetV1 because it has a  $1 \times 1$  conv across layers and a skip connection between two  $1 \times 1$  layers, which lowers the number of parameters (30% lower than MobileNetV1) and operations ( $2 \times$  lower than MobileNetV1) needed.

**CNN from Scratch.** After analyzing various transfer learning-based models, Convolution Neural Network (CNN) from scratch was also introduced to determine the human age from face images to obtain the lower mean absolute error. The scratch model is mainly built using four convolutional layers (32, 64, 128, and 256 filters), two dense layers, and a dense output layer. All four convolution layers utilized smaller filters ( $3 \times 3$ ).

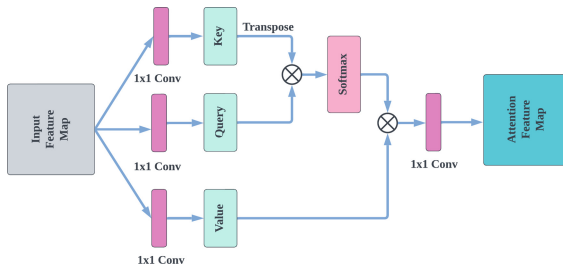
**CNN from Scratch with Attention.** An attention module was added at the end of the global average pooling of the scratch CNN model to reduce mean absolute error more.

### 3.2 Attention Module

The attention mechanism facilitates concentrating on a particular region of an image. This study used self-attention, which considers the entire image to generate attention weight. Scaled Dot-Product was chosen as the attention-scoring function, which performs element-wise dot-product of query and key vectors. The output is then normalized using the softmax function to remove negative weights. This results in a relationship where a pixel value or patch is more important than others. The procedure described above can be expressed as follows:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

where  $\sqrt{d_k}$  represents dimension of the query vector  $Q$  and key vector  $K$ , and  $V$  represents the value vector. The self-attention module's architectural layout is shown in Fig. 3.



**Fig. 3.** Self attention module where  $\otimes$  stands for matrix multiplication.

### 3.3 Model Design and Tuning

CNN from scratch with attention has been selected as our foundation model. An input of a 2D convolution layer paired with a 2D average pooling layer where the number of filters was 32 and the input shape of the image was  $200 \times 200$ . Three pairs of 2D convolutional layers paired with a 2D average pooling layer where the number of filters was 64, 128, and 256 were used as feature extractors. A  $3 \times 3$  kernel is used in all four convolution layers, Relu as the activation function, and pool size was  $2 \times 2$  for 2D average pooling. A 2D global average pooling was placed right after the final 2D average pooling layer. Figure 4 illustrates the architecture of the scratch CNN model.

In order to adjust the average weight of the feature before passing it to the dense layer based on the most significant feature located in the input image for a given problem domain, a self-attention module is added at the end of the 2D global average pooling and before the dense layer. Two Dense layers with several hidden units, 256 and 128, respectively, were used. The softmax activation in the output layer was used to create output probability distributions for human age classes.



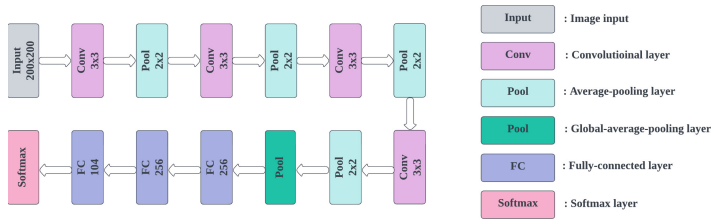


Fig. 4. Scratch CNN model architecture.

## 4 Result and Observation

The experiments were conducted using a computer with an Intel Core-i5 7th generation processor with 8 GB of RAM and a 4 GB dedicated graphics card. The software used was Python 3.7 and Keras 2.11.0.

### 4.1 Dataset and Preprocessing

We develop our dataset by merging two existing datasets, namely, UTKFace<sup>1</sup> and Facial-age.<sup>2</sup> Sample images of the combined dataset are shown in Fig. 5. A total of 33,486 labeled facial images were combined to classify human age from age 1 to 116, as shown in Fig. 6. We had the maximum number of images for the age of 26. We divided our dataset into 70:30 for training (23,440) and testing (10,046). All images were resized to  $228 \times 228 \times 3$  pixels to ensure all the dataset images were the same in size. Then randomly cropped to  $220 \times 220 \times 3$  pixels from the center of the image to make the model more vigorous even if the image is partially seen. As both DenseNet201 and MobileNetV2 were pre-trained with ImageNet, the combined dataset's RGB images ( $200 \times 200 \times 3$ ) were used to train models. In the other two models from scratch, with and without the attention module, images were converted in grayscale ( $200 \times 200 \times 1$ ) to reduce the effect of lighting in color RGB images.

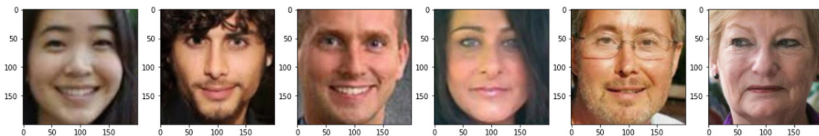
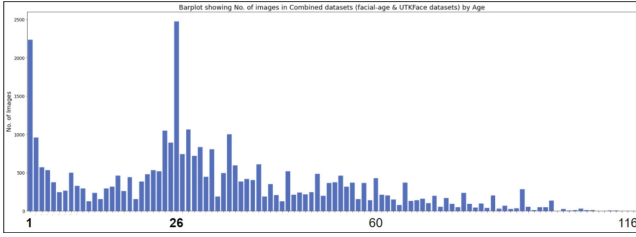


Fig. 5. Sample of our merged dataset of UTKFace and Facial-age.

<sup>1</sup> <https://susanqq.github.io/UTKFace/>.

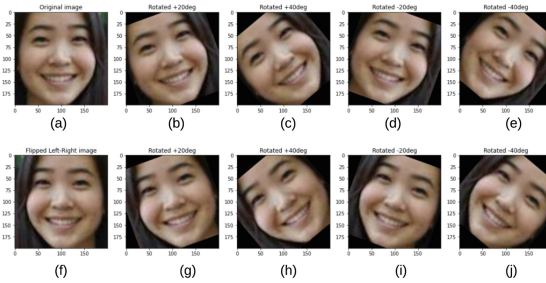
<sup>2</sup> <https://www.kaggle.com/datasets/frabbisw/facial-age>.



**Fig. 6.** Distribution of our merged dataset.

### 4.2 Augmentation Techniques

One of the main disadvantages of deep learning algorithms is the enormous amount of data they demand during the training phase. The model can generalize well with a huge dataset because a learning algorithm must adjust many parameters. Therefore, several augmentation approaches enhance the number of face images in the training set. Since each image was multiplied by ten, our train images went from 23,440 to 234,400. Figure 7 shows the visual representation of a single image after applying augmentation strategies.



**Fig. 7.** Example of augmentation: **a** original image, **b** rotated + 20°, **c** rotated + 40°, **d** rotated - 20°, **e** rotated - 40°, **f** horizontal\_flip, **g** rotated + 20°, **h** rotated + 40°, **i** rotated - 20°, **j** rotated - 40°.

Table 1 lists various image enhancement methods for our dataset, including horizontal flipping, 20° and 40° clockwise and anti-clockwise rotation.

### 4.3 Performance Evaluation and Hyperparameters Tuning

Four CNN models were used to evaluate the performance of the suggested model: MobileNetV2 and DenseNet201, both pre-trained via transfer learning, and two CNN models built from scratch, one of which used an attention mechanism. Mean absolute error (MAE) was chosen as the evaluation metric for age estimation.

**Table 1.** Different augmentation techniques on the combined dataset.

Augmentation technique	Parameter
horizontal_flip	True
rotation_clockwise_20degree	True
rotation_anti-clockwise_20degree	True
rotation_clockwise_40degree	True
rotation_anti-clockwise_40degree	True

**Table 2.** Evaluation of our various models on the merged dataset.

Model	MAE
DenseNet210	5.557
MobileNetV2	4.581
CNN from scratch	3.98
CNN from scratch with attention	<b>3.515</b>

The average of the difference between the predicted age and the actual age is known as the MAE, and it is defined as follows:

$$\text{MAE}(y, \hat{y}) = \frac{\sum_{i=0}^{N-1} |y_i - \hat{y}_i|}{N} \quad (2)$$

where  $N$  denotes the total number of data samples,  $y_i$  denotes the actual age, and  $\hat{y}_i$  denotes the predicted age of the  $i$ -th sample.

We compared our four models over 50 epochs to determine the best model. Table 2 shows the corresponding MAE for each model. CNN from scratch with attention has the lowest MAE among the four models.

We tested different batch sizes (128, 64, 32) and found that using a batch size of 128 resulted in better outcomes for the CNN model built from scratch with attention. See Table 3 for results.

Adam and stochastic gradient descent (SGD) optimizers were tested with different learning rates on the best model, and their mean absolute errors were compared. The results showed in Table 4 that the Adam optimizer with a learning rate of 0.0001 had better performance.

A comparison of our model performance with the state-of-the-art result is shown in Table 5, which lists the model's performance with the state-of-the-art paper. Our proposed model with attention outperforms with the lowest MAE, 3.515. Our deep-learning model accurately estimates ages from the visualization in Fig. 8. It contrasts predicted ages with actual ages on the test set.

**Table 3.** Comparison of various batch sizes on the combined dataset.

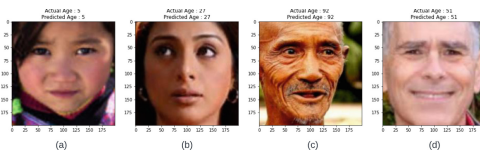
Model	Batch Size	MAE
DenseNet210	128	5.557
DenseNet210	64	5.559
MobileNetV2	128	4.581
MobileNetV2	64	4.584
CNN from scratch	128	3.982
CNN from scratch	64	3.980
CNN from scratch	32	3.984
CNN from scratch with attention	128	<b>3.515</b>
CNN from scratch with attention	64	3.518
CNN from scratch with attention	32	3.527

**Table 4.** Comparison of various optimizers on the combined dataset.

Model	Optimizer	Learning rate	MAE
CNN from scratch with attention	Adam	0.001	3.642
CNN from scratch with attention	Adam	0.0001	<b>3.515</b>
CNN from scratch with attention	SGD	0.001	3.593
CNN from scratch with attention	SGD	0.0001	3.611

**Table 5.** Evaluation of our proposed model with the state-of-the-art method.

Model	Mean absolute error
Gustafsson et al. [13]	4.65
Randomized bins [14]	4.55
MWR [15]	4.37
CNN from scratch with attention (ours)	<b>3.51</b>



**Fig. 8.** Age estimation example from the proposed model.

## 5 Conclusion and Future Work

In this work, experiments on the field of human age estimation, CNN model with attention mechanism, and also leveraging transfer learning from two pre-trained CNN models—DenseNet201 and MobileNetV2 were conducted by us. Images were resized and cropped as part of data pre-processing. A CNN model was chosen as a main feature extractor, and an attention module was used after the CNN model to adjust weight, focusing on the important patch of the image. Lastly, dense layers were used to predict the human age. From the experimental result, it was apparent that attention to the CNN model performs better than conventional CNN architecture for the same amount of data with the lowest MAE (3.515). So this proposed methodology has the ability to predict human age more precisely by increasing the dataset. Yet, the model suffers from class imbalance and needs to increase the amount of old people's image for a robust system. This experiment has created numerous opportunities to improve human age estimation from their facial expression, which is still a complex and hot research topic for researchers. In the future, a more precise algorithm with an attention mechanism will be incorporated to identify human age classification more precisely in the wild for better real-world performance.




## References

1. Al-Shannaq, A.S., Elrefaei, L.A.: Comprehensive analysis of the literature for age estimation from facial images. *IEEE Access* **7**, 93229–93249 (2019)
2. Angulu, R., Tapamo, J.R., Adewumi, A.O.: Age estimation via face images: a survey. *EURASIP J. Image Video Process.* **2018**(1), 1–35 (2018)
3. Fu, Y., Guo, G., Huang, Thomas S.: Age synthesis and estimation via faces: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(11), 1955–1976 (2010)
4. Guo, M.-H., et al.: Attention mechanisms in computer vision: a survey. *Comput. Visual Media* **8**(3), 331–368
5. Huang, G., et al.: Densely connected convolutional networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017)
6. Howard, A.G., et al.: Mobilenets: Efficient Convolutional Neural Networks for Mobile Vision Applications. [arXiv:1704.04861](https://arxiv.org/abs/1704.04861) (2017)
7. Kwon, Y.H., Lobo, N.D.V.: Age classification from facial images. *Comput. Vis. image Understanding* **74**(1), 1–21 (1999)
8. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(6), 681–685 (2001)
9. Geng, X., Yin, C., Zhou, Z.-H.: Facial age estimation by learning from label distributions. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(10), 2401–2412 (2013)
10. Hu, Z., et al.: Facial age estimation with age difference. *IEEE Trans. Image Process.* **26**(7), 3087–3097 (2016)
11. Zhang, B., Bao, Y.: Cross-dataset learning for age estimation. *IEEE Access* **10**, 24048–24055 (2022)
12. Vaswani, A., et al.: Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30** (2017)

13. Gustafsson, F.K., et al.: Energy-based models for deep probabilistic regression. In: Computer Vision-ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XX 16. Springer International Publishing (2020)
14. Berg, A., Oskarsson, M., O’Connor, M.: Deep ordinal regression with label diversity. In: 2020 25th International Conference on Pattern Recognition (ICPR) (2020)
15. Shin, N.-H., Lee, S.-H., Kim, C.-S.: Moving window regression: a novel approach to ordinal regression. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2022)



# Recognition and Classification of Crop Images by Convolutional Neural Network of Hybrid Architecture

K. Tokarev<sup>1</sup> , N. Lebed<sup>1</sup>, and I. Yudaev<sup>2</sup>  

<sup>1</sup> Volgograd State Agricultural University, Universitetskiy Ave., 26, 400002 Volgograd, Russia

<sup>2</sup> Kuban State Agrarian University, Kalinina St. 13, 350044 Krasnodar, Russia  
zirochka2505@gmail.com

**Abstract.** The authors classify modern approaches to the use of computer tools of artificial intelligence and machine learning to solve problems of increasing the productivity of agrophytocenoses, identify and justify the most promising areas. A strategic analysis of the use of approaches to managing the productivity of agricultural systems using artificial intelligence, with the actualization of opportunities and threats, is carried out. A system of operational monitoring of agrocenoses using satellite and aerial photographs obtained from unmanned aerial vehicles is proposed for the subsequent classification of agrophytocenoses, compilation of their heat maps and decision-making by a computer system based on the data obtained in precision farming. The algorithm of the learning process of a convolutional neural network for the system of analysis of remote monitoring images developed by the authors has been developed. Experimental studies on the classification of images of agricultural crops based on remote monitoring by the convolutional neural network of hybrid architecture, evaluation of the results of its training and testing demonstrate high accuracy of classification of the marked data set of the categories “healthy vegetation”, “affected vegetation” and “soil”. The results obtained to improve the accuracy of object classification (1–2%) indicate the prospects of using the CNN for precision farming tasks. However, for the practical use of the obtained classifiers, further research is required to reduce the number of classification errors. This is especially true for the classification of the class “affected vegetation”. The most promising direction for further research is the development and implementation of intelligent systems for operational monitoring of the state of crops based on high-resolution aerial photographs obtained from unmanned aerial vehicles with the ability to visualize the values of vegetation indices on them.

**Keywords:** Agrophytocenosis productivity · Artificial intelligence · Convolutional neural networks · Remote monitoring

## 1 Introduction

A number of state programs are being implemented in Russia for the development of modern highly efficient agriculture: “The Strategy for the development of agricultural machinery of the Russian Federation until 2030”, the Federal Scientific and Technical

Program for the Development of Agriculture for 2017–2025 (“Agricultural machinery and equipment”, “Breeding and seed production”), the Doctrine of Food Security of the Russian Federation”, the Departmental project “Digital Agriculture. One of the main objectives of the programs is the transition to digital agriculture, precision farming, and the active use of digital technologies. According to preliminary estimates, this will increase the contribution of agriculture to the country’s economy to 5.9 trillion rubles by 2024, increase the export revenue of agricultural organizations, and significantly increase the efficiency of agricultural production.

Artificial intelligence technologies are of particular importance among these technologies. Artificial intelligence plays an important role in managing the information lifecycle, including data processing, information flow management and knowledge.

The transition to a highly productive agricultural economy based on precision farming methods requires the use of artificial intelligence to support decision-making in the management of agricultural production using information technology and mathematical modeling of climatic, agrobiological and organizational and technological factors.

It is required to identify by artificial intelligence methods the regularities of the formation of agrobiocenosis productivity in conditions of insufficient heat and moisture supply, plant diseases and pests, weeds and other negative factors, improving the methods of recognition, analysis and processing of aerial photographs of agrobiocenoses for operational management of agricultural production [1–5].

The scientific novelty of the research consists in the creation of intelligent decision support algorithms to increase the productivity of agrophytocenoses based on the recognition and classification of aerial photographs of crops based on convolutional neural network algorithms of hybrid architecture, which will allow the scientist-agronomist to determine the most significant parameters for optimizing the placement of crops, monitor defective areas of crops of various nature, calculate NDVI indices in the real time, implement operational agrotechnical measures, etc.

## 2 Materials and Methods

Based on the analysis of the literature data [6–11], as well as our own research, we have identified the current directions of using approaches to managing the productivity of agrophytocenoses using artificial intelligence (Table 1).

As can be seen from Table 1, the use of artificial intelligence covers all levels of agricultural production planning, allowing for high-precision monitoring, as well as making operational decisions independently, providing autonomous management of agrophytocenosis productivity management.

It is necessary to highlight the most promising approaches to the use of artificial intelligence in agriculture:

1. Operational monitoring of agrocenoses using satellite and data obtained from unmanned aerial vehicles. This approach provides real-time search, processing and visualization of problem areas of crops in precision farming with the help of specialized libraries, followed by an increase in the accuracy of information processing.;



**Table 1.** Directions for using approaches to managing the productivity of agricultural systems using artificial intelligence

Satellite monitoring, using unmanned aerial vehicles, sensors	Computer vision technologies and audio analysis	Automated technologies and tools
<p>Diagnostics of pathologies and diseases of agricultural plants and animals; monitoring of soils for the optimal amount of trace elements necessary for the cultivation of high-quality crops; forecasting of natural and climatic conditions and taking appropriate measures based on this; operational monitoring of agrocnoses using satellite and aerial photographs obtained from unmanned aerial vehicles for subsequent classification of agrocnoses, compilation of their heat maps and subsequent decision-making by the system based on the data obtained</p>	<p>Monitoring the activities of animals in order to minimize their stress and take operational measures in case of critical situations; automation of agricultural technical systems, which allows taking appropriate measures in real time in case of a sharp change in natural and climatic conditions; monitoring of the state of nutrient media/plant or animal object in closed ecosystems to prevent single and mass infection using computer vision technologies</p>	<p>Technical automation of agricultural processes and phenomena, which allows, with the accumulation of relevant data, to optimize the standard procedures performed, accelerate sowing and harvesting operations, eliminate human heavy manual labor; treatment of plants and animals with substances dangerous to human health and life; urbanized crop production performed by vertical high-tech fully automated farms located inside buildings on the territory of cities; the system of risk forecasting and diagnostics of production processes in the management of agricultural enterprises of various capacities based on the accumulated knowledge of the system about the action in similar situations by a specialist</p>

2. Monitoring of the state of nutrient media/plant or animal object to prevent single and mass infection using computer vision technologies. This monitoring is extremely necessary in case of prevention of bacterial or mycotic infection with biotechnological methods of agriculture, since such an infection in a short time is able to infect the entire material of plant cells or tissues, which will lead to the rejection of a number of samples and reduce the yield of finished crop production. In addition, in the worst case, infection with bacterial or mycotic infection of the entire sterile chamber or laboratory room is not excluded, which will not allow the technology of microclonal reproduction of plants to be implemented until the contamination problem is completely solved and will lead to production losses;
3. Predictive analytics in combination with invariant data analysis tools in the management of agricultural enterprises of various capacities. This approach will ensure effective management of the company's production processes by predicting the behavior

of the system, as well as expanding the options for solving problems by the system itself by accumulating data from training while monitoring the actions of human specialists and past events [12–18].

However, despite the unconditional advantages of applying modern approaches to the use of artificial intelligence in the agro-industrial complex, there are weaknesses, as well as threats when implementing such systems (Table 2).

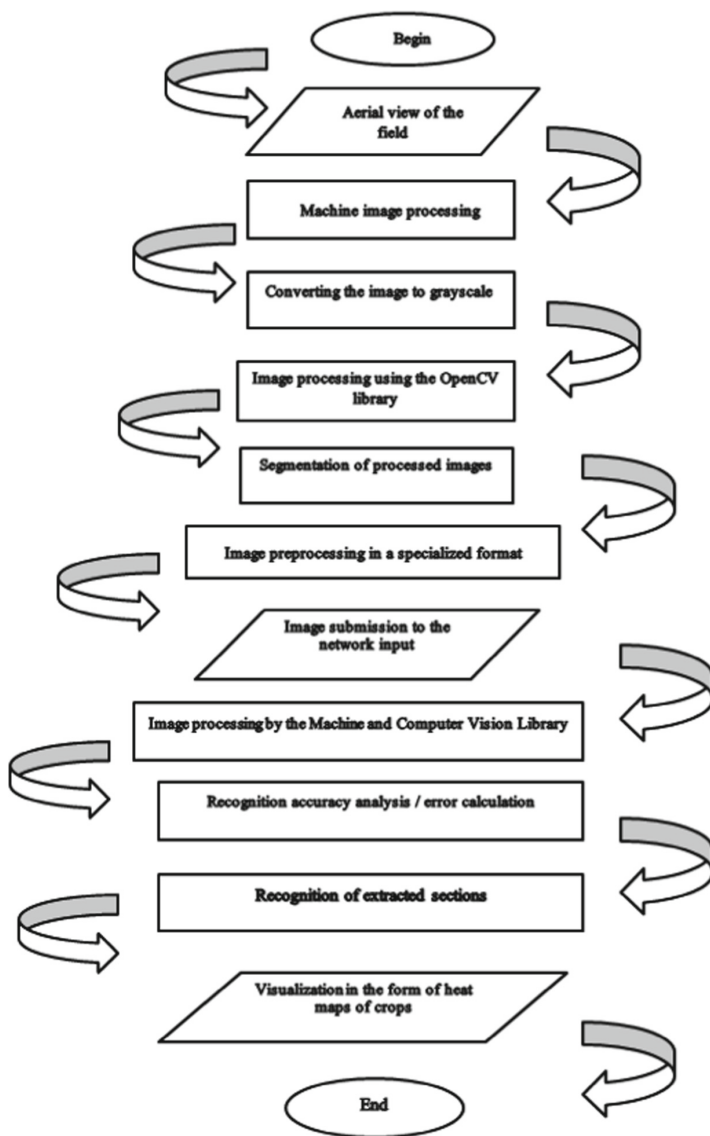
**Table 2.** Analysis of factors of the external and internal environment, as well as opportunities and threats of using artificial intelligence to solve problems of increasing the productivity of agroecosystems

Strength	Weakness
Improving labor efficiency in agricultural organizations using AI-technologies	The need for long-term research and significant investments in the development of AI-technologies for agriculture
Improving the efficiency of management decisions, as well as increasing the level of knowledge and access to information	The duration of AI-technologies entering the market, the complexity of determining the commercial effectiveness of these technologies
Increasing the attractiveness of the industry for young highly qualified specialists	The need to process huge amounts of data, energy costs and expensive digital equipment
Opportunity	Treats
Attracting highly qualified specialists to work in the field. It is achievable by the development of urbanized crop production inside heated vertical farms inside multi-storey buildings of cities, as well as work with remote access	Lack of methodological recommendations on the introduction of digital intellectual technology in the agro-industrial sector. Lack of a training system capable of mastering digital technologies in agriculture
A significant increase in progress in the development of AI-technologies in agriculture based on machine learning, the use of big data, neural networks, etc	There is no desire of agricultural producers to switch to new technologies for conducting highly productive agricultural production. The absence of a domestic component base

The most promising method of applying the approach to managing the productivity of agrophytocenoses through AI-technologies is the use of artificial neural networks, in particular Convolutional Neural Networks (CNN), belonging to the third generation of neuromodels. Convolutional neural networks (CNN) and deep convolutional neural networks (DCNN) are very different from other types of networks. They are usually used for image processing, less often for audio.

This method is also applicable for operational monitoring of agrocenoses using satellite and aerial photographs obtained from unmanned aerial vehicles for subsequent classification of agrocenoses, compilation of their heat maps and subsequent decision-making by the system based on the data obtained [19–24].

The algorithm developed by the authors in the course of research (Fig. 1) classification of aerial photographs of crops, identification and analysis of their condition with

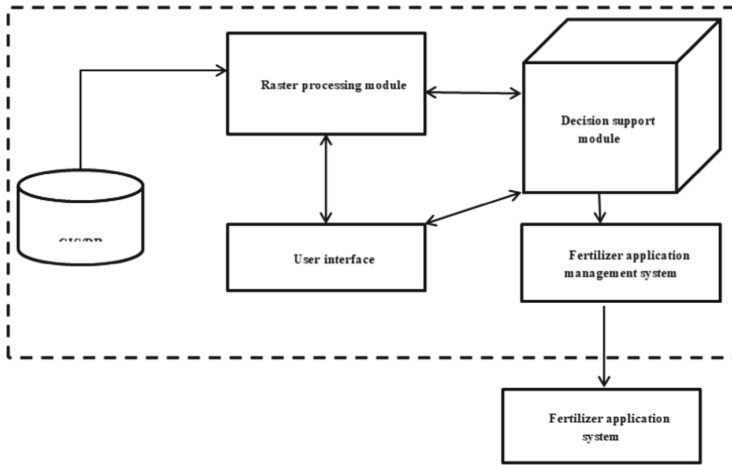


**Fig. 1.** Algorithm for classification of aerial photographs of agricultural crops

the possibility of localization of extracted sites, allowed to implement a software system based on a web interface with the ability to monitor crops online using mobile devices.

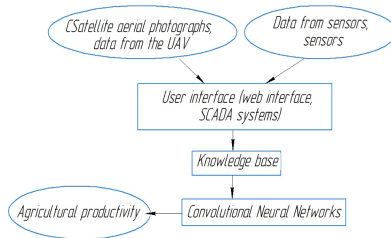
The conceptual scheme of the hardware and software complex for increasing the productivity of agrophytocenoses, the design and software implementation of which is provided during the research is shown in Fig. 2.

In the course of research on the problems of increasing bioproductivity, an algorithm for operational monitoring of agrocenoses using satellite and aerial photographs obtained



**Fig. 2.** The conceptual scheme of the hardware and software complex for increasing the productivity of agrophytocenoses

from unmanned aerial vehicles is proposed for subsequent classification of agrocenoses, compilation of their heat maps and subsequent decision-making by the system based on the data obtained (Fig. 3).



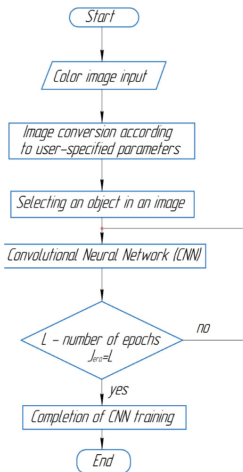
**Fig. 3.** Scheme of the operational system monitoring of agrocenoses

A block diagram of the neural network learning process for the system of remote sensing and satellite monitoring images analysis proposed by the authors is shown in Fig. 4. As an example of how the system works with images, real satellite photographs of crops are presented.

The software implementation of the process of forming training and control samples was carried out by means of the Python programming language using open libraries of deep machine learning for the analysis and subsequent classification of images [9–11]. A fragment of the listing of the source code of the module for the formation of training and control samples is shown in Fig. 5.

### 3 Results and Discussion

On a common database of class images: training + validation sample.



**Fig. 4.** The proposed algorithm (flow diagram) for training (on the example of processing satellite and aerial photographs obtained from unmanned aerial vehicles for the analysis and classification of agrocenoses)

```

PATH_DATASET = ''

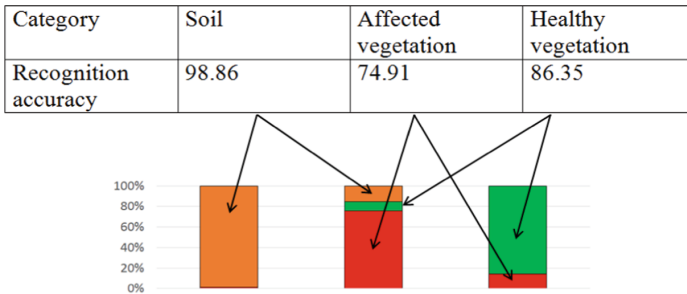
frame = cv2.imread("Image_test_1.jpg")
img = cv2.imread('Image_test_1.jpg',0)
equ = cv2.equalizeHist(img)
res = np.hstack((img,equ)) #stacking images side-by-side

gray = cv2.cvtColor(frame, cv2.COLOR_BGR2GRAY)
blurred = cv2.GaussianBlur(gray, (11, 11), 0)
thresh = cv2.threshold(blurred, 100, 255, cv2.THRESH_BINARY)[1]
thresh = cv2.erode(thresh, None, iterations = 16)
thresh = cv2.equalizeHist(thresh)
thresh = cv2.dilate(thresh, None, iterations = 12)
cnts = cv2.findContours(thresh.copy(), cv2.RETR_EXTERNAL, cv2.CHAIN_
cnts = cnts[0] if imutils.is_cv2() else cnts[1]
cnts = contours.sort_contours(cnts)[0]

cand_rect = []
cnts_small_area = []
for count, item in enumerate(cnts):
    rect = cv2.contourArea(item)
    if rect < 265 * 265 and rect > 20 * 20:
        (x1, y1, w, h) = cv2.boundingRect(item)
        y2 = y1 + h
        x2 = x1 + w
        match_part = frame[y1:y2, x1:x2]
        image_path = "match_part" + str(count) + ".jpg"
        image_data = tf.gfile.FastGFile(image_path, 'rb')
        plt.imshow(np.asarray(match_part), cmap=plt.cm.gray)
  
```

**Fig. 5.** A fragment of the source code of the program for the formation of training and control samples

During the testing of classifiers, a diagram is constructed that depicts the correspondence of the obtained classes to the real ones (Fig. 6).



**Fig. 6.** Classifier testing results

As a result of the study, the authors proposed and implemented using specialized machine learning libraries an algorithm for recognizing the state of vegetation with subsequent processing of aerial photographs by a convolutional neural network. The proposed algorithm makes it possible to reduce the influence of noise factors (lighting, solar glare, distortion of the color characteristics of images of pixel neighborhoods when vegetation and soil get into it) on the quality of the maps of the affected vegetation.

The conducted research and the results obtained will allow not only to quickly obtain information about the heterogeneities of soil and climatic zones, the general state of the fertility level of a particular field, but also, due to the intellectual component, will allow us to develop scientifically sound recommendations for improving soil fertility in the conditions of precision agriculture of the agriculture of the Russian Federation.

As a direction for further research, it is planned to improve the database and the program being developed to enable the analysis of time series of retrospective images in order to identify characteristics and intellectual assessment of the dynamics of plant development using neural network technologies.

## 4 Conclusion

The results obtained to improve the accuracy of object classification (1–2%) indicate the prospects of using the CNN for precision farming tasks. However, for the practical use of the obtained classifiers, further research is required to reduce the number of classification errors. This is especially true for the classification of the class “affected vegetation”.

**Acknowledgements.** The article is prepared with the financial support of the Russian Science Foundation, project №. 22-21-20041 and Volgograd region.

## References

1. Li, Q., Liu, J., Mi, X., et al.: Object-oriented crop classification for GF-6 WFV remote sensing images based on convolutional neural network. **25**(2), 549–558 (2021). <https://doi.org/10.11834/jrs.20219347>
2. Guo, Z., Qi, W., Huang, Y., et al.: Identification of crop type based on C-AENN using time series sentinel-1A SAR data. *Remote Sens.* **14**(6) (2022). <https://doi.org/10.3390/rs14061379>

3. Khamparia, A., Saini, G., Gupta, D., et al.: Seasonal crops disease prediction and classification using deep convolutional encoder network. **39**(2), 818–836 (2020). <https://doi.org/10.1007/s00034-019-01041-0>
4. Khryashchev, V., Pavlov, V., Priorov, A., Kazina, E.: Convolutional neural network for satellite imagery. *Conf. Open Innov. Assoc. FRUCT* **22**, 344–347 (2018)
5. Garge, N.R., Bobashev, G., Eggleston, B.: Random forest methodology for model-based recursive partitioning: the mobForest package for R. *BMC Bioinform.* **14**, 125 (2013)
6. Kurmi, Y., Saxena, P., Kirar, B.S., et al.: Deep CNN model for crops' diseases detection using leaf images. *Multidimensional Syst. Signal Process.* (2022). <https://doi.org/10.1007/s11045-022-00820-4>
7. Prottasha, S.I., Reza, S.M.S.: A classification model based on depthwise separable convolutional neural network to identify rice plant diseases. **12**(4), 3642–3654 (2022). <https://doi.org/10.11591/ijece.v12i4.pp3642-3654>
8. Turkoglu, M.O., D'aronco, S., Schindler, K., et al.: Crop mapping from image time series: deep learning with multi-scale label hierarchies. **264**, 112603 (2021). <https://doi.org/10.1016/j.rse.2021.112603>
9. Tokarev, K.E.: Agricultural crops programmed cultivation using intelligent system of irrigated agrocoenoses productivity analyzing. *J. Phys. Conf. Ser.* **1801**, 012030 (2021)
10. Plant, R.E., et al.: Relationship between remotely sensed reflectance data and cotton growth and yield. *Trans. ASAE* **43**(3), 535–546 (2000)
11. Tokarev, K., Lebed, N., Prokofiev, P., Volobuev S., Yudaev, I., Daus, Y., Panchenko, V.: Monitoring and Intelligent Management of Agrophytocenosis Productivity Based on Deep Neural Network Algorithms. *Lecture Notes in Networks and Systems* 569, pp. 686–694 (2023)
12. Tokarev, K.E.: Raising bio-productivity of agroecosystems using intelligent decision-making procedures for optimization their state management. *J. Phys. Conf. Ser.* **1801**, 012031 (2021)
13. Petrukhin, V., Feklistov, A., Yudaev, I., Prokofiev P., Ivushkin D., Daus, Y., Panchenko, V.: Modeling of the Device Operating Principle for Electrical Stimulation of Grafting Establishment of Woody Plants. *Lecture Notes in Networks and Systems* 569, pp. 667–673 (2023)
14. Isaev, R.A., Podvesovskii, A.G.: Application of time series analysis for structural and parametric identification of fuzzy cognitive models. *CEUR Works. Proc.* **2212**, 119–125 (2021)
15. Churchland, P.S.: *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. MIT Press, Cambridge, USA (1986)
16. Aleksander, I., Morton, H.: *An Introduction to Neural Computing*. Chapman & Hall, London, UK (1990)
17. McCulloch, W.S., Pitts, W.A.: Logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133 (1943)
18. Ivushkin, D., Yudaev, I., Petrukhin, V., Feklistov, A., Aksenov, M., Daus, Y., Panchenko, V.: Modeling the Influence of Quasi-Monochrome Phytoirradiators on the Development of Woody Plants in Order to Optimize the Parameters of Small-Sized LED Irradiation Chamber. *Lecture Notes in Networks and Systems* 569, pp. 632–641 (2023)
19. Yudaev, I., Eviev, V., Sumyanova, E., Romanyuk N., Daus, Y., Panchenko, V.: Methodology and Modeling of the Application of Electrophysical Methods for Locust Pest Control. *Lecture Notes in Networks and Systems*, 569, pp. 781–788 (2023)
20. Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65**, 386–408 (1958)
21. Cheng, G., Li, Z., Yao, X., Guo, L., Wei, V.: Remote sensing image scene classification using bag of convolutional features. *IEEE Geosci. Remote Sensing Lett.* **14**(10), 1735–1739 (2017)
22. Bian, X., Chen C., Tian L., Du Q.: Fusing local and global features for high-resolution scene classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **10**(6), 2889–2900 (2017)

23. Kadhim, M.A., Abed, M.H.: Convolutional neural network for satellite image. In: *Classification Studies in Computational Intelligence*, 165–178 (2020)
24. Tokarev, K.E.: Overview of intelligent technologies for ecosystem bioproductivity management based on neural network algorithms. *IOP Conf. Ser. Earth Environ. Sci.* **1069**, 012002 (2022)
25. Lebed, N.I., Makarov, A.M., Volkov, I.V., Kukhtik, M.P., Lebed, M.B.: Mathematical modeling of the process of sterilizing potato explants and obtaining viable potato microclones. *IOP Conf. Ser. Earth Environ. Sci.* **786**, 012035 (2021)





# A Comparative Study of Deep Learning Algorithms and SARIMA Models for Forecasting Monthly Solar Radiation and UV Index: Case Study for Mauritius

Janvee Dabeedoal<sup>1</sup>, Ravindra Boojhawon<sup>1(✉)</sup>, Oomesh Gukhool<sup>1</sup>,  
and Deepanjal Shrestha<sup>2</sup>

<sup>1</sup> University of Mauritius, Reduit, Mauritius  
, [r.boojhawon@uom.ac.mu](mailto:r.boojhawon@uom.ac.mu)

<sup>2</sup> School of Business, Pokhara University, Lekhnath, Nepal

**Abstract.** Forecasting solar radiation and UV index accurately have proven in various studies to be crucial in this evolving world where solar energy is being exploited to its core for various applications. The objective of this paper is to apply different artificial neural network architectures to the modelling of the two monthly time series data aforementioned, with their structures and training processes fully described. A comparative analysis of the results, including that of a seasonal autoregressive auto moving model, are afterwards thoroughly carried out, which is finally followed by a forecast for the next twelve months. The optimal models for modelling solar radiation and UV are concluded to be the feedforward neural network, with a RMSE of 13.5 and the hybrid model of gated recurrent unit and long short term memory network, with a RMSE of 0.198 respectively for this study.

**Keywords:** Time series forecasting · Solar radiation · UV index · Artificial neural network

## 1 Introduction

In this era of advanced technology, renewable sources of energy, especially solar energy, are being heavily exploited by people all around the world for their lucrative properties to reduce our dependence on fossil fuels. Mauritius, a tropical island in the Indian Ocean that receives an abundance of solar radiation and UV light throughout the year, is no exception; several photovoltaic farms are being established in different regions of the country for the generation of electricity amongst other various uses. So, it is imperative to accurately model and predict these two components for the efficient management of these farms and also for the country's population to have a better insight into the intensity of the UV index since a high intensity (a UV index greater than 6) has scientifically been proven to be detrimental to the skin and eyes.

Several studies on the application of traditional statistical methods, such as exponential smoothing [1] and autoregressive moving average models [2], to the modeling and prediction of solar radiation [3,4] and UV index [5], have been carried out for many decades; it is quite recently that deep learning algorithms have been implemented for this particular purpose because the aforementioned statistical models have been shown by Sharma et al. [6] to be inaccurate due to the non-stationarity of the time series data corresponding to solar irradiation and UV index and that deep learning algorithms are better at capturing the non-stationarity property. For instance, Justin et al. and Mukhoty et al. (see [7,8]) have tried using different types of deep learning algorithms, with some of them being hybrid models, to model solar radiation and they each concluded that the hybrid model of stacked LSTM with principal component analysis and the encoder-decoder networks of LSTM are the most accurate models respectively for their research. As for the UV index, Latosińska et al. [9] has found that the feedforward neural network was the most accurate model. Moreover, it has been deduced by Leal et al. [10] that there exist linear and polynomial correlations between solar radiation and UV index.

The main emphasis of this paper lies in finding the model which best represents and predicts monthly solar radiation and UV index at a particular location in the northern region of Mauritius, Petit Raffray, through a comparative study involving different well-known artificial neural networks and the SARIMA models.

## 2 Deep Learning Algorithms

Deep learning algorithms comprise of primarily artificial neural networks having many layers of nodes or neurons interconnected with each other such that each node in a certain layer is connected to all nodes in the subsequent layer and so on. Innumerable software are available today that simplify their implementations to different applications despite their complex structures. Any ANN model is quintessentially made up of vertically stacked layers of nodes which should include the following, as discussed by Roza Dastres and Mohsen Soori [11]:

1. Input layer

All the independent variables are to be included in the model are added in this layer where the total number of neurons refers to the number of regressor variables in the model.

2. Output layer

It is the last layer predicting the value of the response variable(s), which is denoted by the number of nodes in that specific layer.

3. Hidden layer(s)

The intermediate layer(s) make up the hidden layer(s), helping in training the ANN to model and predict the dependent variables accurately.

Every connection from one node to the other in an ANN is assigned a certain weight defining its strength, a bias that is analogous to a line intercept and

therefore allows adjustments to be carried out in the nodes for better results, and thirdly an activation function that converts the linear relations of the combinations of the weights and biases to a non-linear one; it is this value that is passed onto the next node in the subsequent layer. It is to be noted that an activation function is, most of the times, a non-linear function, differentiable almost everywhere so that backpropagation, a crucial step in the training of the ANN, can be carried out (where the derivatives of the activation functions are required); it is normally used in the hidden layers and in the output layer to obtain the output of a certain node from one layer to afterwards be fed into another node of the next layer. The activation function chosen to train an ANN has an effect on its performance; so, it must be cautiously chosen. Some common activation functions, as studied by Feng et al. [12] are the sigmoid function, ReLU (Rectified Linear Unit), Tanh (Tangent Hyperbolic), and leaky ReLU.

### 2.1 Training of an ANN

**Forward propagation** and **backward propagation** are the main processes in training any ANN model available. However, some slight changes may take place in some networks.

For an ANN with  $L$  layers and  $n^{[l]}$  neurons in each layer, where  $l = 1, 2, 3, \dots, L$ , two distinct processes take place for every neuron  $j$  in a specific layer  $l$  corresponding to a training example:

1. A weighted sum of the activation values (output from the activation function) from the previous layer added to a bias is calculated as

$$z_j^{[l]} = w_j^{[l]T} a^{[l-1]} + b_j^{[l]}, \tag{1}$$

$1 \times 1 \qquad 1 \times n^{[l-1]} n^{[l-1]} \times 1 \qquad 1 \times 1$

$$z_j^{[l]} = \sum_{k=1}^{n^{[l-1]}} w_{j,k}^{[l]} a_k^{[l-1]} + b_j^{[l]}, \tag{2}$$

where  $n^{[l]}$  is the number of neurons in the following  $l$ th layer, and  $w_{j,k}^{[l]}$  is the weight from the  $k$ th neuron in the  $(l - 1)$ th layer to the  $j$ th neuron in the  $l$ th layer. Additionally, it should be taken into consideration that the activation vector  $a^{[0]}$  refers to the values of the explanatory variables input in the neurons of the input layer of the neural network, also denoted by the vector  $X$  in some cases.

2. Then, an activation function, that is the same for all neurons in one layer, converts this value to an output which, when exceeding some threshold, activates the neuron which then passes the output to the next neuron found in the next layer. The activation value is calculated as

$$a_j^{[l]} = g^{[l]}(z_j^{[l]}), \tag{3}$$

$$a_j^{[l]} = g^{[l]} \left( \sum_{k=1}^{n^{[l-1]}} w_{j,k}^{[l]} a_k^{[l-1]} + b_j^{[l]} \right), \tag{4}$$

where  $g^{[l]}$  is a certain activation function for the  $l$ th layer and it is to noted that each layer can have different activation functions.

Backpropagation is the process during which the optimal weights and biases for each node in all the layers are to be found such that the loss function for the neural network is minimized, ultimately leading to an accurate estimate(s) for the response variable(s). Gradient descent is the key to this; it needs the derivatives of the cost function with respect to all the biases and weights for the different layers corresponding  $m$  training examples, given as

$$\frac{\partial C}{\partial w^{[l]}} = \frac{1}{m} \sum_{n=1}^m \delta^{(n)[l]} a^{[l-1]T}, \tag{5}$$

$$\frac{\partial C}{\partial b^{[l]}} = \frac{1}{m} \sum_{n=1}^m \delta^{(n)[l]}, \tag{6}$$

where

$$\delta^{[l]} = \frac{\partial \mathcal{L}}{\partial b^{[l]}} = w^{[l+1]T} \delta^{[l+1]} \odot g^{[l]'}(z^{[l]}). \tag{7}$$

and  $\odot$  refers to the hadamard product of the matrices and  $\delta^{[l]}$  the vector of errors associated with a certain layer  $l$ , where each entry in the vector corresponds to an error  $\delta_j^{[l]}$  for the  $j$ th neuron in the  $l$ th layer, which is defined as

$$\delta_j^{[l]} = \frac{\partial \mathcal{L}}{\partial z_j^{[l]}}, \tag{8}$$

Gradient descent, as studied by Ruder [13], is then worked out as follows

$$w^{[l]} \rightarrow w^{[l]} - \frac{\alpha}{m} \sum_{n=1}^m \delta^{(n)[l]} a^{[l-1]T}, \tag{9}$$

$$b^{[l]} \rightarrow b^{[l]} - \frac{\alpha}{m} \sum_{n=1}^m \delta^{(n)[l]}, \tag{10}$$

where  $\alpha$  is the learning rate.

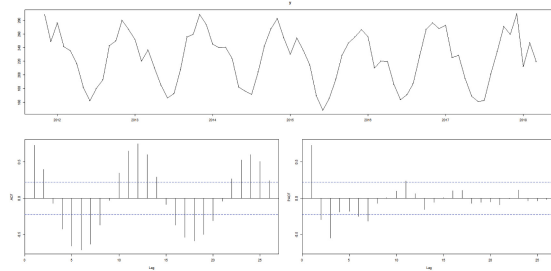
### 3 Simulation Results and Discussions

#### 3.1 Solar Radiation and UV Index Data

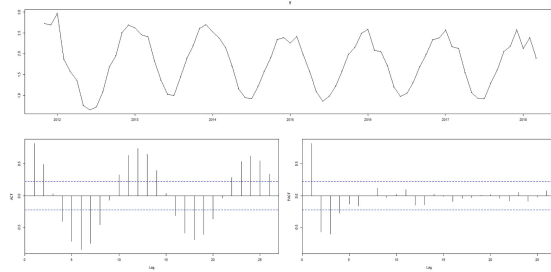
The solar radiation and UV index data, utilized in a comparative analysis of the performance of the different ANN, are both a monthly time series data spanning over a 110- month long period of **November 2011 to December 2020** for a village situated in the North of Mauritius, **Petit Raffray**; the monthly data were compiled from a large dataset of minute data for both variables. Since the latter are seasonal time series data, the traditional seasonal autoregressive

integrated moving average, SARIMA model, has also been included in this comparative study to determine if the neural network architectures are capable of outperforming the former.

The time series plots, along with their corresponding autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, for the monthly data sets corresponding to the two variables, are presented in Fig. 1. The presence of repeated patterns of significant lines at the different lags for the two ACF plots for each of the variable confirms the presence of seasonality in the two data sets.



(a) Time series, ACF and PACF plots for monthly solar radiation.



(b) Time series, ACF and PACF plots for monthly UV index.

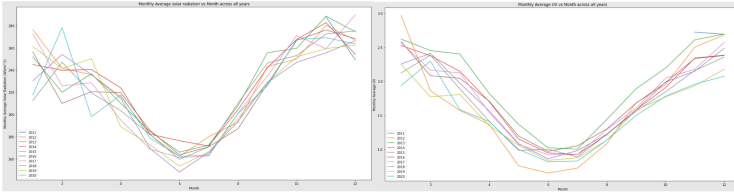
**Fig. 1.** Time series, ACF and PACF plots.

Moreover, we find that the two data sets seem to be fitting a **cosine model** from the time series plot and the seasonal plots depicted in Fig. 2. So, two cosine models, denoted by (11) and (12), are seen to approximately model the data in the least square sense.

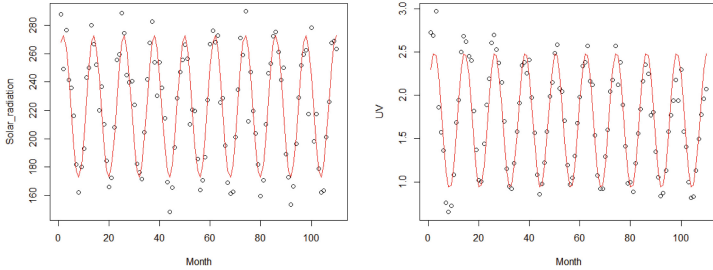
$$\text{Solar Radiation} = 222.57 + 50.16 \cos\left(\frac{\pi}{6}t + 0.0802\right), \tag{11}$$

and

$$\text{UV} = 1.709 + 0.7845 \cos\left(\frac{\pi}{6}t - 0.254\right), \tag{12}$$



(a) Seasonal plot for solar radiation. (b) Seasonal plot for UV index.



(c) Cosine model for solar radiation. (d) Cosine model for UV index.

**Fig. 2.** Seasonal plots and cosine models for both time series data.

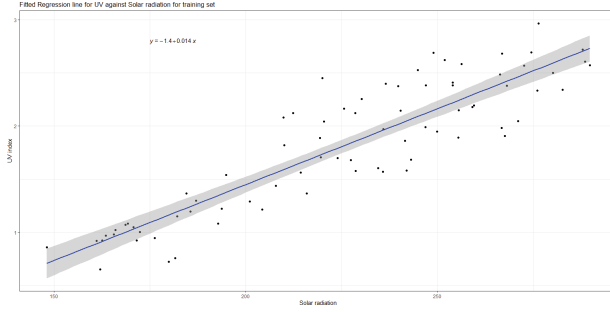
where  $t$  is the time in month.

It may also be of interest to find whether there exists a linear relation between solar radiation and UV index as we find their distributions to be similar from the seasonal and time series plots. A simple linear regression line is therefore fitted, with the UV index as the response variable and solar radiation as the independent variable. The data is divided into the training and test sets under the **70–30%** split. The equation of the fitted line acquired from the training set is

$$y = -1.4093 + 0.0143x, \tag{13}$$

where  $y$  is the UV index and  $x$  the solar radiation.

Both coefficients corresponding to the linear model are concluded to be significant due to their very small p-values;  $1.0e - 10$  for the intercept and  $2.0e - 16$  for the slope, where  $e$  refers to base 10. Moreover, the correlation of the linear relation being 0.8950 indicates a strong positive linear relationship as supported by the least square best fitted line is shown in Fig. 3. The accuracy of the line is evaluated by the test set, whose root mean squared error (RMSE) is 0.2865 and also a high coefficient of determination of 0.8914. So, we reach to the conclusion that the line of regression is indeed an accurate model.



**Fig. 3.** Fitted line of UV against solar radiation.

### 3.2 Result Analysis for Monthly Mean Solar Radiation and UV Index

Eight different deep learning models along with a SARIMA model have been implemented in modeling and predicting monthly solar radiation and UV index respectively. The evaluation metrics used to find the optimal model are the mean absolute mean error (MAE) and root mean squared error (RMSE). It is to be noted that the percentage split used in this case study is the **70–30%** split and the data is normalized between (0, 1) to make the training process easier and faster.

The SARIMA models which best fit the monthly solar radiation and UV index data, based on AIC model selection criterion are represented by (14) and (16) respectively.

SARIMA(1, 0, 0)(2, 1, 1)<sub>12</sub> for Solar Radiation:

$$\begin{aligned} &(1 - B^{12})(1 + 0.2490B)(1 + 0.4383B^{12} + 0.4619B^{24})y_t \\ &= -0.1629 + (1 - 0.5212B^{12})z_t, \end{aligned} \tag{14}$$

where,

$$\{z_t\} \sim \mathcal{N}(0, 140), \tag{15}$$

and monthly mean solar radiation is denoted by  $y_t$ .

SARIMA(1, 1, 0)(0, 1, 1)<sub>12</sub> for the log UV index (Box-Cox transformation has been implemented to ensure that the normality assumption of residuals is not violated):

$$(1 - B)(1 - B^{12})(1 + 0.4413B)y_t = (1 - 0.6528B^{12})z_t, \tag{16}$$

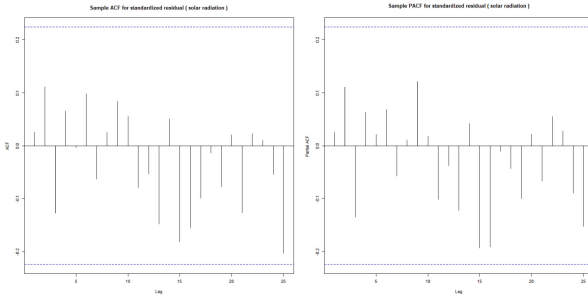
where,

$$\{z_t\} \sim \mathcal{N}(0, 0.009494), \tag{17}$$

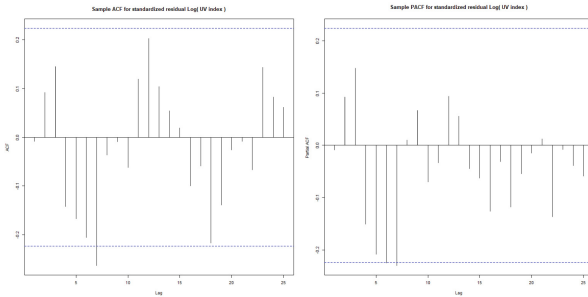
and monthly mean UV index is denoted by  $y_t$ .

To ascertain that the SARIMA models are in fact a good fit, the residual analysis for its training set need to be performed to check whether the different

assumptions of the residuals hold in this case. First, the ACF and PACF for the standardized residuals for each of the variables are worked out. Since no significant lag is present in the ACF or PACF for both solar radiation and UV index represented by Fig. 4, we conclude that their residuals are random and are confirmed to be noise.



(a) ACF and PACF for the residuals for solar radiation.



(b) ACF and PACF for the residuals for UV index.

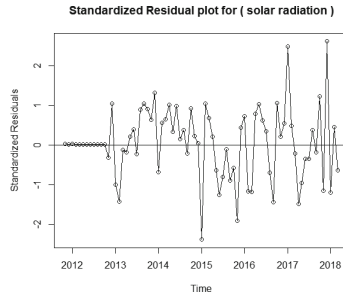
**Fig. 4.** ACF and PACF for the residuals.

To test for the normality of the residuals, Shapiro-Wilk’s test and Jarque Bera test are carried out for both time series data. Since the p-values corresponding to solar radiation for the Shapiro-Wilk’s and Jarque Bera tests are 0.539 and 0.5199 respectively (they are greater than 0.05), the null hypothesis of normality of the residuals is not rejected at 5% significance level and we hence conclude that the latter are normal. As for UV index, since the p-values for the Shapiro-Wilk’s and Jarque Bera tests are 0.0009795 and  $5.903e-05$  respectively (they are less than 0.05), the null hypothesis of normality is rejected, indicating that the residuals are not normally distributed. As the normality assumption is violated, this SARIMA model is concluded as not a good fit and thus modeling and predicting the monthly UV index could be less reliable.

To test if the variance of the residuals for solar radiation is constant, the standardized residual plot is used, as shown in Fig. 5. Since the residuals are



scattered and do not have an underlying pattern, we conclude that the variance of the residuals is constant. Finally, to test for the independence of its residuals, the runs test is carried out where the p-value is  $0.727 > 0.05$ , and therefore the null hypothesis of independence is not rejected at 5% level of significance and we conclude that the residuals are in fact independent of each other. As all the assumptions of the SARIMA model for solar radiation are satisfied, we reach the conclusion that it is a good model.



**Fig. 5.** Standardised residual plot for solar radiation.

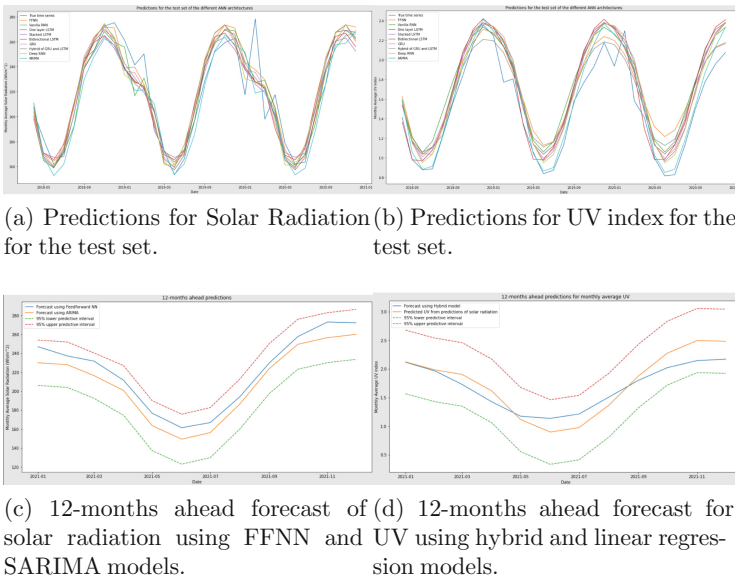
The accuracy metrics for the test set for each of the two time series data are summarized in Table 1.

**Table 1.** Accuracy metrics for solar radiation and UV index.

Model	Solar radiation		UV index	
	RMSE	MAE	RMSE	MAE
Feedforward NN	13.49315	9.76104	0.22893	0.20078
Simple RNN	15.39646	11.06023	0.24471	0.22742
One layer LSTM	15.65280	12.01297	0.22118	0.18997
Stacked LSTM	14.02874	10.13109	0.21808	0.19442
Bidirectional LSTM	14.50322	10.43551	0.23634	0.21381
GRU	14.88307	11.57955	0.22347	0.19747
Hybrid of GRU and LSTM	14.83550	11.13953	0.19824	0.17603
Deep RNN	15.80181	11.85836	0.20515	0.18059
SARIMA	17.39390	12.38790	0.14993	0.11233

Each of the models implemented in this study was then used to predict solar radiation and UV index for the test set, i.e., the interval between April 2018 and December 2020, a visual representation provided in Fig. 6. The optimal models for monthly solar radiation and UV index are concluded to be FFNN and

the hybrid model of GRU and LSTM respectively even though the RMSE and MAE for the SARIMA corresponding to UV index were the least. The optimal models were then applied to predict solar radiation and UV index for the next twelve months. For solar radiation, the SARIMA model was implemented for the forecast as well and for UV index, the linear relation obtained from (13) on the predicted values for the solar radiation for FFNN model was used for the forecast, as shown in Fig. 6. We observe that the forecasted UV index does not have a tendency to exceed an index of 4 and also from the original data in Fig. 1, we infer that the UV index is not greater than 3. This indicates that Petit Raffray normally receives UV radiation of low risk of harm.



**Fig. 6.** Predictions for the test set and forecast for 12 months ahead.

## 4 Conclusions

Nine popular ANN models together with the SARIMA models were applied to fit two distinct time series data corresponding to solar radiation and UV index. The most optimal model for monthly solar radiation and UV index in this study was found to be the FFNN and the hybrid model of GRU and LSTM respectively, which were then used for a year ahead forecast. From the results, we were able to infer that the UV index is always less than 4, thereby concluding that any harm from UV radiation is quite low for Petit Raffray. Moreover, we infer that ANN models are indeed capable of modeling and predicting time series data accurately in this project.

As future study, it is proposed to model and predict solar radiation and UV index at higher resolutions such as at hourly or even minutely scale while including some other weather parameters as well. Furthermore, other models such as convolutional neural networks, and general regression neural networks, can be implemented.

**Acknowledgements.** We are thankful to Mr. Silvio Chiara of the Meteorological Station of Petit Raffray for providing the solar radiation and UV index data.

## References

1. Huang, G.Y., Lai, C.J., Pai, P.F.: Forecasting hourly intermittent rainfall by deep belief networks with simple exponential smoothing. *Water Resour. Manage.* **36**(13), 5207–5223 (2022)
2. Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J.: Intelligent computing & optimization. In: *Proceedings of the 5th International Conference on Intelligent Computing and Optimization 2022 (ICO2022)*, vol. 569, Springer Nature (2022)
3. Doorga, J., Rughooputh, S., Boojhawon, R.: Geospatial Modelling of Solar Radiation Climate. In *Geospatial Optimization of Solar Energy: Cases from Around the World*, pp. 19–33. Springer International Publishing, Cham (2022)
4. Doorga, J.R., Rughooputh, S.D., Boojhawon, R.: Modelling the global solar radiation climate of Mauritius using regression techniques. *Renew. Energy* **131**, 861–878 (2019)
5. Fioletov, V., Kerr, J.B., Fergusson, A.: The UV index: definition, distribution and factors affecting it. *Can. J. Publ. Health* **101**, I5–I9 (2010)
6. Sharma, V., Yang, D., Walsh, W., Reindl, T.: Short term solar irradiance forecasting using a mixed wavelet neural network. *Renew. Energy* **90**, 481–492 (2016)
7. Justin, D., Concepcion, R.S., Calinao, H.A., Alejandrino, J., Dadios, E.P., Sybingco, E.: Using stacked long short term memory with principal component analysis for short term prediction of solar irradiance based on weather patterns. In: *2020 IEEE Region 10 Conference (TENCON)*, pp. 946–951. IEEE (2020)
8. Mukhoty, B.P., Maurya, V., Shukla, S.K.: Sequence to sequence deep learning models for solar irradiation forecasting. In: *2019 IEEE Milan PowerTech*, pp. 1–6. IEEE (2019)
9. Latosińska, J.N., Latosińska, M., Bielak, J.: Towards modelling ultraviolet index in global scale. Artificial neural networks approach. *Aerosp. Sci. Technol.* **41**, 189–198 (2015)
10. Leal, S.S., Tíba, C., Piacentini, R.: Daily UV radiation modeling with the usage of statistical correlations and artificial neural networks. *Renew. Energy* **36**(12), 3337–3344 (2011)
11. Dastres, R., Soori, M.: Artificial neural network systems. *Int. J. Imag. Robot. (IJIR)* **21**(2), 13–25 (2021)
12. Feng, J., Lu, S.: Performance analysis of various activation functions in artificial neural networks. *J. Phys. Conf. Ser.* **1237**(2), 022030 (2019)
13. Ruder, S.: An overview of gradient descent optimization algorithms. [arXiv:1609.04747](https://arxiv.org/abs/1609.04747) (2016)



# A Study on Fault Classification and Location Using Supervised Machine Learning

Nanda Kumari<sup>(✉)</sup> and Channarong Banmongkol

Department of Electrical Engineering, Chulalongkorn University, Bangkok, Thailand  
6470201921@student.chula.ac.th, channarong.b@chula.ac.th

**Abstract.** Over the years supervised machine learning (SML) has proven its ability to identify patterns in labeled data sets and predict outcomes. The purpose of the paper is to develop an algorithm applying SML's classification and regression features to classify and locate faults in transmission and distribution lines utilizing MATLAB. The algorithm uses the RMS values of three-phase and zero sequence components measured at a single end. It combines classification techniques such as linear discriminant analysis (LDA) and neural networks (NN) with regression techniques like a decision tree (DT) and the least squares method (LS). The algorithm's effectiveness was assessed using IEEE 14 bus system with a classification accuracy of 100% and the maximum location accuracy of 97.26%.

**Keywords:** Linear discriminant analysis · Neural network · Decision tree · Supervised machine learning

## 1 Introduction

Transmission and distribution lines around the globe are subjected to various faults due to the following reasons: switching surges, insulation failure, snow, conducting path failure, excessive growth in the right of way, falling of trees, creepers on the towers and poles, sudden changes in load parameters at the customer end leading to short circuits, under/over current, under voltage, unbalanced phase voltage, trespassing of animals and often surge leads to fire, loss of service and damages the equipment.

The frequent faults cause wear and tear causing insulation failure, and the life span of the line and the substation equipment becomes a major concern. When such issues are not resolved, it leads to blackouts which cause major socio-economic crises. In the event to find the root cause of the blackouts the reference [1] classifies the root cause as natural, malicious, accidental, and cascading. The major sources of faults are equipment failures, excessive growth of vegetation, human error, drastic climatic conditions, and accidental fires in homes and industries.

## 2 Classification of Faults

Power systems around the globe are subjected to a wide range of faults classified and the faults associated with transmission and distribution lines are shunt faults which can be further classified as balanced and unbalanced. The transmission and distribution lines'

major concern is the shunt faults, also known as short circuit faults, which add to major reliability issues. The most frequent faults in transmission and distribution lines are SLG (Single line to ground) which constitutes 80% of total faults [1].

## 2.1 Shunt Faults (Short Circuit Faults)

Short circuits are dangerous and can cause arcing, fires, and equipment explosions. They can also lead to abnormal currents, overheating, and a shortened equipment lifespan. Short circuits can also disrupt operating voltages, affecting the quality of service for customers. If a short circuit persists and cannot be located, it can lead to major power interruptions and equipment failure. Shunt faults can be classified as follows [2].

- (a) Single line to ground fault (SLGF): For example, one of the phases of a conductor contacts the ground or neutral wire on a distribution line.
- (b) Line-to-line fault (LLF): For example, strong wind causes a short circuit between two phases of the conductor.
- (c) Double line to ground fault(LLGF): For example, the fault is associated with the falling tree which connects two-phase to the ground.
- (d) Triple Line fault (LLLFF): For example, the fault is associated with a falling tree connecting three phases of the conductor.

## 3 Machine Learning Techniques

AI is used through machine learning, which is the process of using data-driven mathematical models to help a computer learn without explicit instructions. As a consequence, a computer system continues to learn new abilities and improve on its own, according to [3] and an overview of machine learning techniques is displayed in Table 1.

**Table 1.** Supervised and unsupervised ML

Supervised classification	Regression	Unsupervised
Support vector machine (SVM)	Linear regression	K-means, K-median
Linear discriminant (LDA)	Ensemble methods	Fuzzy, C-means
Naïve Bayes	Decision trees	Hierarchical
K-nearest neighbor (KNN)	Least squares	Gaussian mixture
Neural Networks	Neural networks	Neural networks

### 3.1 Supervised Machine Learning (SML)

Supervised learning requires correctly labeled input and output sample data as an example to train a network or model as per reference [3]. Supervised learning is of two types: classification and regression. The classification classifies the input into a predetermined

output, such as genuine or spam mail. The regression method is the most common machine language used across various fields and it predicts continuous responses, for example, the relationship between the effect of sales after advertisement or reckless driving and road accidents. This research focuses on classification (discriminant, neural network) and regression (least squared, decision tree) features of supervised learning.

### 3.1.1 Linear Discriminant Analysis (LDA)

A statistical technique called discriminant analysis is used to categorize items into pre-determined categories (classes) based on several predictor factors. It is applied to categorize fresh observations based on their combination of predictor values and to identify the set of variables that best distinguishes across classes. Building a discriminant function  $D(x)$  that properly distinguishes the various classes and can be used to foretell the class membership of fresh data is the aim of discriminant analysis.

$$D(x) = w'x + w_0 \quad (1)$$

where

“w” denotes the vector of coefficients,

“x” is the vector of predictor values for a particular observation, and

“w<sub>0</sub>” denotes the intercept term.

By calculating the value of the function for each observation and allocating it to the class with the biggest value, the discriminant function distinguishes the classes. Using techniques like maximum likelihood estimation, the coefficients and the intercept term are inferred from the training data. Depending on the type of discriminant analysis performed, the specific shape of the discriminant function may change (e.g., linear, quadratic, Mahala Nobis).

### 3.1.2 Neural Network (NN)

Inspired by the working of the human brain, a neural network consists of several nodes known as artificial neurons organized in layers, and the neural network can be represented as

$$y = f(z) \quad (2)$$

where

“y” represents the neuron’s output,

“z” represents the inputs’ weighted sum, and

“f” represents the activation function.

With time there is a gradual shift of focus to AI-based techniques with an improved platform to perform data analysis research. AI-based methods are on the rise and all the above methods and techniques are used for data analysis across various fields.

In fault analysis, artificial neural networks are in use owing to their accuracy and their ability to understand the system behavior through existing data. It analyses the inputs and assigns them to predetermined outputs as indicated [1].

### 3.2 Unsupervised Machine Learning (UML)

Whereas unsupervised learning doesn't need the example. The system uses the data and groups them with shared characteristics known as clustering and it's widely used in gene sequencing, market research, and object orientation recognition like digital image processing [3].

## 4 Related Work

Some of the existing methods in power systems are discussed here. The fault classification Distribution Management System-based fault location as per [4] demonstrates how the DMS and network information can be integrated to form a distribution automation system using the existing microprocessor-based relay without much cost implication. The reference [5] uses an improved cuckoo search algorithm to find the fault, use current data from the field terminal unit(FTU), and perform a generic switching function. In this case, the algorithm's accuracy depends on the accuracy of data obtained from the FTU. The reference [6] uses the combined approach of Wavelet and ANFIS (adaptive neuro-fuzzy interference system) to detect and classify the high impedance fault and it was effective. The classification of fault causes is explored by reference [7] using the SML for the distribution system in PEA. The algorithm uses waveforms generated by the disturbance recorder as the input data and fits with the labeled data. The method was proven efficient as causes such as vegetation, animals, and devices were effectively classified. The method was further compared with the neural networks and indicated improved accuracy. The reference [8] uses the GK clustering feature of machine learning to identify the fault in the power system and compares it with fuzzy clustering.

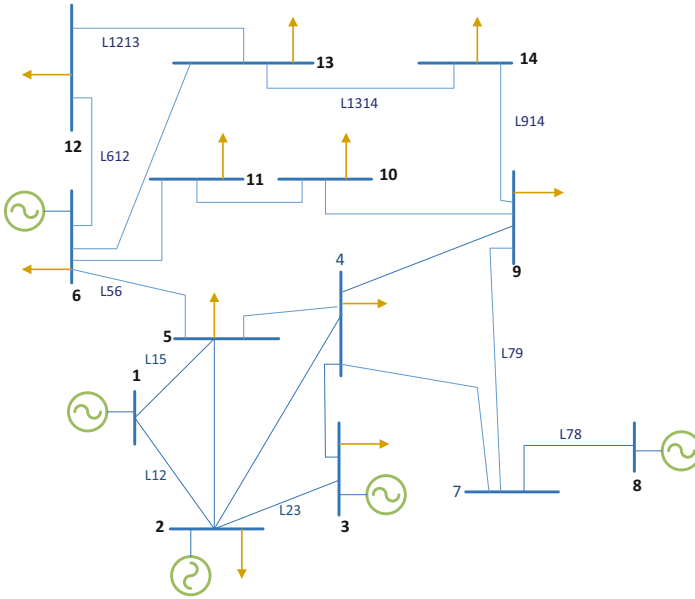
The SML and neural networks have outperformed conventional methods due to their ability to recognize the patterns in given data sets and their improved accuracy when deployed to carry out the task. Thus, this paper further emphasizes the use of supervised learning in fault classification and location.

## 5 Methodology

The methodology consists of a test system, database generation, algorithm, and output display/result.

### 5.1 Test System and Data Base Generation

The IEEE 14 bus system consists of 5 generators, and 11 loads from reference [9] (see Fig. 1). Various faults were applied at lines L12, L15, and L56 to create the database. The fault resistance ranging from 0.01 to 150 ohms was applied at line and RMS values of three-phase voltage and current (Ia, Ib, Ic, Va, Vb, Vc) and zero sequence current, and voltage (I0, V0) collected from bus 1.



**Fig. 1.** IEEE 14 bus system

## 5.2 Algorithm

The algorithm process flow includes the following steps:

1. Start: Create/Access data: Data history (excel files)/create
2. Preprocess data: Data sorting
3. Input: RMS values of current, voltage, zero sequence Voltage, current
4. Data analysis: Data preprocessing.
5. Fault detector: classifies normal or faulty.
6. Fault classifier: Classifies the fault type based on assigned bits.eg 1001-AG.
7. Fault locator: Locates the fault correlating inputs and outputs.

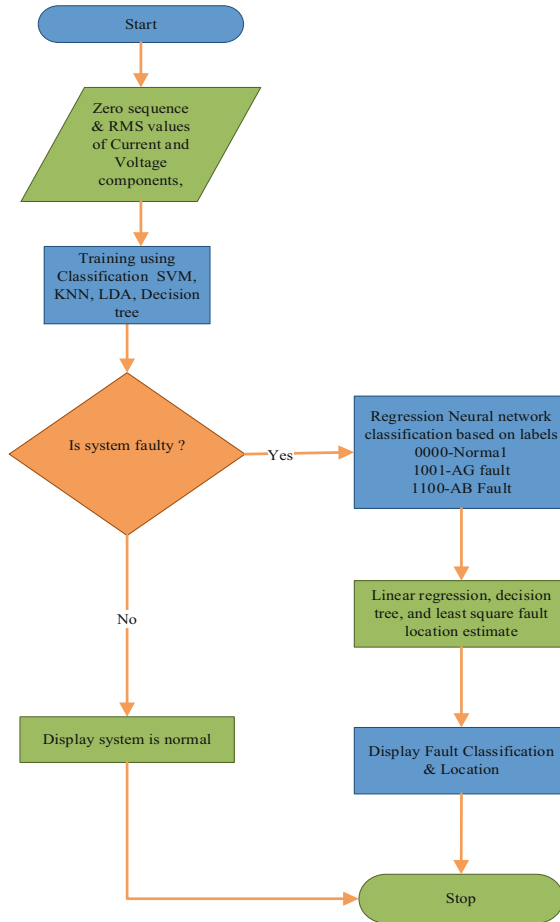
## 5.3 The Flow Chart

The process flow of the algorithm is shown in Fig. 2. The algorithm uses different SML techniques for detection, classification, and location.

## 5.4 Results and Discussions

The database generated is preprocessed and categorized into training, validation, and testing sets, and 70% of the total data is used for training, 15% for validation, and 15% for testing (see Table 2.)





**Fig. 2.** Process flow chart

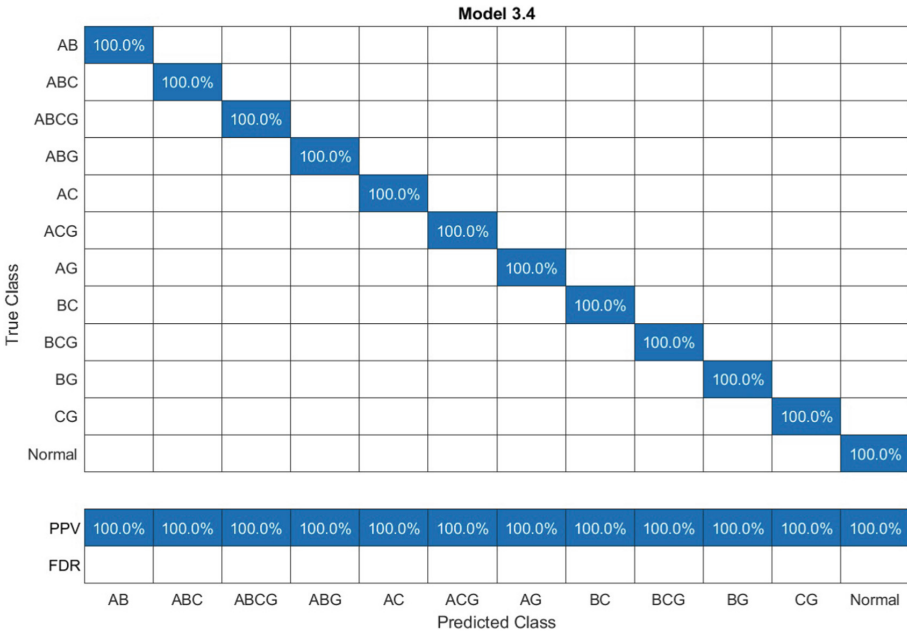
**Table 2.** Data distribution.

The training set 70% of 5200	Validation set. 15% of 5200	Testing set 15% of 5200
3640	780	780

#### 5.4.1 Training and Testing via Linear Discriminant Analysis

Using the 5000 data sets the training and testing was executed. The model was able to identify the faulty case, and normal case efficiently (see Fig. 3).

The data was further trained using other classification techniques as per Table 3 and the accuracy is compared. The KNN showed an accuracy of 91.7% followed by SVM, decision tree, and linear discriminate analysis with an accuracy of 100%.



**Fig. 3.** Result in positive predictive values (PPV) and false discovery rate (FDR)

**Table 3.** Comparison with other SML methods

S. No.	Machine learning models	Accuracy %
1	KNN	91.70
2	SVM	100
3	Decision tree	100
4	Linear discriminant	100

**5.4.2 Training and Testing via Regression Neural Network**

The regression neural network models are used to classify based on the assigned label and 100% accuracy is achieved, and (see Fig. 4) the regression R = 1 indicates 100% accuracy as the model could identify all 12 different fault types accurately in all the different fault conditions.

The classification pattern recognition tool was also used to classify the labeled data set, but the regression neural network outperformed the classification NN.

**5.4.3 Location Estimation.**

The fault data collected at different fault locations are used and the algorithm predicts the output. The testing data consists of faults applied at the step of 4 km at fault resistance ranging from 0.01 to 150 ohms and the total line length is 44.47km. The plot is generated

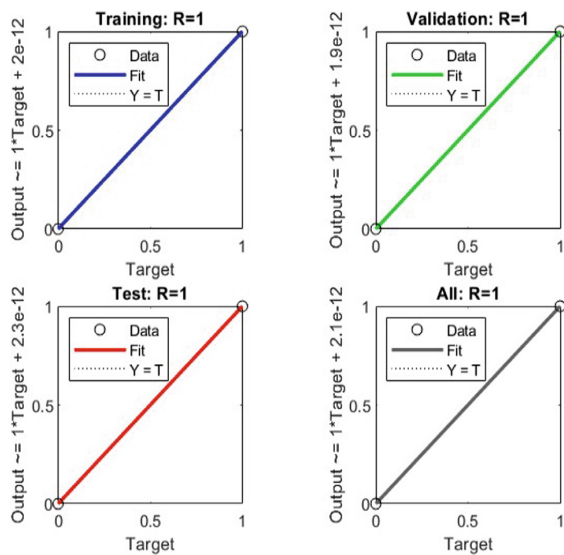


Fig. 4. Regression plot using neural network.

by an algorithm (see Fig. 5) that compares the overall accuracy using the three methods. The accuracy of least squares is 79.85%, followed by linear regression at 85.40% and the decision tree at 97.26%, and the fault estimation error is computed for the regression decision tree as shown in Table 4.

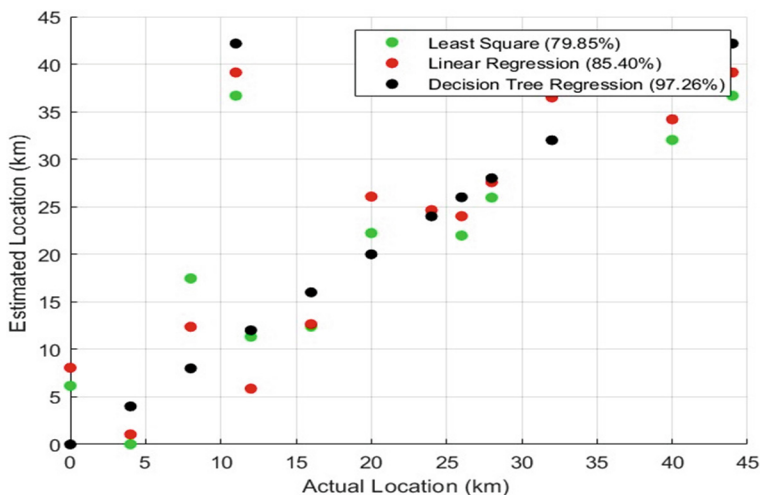


Fig. 5. Overall accuracy for all fault locations.

**Table 4.** Fault estimation error using decision tree for SLG

Actual fault km	Error = (Actual-predicted)/line length			
	R = 0.01 Ω	R = 10 Ω	R = 50 Ω	R = 150 Ω
4	0.00	0.00	0.00	0.00
8	0.00	0.00	0.00	0.00
12	0.00	0.00	0.00	0.00
16	0.00	0.00	0.00	0.00
20	0.00	0.00	0.00	0.00
24	0.00	0.00	0.00	0.00
28	0.00	0.00	0.00	0.00
32	0.00	0.00	0.00	0.00
36	0.22	0.22	0.22	0.22
40	0.00	0.00	0.00	0.00

## 6 Conclusion

The supervised machine learning concept is implemented to detect, classify, and locate the fault in the power system regardless of various fault types. The ability of the AI model is explored and found that if the data consist of a certain distinctive pattern SML can understand the pattern and further fitting features enabling to design predictive model without the need for system parameters like line impedance. The proposed method has reduced the vigorous calculation process of the conventional method saving time and energy. Future work involves applying this technique to real field data of the utility and checking the efficiency of the algorithm.

**Acknowledgment.** The author would like to thank the ASEAN/NON-ASEAN graduate scholarship program of Chulalongkorn University, Bangkok, Thailand.

## References

1. Stefanidou-Voziki, P., Raison, N.S.B., Dominguez-Garcia, J.L.: A review of fault location and classification methods in distribution grids. *209* (2022). <https://doi.org/10.1016/j.epr.2022.108031>
2. Abbas, A., Al-Tak, M.: A review of methodologies for fault location techniques in distribution power system. *Iraqi J. Electr. Electron. Eng.* **17**(2), 27–37 (2021)
3. Murphy, K.P.: *Probabilistic Machine Learning: An Introduction*. MIT Press (2022)
4. Jarventausta, P., Verho, P., Partanen, J.: Using fuzzy sets to model the uncertainty in TBE fault location process of distribution works. *EEE Trans. Power Del.* **9**(2) (1994)
5. Huang, X., Xie, Z., Huang, X.: Fault location of distribution network base on improved Cuckoo search algorithm. *IEEE Access* **8**, 2272–2283 (2020)

6. Veerasamy, V., et al.: High Impedance Fault Detection in MV Distribution Network using Discrete Wavelet Transform and Adaptive Neuro-Fuzzy Inference System. <https://doi.org/10.20944/preprints201810.0687.v1>
7. Promrat, W., Pupatanan, W., Benjapolakul, W.: Fault cause classification on PEA 33 kV distribution system using supervised machine learning compared to artificial neural network. In: 2021 9th International Electrical Engineering Congress (iEECON), 2021, pp. 5–8
8. Amalina Abdullah, C.B., Hoonchareon, N., Hidaka, K.: A study on the Gustafson-Kessel clustering algorithm in power system fault identification. *J. ElectrEngTechnol.* **12**(5), 1798–1804 (2017). <https://doi.org/10.5370/JEET.2017.12.5.1798>. ISSN (Print): 1975-0102. ISSN (Online): 2093-7423
9. Kersting, W.H.: Radial distribution test feeders. *IEEE Trans. Power Syst.* **6**(3), 975–985 (1991)



# Deep Learning-Based Time Series Forecasting for CO<sub>2</sub> Emission

Abhishek Anilkumar<sup>1</sup> (✉), V. Yadukrishnan<sup>1</sup>, M. Nimal Madhu<sup>2</sup>, V. Hareesh<sup>1</sup>,  
and B. Premjith<sup>1</sup>

<sup>1</sup> Center for Computational Engineering and Networking, Amrita School of Artificial Intelligence, Amrita Vishwa Vidyapeetham, Coimbatore 641112, India

abhishek18.anilkumar@gmail.com, b\_premjith@cb.amrita.edu

<sup>2</sup> Department of Electrical Engineering, National Institute of Technology Calicut, Kozhikode, Kerala, India

**Abstract.** The research paper provides a comprehensive and effective comparison of various deep learning (DL) models applied to univariate data of carbon dioxide (CO<sub>2</sub>) concentrations spanning from 1870 to 2021, focusing on five European countries: the United Kingdom (UK), Germany, France, Ukraine, and Italy. The primary aim of the study is to contribute to a better understanding of climate change and to facilitate the development of policies and actions to mitigate CO<sub>2</sub> emissions. Three DL models, namely long short-term memory (LSTM), convolutional neural network-long short-term memory (CNN-LSTM), and dynamic mode decomposition (DMD), are employed and thoroughly evaluated in this research. A novel aspect of this paper lies in the spatiotemporal modeling and forecasting capabilities, which enable simultaneous analysis for multiple countries using DMD. The results show that DMD outperforms other DL models, such as LSTM and CNN-LSTM, in accurately capturing data trends and forecasting future CO<sub>2</sub> values with the highest precision. The findings presented in this research paper have significant implications for advancing our understanding of climate change dynamics and guiding more effective measures to tackle CO<sub>2</sub> emissions.

**Keywords:** CO<sub>2</sub> emissions · Time series forecasting · Long short-term memory · Dynamic model decomposition · Convolutional neural network

## 1 Introduction

Burning of fossil fuels and energy production is the major contribution towards the increasing levels of CO<sub>2</sub> emission which is the major contributor to the world environmental crisis. To mitigate the effects of these emissions, the European Union has introduced various policies and initiatives aimed at reducing emissions. For assessing the success of these initiatives, we must have a reliable and accurate model that predicts the annual CO<sub>2</sub> emissions. In this research paper, we aim to develop a comprehensive model to predict CO<sub>2</sub> emissions for European countries.

The data used in this study will be sourced from “our world in data”, which contains the annual CO<sub>2</sub> emission data or information for the whole world. The models will be

trained using historical data and model has been tested and performance analysis for comparative study is conducted. The outcome of this study provides valuable insights into the effectiveness of the proposed models in predicting CO<sub>2</sub> emissions for European countries and informs future policy decisions aimed at reducing emissions.

In conclusion, this research paper will contribute to the ongoing efforts to mitigate CO<sub>2</sub> emissions by developing an accurate and reliable model that predicts emissions for European countries. The results of this study will provide crucial information for policymakers and stakeholders, and pave the way for the development of more advanced models to tackle the environmental crisis.

## 2 Literature Review

A multitude of research has been carried out utilizing machine learning models, deep learning models, and statistical models on time series data. Some notable studies include Kumari and Singh [1] researched the prediction of CO<sub>2</sub> emissions in India using statistical, machine learning, and deep learning models. The study found that the Long Short-Term Memory (LSTM) model outperformed other models, achieving an R<sup>2</sup> score of 0.990 for univariate time series data. This paper fails to address how ML/DL frameworks can understand the non-linearity present in the underlying data. Li et al. [2] compared the performance of deep learning algorithms such as LSTM, CNN, CNN-LSTM, and Sparrow search Algorithm with LSTM in predicting carbon emission intensity and per capita emissions from 2023 to 2033. They found that CNN-LSTM outperformed other models. The performance comparison was based on a small sample of data from 2009 to 2019. In 2017, Kuttichira et al. [3] proposed a method to predict stock prices using DMD. The authors compared three methods for stock value prediction. DMD outperformed forecasting models like ARIMA in both univariate and multivariate time series forecasting, from this paper we are referring to the forecasting ability of the DMD model. Olszewska et al. [4] have developed a system incorporating a LSTM with IOT for CO<sub>2</sub> forecasting in indoor air quality monitoring. The developed system will be able to predict future CO<sub>2</sub> concentrations. The authors claim that the system performs with good accuracy with a marginal error of 5.5%. In recent years, various DL forecasting models are developed by using standalone DL [5, 6] and hybrid DL [7–9] models.

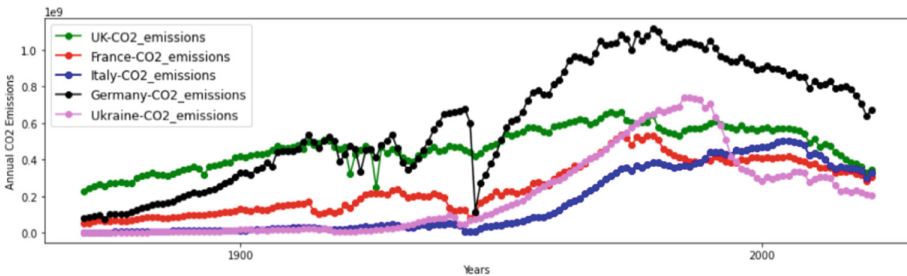
Bakay and Ağbulut [5] used Support Vector Machine (SVM), Artificial Neural Network (ANN), and deep learning models to forecast CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O, F-gases, and total GHG emissions in Turkey from 1990–2014. Results showed that the deep learning models performed better than SVM and ANN models in terms of Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) performance metrics. Pino-Mejías et al. [6] developed models to predict energy consumption and CO<sub>2</sub> emissions of office buildings in Chile. The authors have proved that a mathematical model that depends on multi-layer perceptron (MLP) performs better compared to models that depend on linear regression.

Kim et al. [7] developed a novel approach for predicting electrical energy consumption. The authors created a CNN-LSTM network that will be able to extract spatiotemporal attributes to effectively predict electrical energy prediction. The CNN-LSTM model can predict consumption effectively with a Mean Square Error (MSE) score of 0.37 for

the time granularity in terms of minute, hour, daily, and weekly, the forecasting capability and pattern identification ability of the CNN-LSTM model of this project is used in this research project. The Livieris et al. [8] in a research paper propose two types of CNN-LSTM architecture for predicting the gold price and movement. From the research, the authors concluded that the CNN-LSTM model having a convolutional layer with a lesser filter count performs better for regression problems and a convolutional layer having a higher filter count compared to the previous one performs better in classification problems. Imamverdiyev et al. [9] developed a CNN-LSTM model for predicting oil production around the world. The proposed model has been compared with the performance of standalone CNN and LSTM models and found that CNN-LSTM performed better compared to the other two by using the Root Mean Squared Logarithmic Error (RMSLE) performance metric along with the reduction in loss function for DL models.

The proposed research project aims to compare the performance of three deep learning models (LSTM, CNN-LSTM, and DMD) in predicting CO<sub>2</sub> emissions using historical data for 5 European countries. The LSTM model is particularly suited for analyzing sequential data, while the CNN-LSTM hybrid model combines features of both CNN and LSTM models to provide strong performance in feature extraction and prediction. The DMD model captures patterns and trends in the data to implement space-time relations from underlying data. This project also explores the spatiotemporal relation that DMD can establish between CO<sub>2</sub> emissions and countries. The project utilizes historical data on CO<sub>2</sub> emissions for 151 years for 5 European countries and aims to generate accurate predictions for future emissions.

### 3 The Dataset



**Fig. 1.** Annual CO<sub>2</sub> emission plot for the UK, France, Italy, Germany and Ukraine

For this project, the CO<sub>2</sub> emission data from Our World in Data is used. The data used for this research is of type univariate from the year 1870 to 2021. The data is measured in terms of billion tonnes. For performing the prediction and comparative study of the performance of deep learning models we are considering the annual CO<sub>2</sub> emission of 5 countries from the Europe continent, namely the United Kingdom (UK), Italy, France, Germany and Ukraine. Figure 1 show the annual CO<sub>2</sub> emission for 5 European countries.



From Fig. 1, we can understand that the data is following a non-linear trend. The non-linearity present in the dataset is being used to design a DL model which can predict future CO<sub>2</sub> emission values accurately.

## 4 Methodology

As we can see from Fig. 1 there is a non-linearity existing between the year field and annual emission. To effectively predict the future value for time series forecasting, we are taking data from 1870 to 2021. The data used for this project is of type univariate, that is it has only 2 columns one for the year and the other for Annual CO<sub>2</sub> emissions. As part of the preprocessing step, we are selecting the countries which, are having values from the year 1870 to 2021. Out of all the data for 22 European countries, the annual CO<sub>2</sub> emission for 5 countries is used for this research. Once the data is loaded into the data frame, data normalization is applied if LSTM and CNN-LSTM models are used, data normalization is not performed if data is fed into DMD since DMD models can identify the hidden trend within the dataset without normalization. After this stage then data is split into testing and training sets. The training dataset is used for training the data, and the testing set is used for testing the model. The architecture flow diagram is shown in Fig. 2.

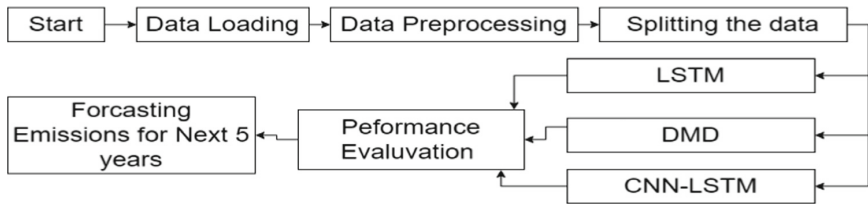


Fig. 2. Architecture

### 4.1 Data Normalization

Data Normalization is the process of converting the values within a data frame or dataset into a common scale. The normalization process is done to improve the quality of data and help reduce the computational overhead on the algorithm to understand and find the trend lying within the data. For this project, we are using MinMaxScaler normalization. The scale of normalization used in this project is between [0,1]. To obtain the predicted values in form of actual numerical values present in the data, an inverse transform operation is performed. The formula for MinMaxScaler is given below.

$$X_{scaled} = \frac{X - \min(X)}{\max(X) - \min(X)}. \quad (1)$$

$$X_{inverse\_transform} = X_{scaled} * (\max(X) - \min(X)) + \min(X). \quad (2)$$

## 4.2 Sliding Window

The Sliding Window [10] process is a technique used for processing sequential data such as time series. The basic working principle of sliding windows is to break the data points into smaller chunks and use these windowed chunks for predictions. The window size used is 5. One of the advantages of the sliding window model is that it can handle nonlinear data, where its statical property can change over time. The sliding window adjusts to these changes by fitting a new model for each new window of data.

## 4.3 Deep Learning Model-1 - Long Term-Short Memory

LSTM [10] is a class of Recurrent Neural Networks (RNN) that is adept to solve problems having vanishing gradients. The LSTM contains a series of components namely memory cells, input gates, forget gates, and output gates. All these components will work in synergy to regulate and flow of information through the neural network allowing it to model long-term dependencies. A memory cell in an LSTM network is a small neural network structure used to store and recall information over an extended period. The function of the input gate is to control the inflow of information into the memory cell. The function of the forget gate is to control information flow from the memory cell. The function of the output gates is to regulate the flow of information from the memory cell to the rest of the network. For this project, a LSTM architecture has one input layer, 2 LSTM layers with size 64, and 2 Dense layers with sizes 8 and 1, The layer and neuron count are attained based on trial and error. The model parameters which give the best R2 score are considered for this project.

## 4.4 Deep Learning Model-2 - Convolutional Neural Network – Long Short-Term Memory

The CNN-LSTM [10, 11] is a hybrid framework, that is used for forecasting applications. In this architecture, the CNN part of the model consists of one or multiple layers of convolutional that can extract local attributes from the loaded data. The CNN layer is fed into one or more layers of LSTM which has the same architecture as above/ Finally, the results from the LSTM are fed into a dense layer which is used for prediction. By combining the strengths of CNN and LSTM, this hybrid architecture can effectively capture both local features and temporal dependencies in the input data. In this project, a Conv\_1D layer with size 128 and 1 LSTM layer with size 64, and 2 dense layers with sizes 8 and 1. The model parameters which give the best R2 score are considered for this project.

## 4.5 Deep Learning Model-3 - Dynamic Mode Decomposition (DMD) Using PyDMD

Dynamic Mode Decomposition (DMD) [3, 12] is a technique developed by Schmid for understanding the dynamic or pattern of the data with the flow of time. In simple words, DMD can be explained as a process directed by data, free from all complex equations which can decompose a system that is intricate into its corresponding spatiotemporal

components, which can be used to predict future values. The data we used is yearly data starting from the year 1870 to 2021 for 5 European countries. With 5 being the number of rows for the DMD matrix and 151 number of columns which is the CO<sub>2</sub> emission for the mentioned years. The DMD dataflow is as follows,

Collect the snippet of data for a large duration of time. Here we will have two matrices one with the size of 1 to T-1 and the other with having a matrix of size 2 to T. The matrix shape is as follows:

$$Y_1 = [y_1, y_2, \dots, y_{T-1}] \tag{3}$$

$$Y_2 = [y_2, y_3, \dots, y_T] \tag{4}$$

Take the matrix  $Y_1$  and perform Singular Value Decomposition.

$$Y_1 = U * \sum * V \tag{5}$$

The SVD will yield 3 matrices where  $U$  and  $V$  are left and right singular vectors and  $\sum$  is a matrix with singular values as diagonal elements.

Compute the Koopman matrix by performing truncated SVD on a certain predefined matrix.

$$A' = [\text{truncated } U]^T * Y_2 * [\text{truncated } V] * [\text{truncated } \sum]^{-1}. \tag{6}$$

Perform the Eigen Value Decomposition on the above obtained Koopman matrix, where  $\Phi$  is eigen values and  $Q$  is the eigenvectors of the Koopman matrix.

$$A' = Q * \Phi * Q^T \tag{7}$$

After obtaining the Koopman matrix, and DMD spectrum we will be able to perform time series forecasting using the below two equations.

$$\Psi = Y_2 * [\text{truncated } V] * [\text{truncated } \sum]^{-1} \tag{8}$$

$$A = \Psi * \Phi * \text{pseudoinverse}(\Psi) \tag{9}$$

one of the most notable features of DMD is its ability to develop the state of the system using its unique data-driven spectral decomposition, hence the above equation can be further simplified into the below equation and represented in terms of Fourier modes.

$$Y_{DMD}(t) = \sum_{j=1}^n Q_j * \exp(\omega_j * t) * b_j \tag{10}$$

where  $Y_{DMD}(t)$  is the state of the system at time  $t$ ,  $b_j$  is the first amplitude of DMD modes,  $Q_j$  is a matrix where columns are the column of eigenvectors and  $\omega_j$  is a diagonal matrix with eigenvalues

After calculating the A matrix, we will be able to implement the DMD for time series forecasting. For implementing DMD, the PyDMD package available in python is used. PyDMD provides an easy-to-use implementation of DMD for dynamic data analysis. The library allows users to perform DMD on data sets and obtain a low-dimensional representation of the dynamics in the data. The PyDMD module takes a single data point as input to predict a single data point output. Since PyDMD is an inbuilt function the datapoint count is fixed. Here we are training the DMD model with 140 years of data and testing the DMD model with 11 years of data with the Singular Value Decomposition (SVD) rank of 5. Since the predicted value contains complex numbers, we are calculating the absolute value of the predicted values. Since the DMD model works with single input-single output architecture we are using the previous year's value to predict the next year's value. The PyDMD library also provides tools for visualizing and post-processing the results of the DMD analysis.

#### 4.6 Analysis of Models Using Performance Evaluation Metrics

In this portion, we will be explaining the process of analyzing the efficiency of the model by using various evaluation metrics. As part of the performance evaluation, we are using 3 metrics to calculate the effectiveness of the model. The metrics we are using for performance evaluation are R2 Score, Mean Squared Log Error (MLSE) and MAPE.

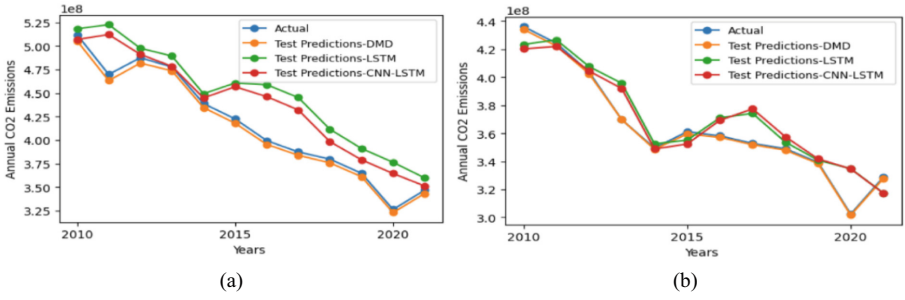
## 5 Results and Discussions

The error rate for various models is given in Table 1. For LSTM and CNN-LSTM we have individually calculated the metric values and took the average of the metric values.

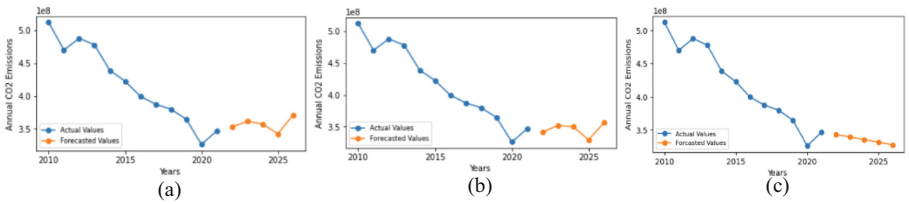
**Table 1.** Performance metric details table.

Metric	LSTM	CNN-LSTM	PyDMD
R2 score	0.825988	0.861511	0.996008
Mean square log error	0.00014641	0.00132	0.000355047430
Mean absolute percentage error	0.04692240	0.044082	0.015354309768

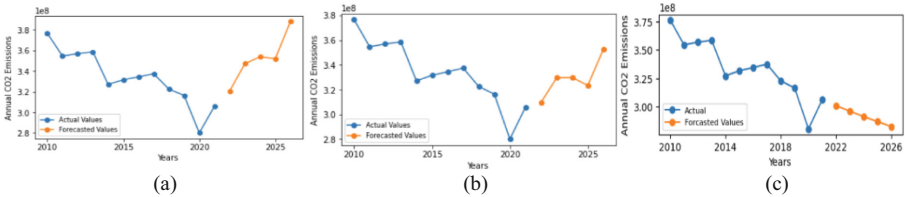
Table 1 shows the performance metric of the 3 DL models used in this project. Taking the R2 score, MSLE, and MAPE along by comparing the comparative performance plot of 3 DL models in the above plot, we can understand the DMD models outperforms LSTM and CNN-LSTM models. Moreover, from the above plot, we can understand that the DMD model can identify the non-linearly present in the original dataset and predict the future with more accuracy compared to other DL models. The above clearly states that the DMD model is the best-performing model as actual values and predicted values are close to each other, while CNN-LSTM and LSTM predict the values which not as close as DMD predicted values. Hence DMD is better at predicting CO<sub>2</sub> emissions compared to the other 2 models.



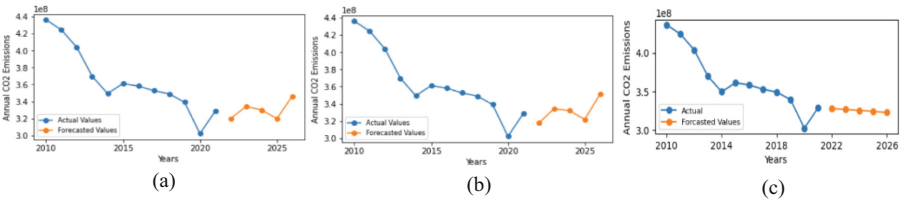
**Fig. 3.** Performance analysis plot for actual and predicted data for DMD, LSTM, and CNN-LSTM for the countries (a) UK (b) Italy



**Fig. 4.** Actual-forecasted plot for the UK using different DL algorithm (a) LSTM (b) CNN-LSTM (c) DMD

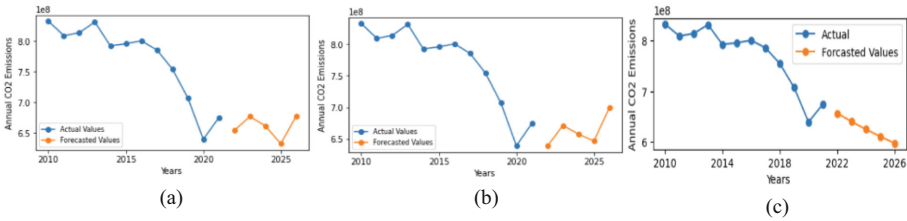


**Fig. 5.** Actual-forecasted plot for France using different DL algorithm (a) LSTM (b) CNN-LSTM (c) DMD

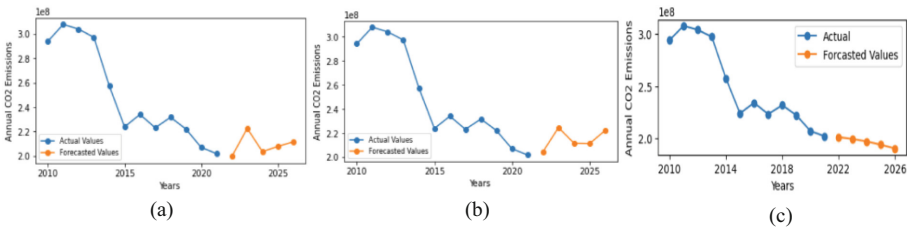


**Fig. 6.** Actual-forecasted plot for Italy using different DL algorithm (a) LSTM (b) CNN-LSTM (c) DMD

In Figs. 4, 5, 6, 7 and 8, we are performing a comparison of forecasted value between the 3 DL algorithms. Based on Fig. 3 we are understanding that DMD prediction is



**Fig. 7.** Actual-forecasted plot for Germany using different DL ALGORITHM (a) LSTM (b) CNN-LSTM (c) DMD



**Fig. 8.** Actual-forecasted plot for Ukraine using different DL algorithm (a) LSTM (b) CNN-LSTM (c) DMD

more accurate in prediction in comparison with LSTM and CNN-LSTM models. And Figs. 4, 5, 6, 7 and 8 show the decreasing trend in the DMD forecast while LSTM and CNN-LSTM show non-linearity in forecasted value. The DMD module can identify the non-linearity present in the data and accurately predict the future value. DMD is also able to spatiotemporal analysis on the CO<sub>2</sub> Emission data of the 5 European countries with an accuracy of 99.6% which is a novel approach. The DMD model takes less time to execute while the LSTM and CNN-LSTM take more time to run the epochs. Hence the computational complexity for DMD is less compared to the other models. For LSTM and CNN-LSTM we are using windowing techniques to make the framework understand the nonlinearity in the data, due to which there will be a slight error in the predicted output. This error will be carried out to the future predicted value as the cumulative error, due to which accuracy is reduced compared to the DMD framework.

## 6 Conclusion

The comparative analysis of various deep learning algorithms reveals that the inbuilt Dynamic Mode Decomposition (DMD) model outperforms other DL algorithms, such as CNN-LSTM and LSTM Models, in predicting future CO<sub>2</sub> emissions. Notably, the DMD model not only provides accurate future emission values but also establishes a crucial spatial-time relationship between countries and years based on annual CO<sub>2</sub> emissions. As part of future research, there is a promising opportunity to expand this work by developing a multivariate DL model that incorporates additional parameters, such as GDP, renewable energy usage, technological advancements, and future governmental policies. This extended approach would enable researchers to create novel models

with enhanced predictive capabilities, ultimately leading to more precise and effective predictions of CO<sub>2</sub> emissions.

## References

1. Kumari, S., Singh, S.: Machine learning-based time series models for effective CO<sub>2</sub> emission prediction in India. *Environ. Sci. Pollut. Res.* 1–16 (2022). <https://doi.org/10.1007/s11356-022-21723-8>
2. Li, F., Hu, Y., Wang, L.: Carbon emission prediction using CNN-LSTM. *Int. J. Multidiscip. Res. Publ. (IJMRAP)* 5(6), 30–35 (2022)
3. Kuttichira, D.P., Gopalakrishnan, E.A., Menon, V.K., Soman, K.P.: Stock price prediction using dynamic mode decomposition. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 2017, pp. 55–60. <https://doi.org/10.1109/ICACCI.2017.8125816>
4. Zhu, Y., Al-Ahmed, S., Shakir, M.Z., Olszewska, J.: LSTM-based IoT-enabled CO<sub>2</sub> steady-state forecasting for indoor air quality monitoring. *Electronics* 12 (2022). <https://doi.org/10.3390/electronics12010107>
5. Bakay, M.S., Ağbulut, Ü.: Electricity production-based forecasting of green house gas emissions in Turkey with deep learning, support vector machine and artificial neural network algorithms. *J. Clean. Prod.* 285, 125324 (2021). ISSN: 0959-6526. <https://doi.org/10.1016/j.jclepro.2020.125324>
6. Pino-Mejías, R., Pérez-Fargallo, A., Rubio-Bellido, C., Pulido-Arcas, J.A.: Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and CO<sub>2</sub> emissions. *Energy* 118, 24–36 (2017)
7. Kim, T.-Y., Cho, S.-B.: Predicting residential energy consumption using CNN-LSTM neural networks. *Energy* 182 (2019). <https://doi.org/10.1016/j.energy.2019.05.230>
8. Livieris, I., Pintelas, E., Pintelas, P.: A CNN-LSTM model for gold price time series forecasting. *Neural Comput. Appl.* 32 (2020). <https://doi.org/10.1007/s00521-020-04867-x>
9. Imamverdiyev, Y., Abdullayeva, F.: Development of oil production forecasting method based on deep learning. *Stat. Optim. Inf. Comput.* 7, 826–839 (2019). <https://doi.org/10.19139/soic-2310-5070-651>
10. Selvin, S., Vinayakumar, R., Gopalakrishnan, E.A., Menon, V.K., Soman, K.P.: Stock price prediction using LSTM, RNN and CNN-sliding window model. In: 2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Udupi, India, 2017, pp. 1643–1647. <https://doi.org/10.1109/ICACCI.2017.8126078>
11. Pandianchery, M., Vishvanathan, S., Gopalakrishnan, E.A., Soman, K.P.: Long Short-Term Memory-Based Recurrent Neural Network Model for COVID-19 Prediction in Different States of India (2022). <https://doi.org/10.1201/9781003324447-12>
12. Harichandana, M., Vishvanathan, S., Sajithvariya, V., Sivanpillai, R.: Comparison of image enhancement techniques for rapid processing of post flood images. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIV-M-2-2020, pp. 45–50 (2020). <https://doi.org/10.5194/isprs-archives-XLIV-M-2-2020-45-2020>



# Manila City House Prices: A Machine Learning Analysis of the Current Market Value for Improvements

Lejan Daniel I. Perdio<sup>(✉)</sup>, Marife A. Rosales, and Robert G. de Luna

Artificial Intelligence and Automation, Polytechnic University of the Philippines, Manila, Philippines

lejandanieliperdio@iskolarngbayan.pup.edu.ph, {marosales, rgdeluna}@pup.edu.ph

**Abstract.** The goal of the study is to develop an intelligent system that employs machine learning techniques to predict house prices in Manila City. The dataset contains attributes such as floor area, no. of storeys, classification, type of structure, additional floors or areas, and tag ID. Machine learning algorithms such as Linear Regression, Bayesian Ridge, Gradient Boost, and Lasso Regression were utilized to construct the predictive model. Feature selection methods and Genetic Algorithm optimization were also implemented. The model's performance was evaluated using different performance metrics, and the findings indicated that Gradient Boost exhibited the most optimal performance with an optimized R-squared value of 0.7508, and EVS value of 0.7640. Based on the evaluation results, the proposed model with genetic algorithm optimization can provide accurate predictions for the prices of houses in Manila City, and could be a useful tool for both buyers and sellers in the real estate industry.

**Keywords:** Real state appraisal · Current market value · Regression models · Manila City house prices

## 1 Introduction

Property investment is a major financial decision that requires a significant amount of initial capital [1]. Thus, any decision to invest in property should be based on a careful assessment, however, accurately determining the market value of these improvements can be a complex and time-intensive task, often involving factors such as residential, structural and environmental [2].

These various factors are commonly examined by traditional real estate appraisal methods, however, such manual procedures may be susceptible to the biases and interests of appraisers, posing a potential risk for a biased or subjective evaluation of a property. This may lead to losses for investors or households involved in the transaction [3]. The advancement of AI technology increases the likelihood of a transition in labor demand from manual and repetitive duties to digital tasks that necessitate skill sets like data



analysis and algorithm development [4]. Therefore, developing practical algorithms and automated models that can provide objective and impartial real estate appraisals has become increasingly important, not only for improving the accuracy of valuations but also for reducing the burden of labor-intensive work [5].

In this research, we aim to predict house prices in Manila City, Philippines using some of the machine learning techniques. Our analysis will focus on identifying key features that influence house prices in Manila City, and using this information to develop a model that can predict prices for properties in the city. This research will provide valuable insights for real estate investors, property developers, and academic researchers interested in the housing market in Manila City. Additionally, we will also explore the use case of this research on government's urban planning [6].

## 2 Review of Related Literature

Recent advancements in machine learning models and the emergence of big data have provided new opportunities for real estate industry [7]. Research studies utilizing statistical learning approaches to analyze the housing market have been conducted in various locations, such as the United States [1, 2, 5–9], Europe [3, 10], Australia [11, 12], and China [13]. Nonetheless, studies applying data analytics and machine learning models to examine the housing market in the Philippines are not widely available in the literature.

Several studies have been conducted to investigate the variables that contribute to housing prices. For example, Zhang conducted a multiple linear regression of the variables that influences the value of residential properties. The findings revealed that property size, location, no. of floors, no. of rooms, and availability of facilities were key determinants of the property's appraisal value [2].

Satish developed a model for predicting housing costs by utilizing various machine learning algorithms, including Gradient Boost, XGBoost, Lasso regression, and neural networks. The results of the analysis showed that the Lasso regression algorithm consistently outperformed the other models in terms of accuracy for predicting housing costs [9]. Fan et al. analyzed the different machine learning algorithms such as Lasso Regression, Ridge Regression, and XGBoost with the use of optimal parameters [6].

Another study by Garcia compared the performance between linear regression and ensemble models for mass-appraisal algorithm techniques. According to its findings, machine learning algorithms are more suitable for dealing with the nonlinearities present in complex real estate market data than traditional linear models. Bagging-based algorithms, such as random forest and extra-trees regressor, exhibit overfitting issues. In contrast, boosting-based algorithms demonstrate superior performance and exhibit lower levels of overfitting [10].

These studies provide valuable insights into the factors that influence housing prices in Manila City and highlight the capability of machine learning techniques in predicting the market value of house improvements. However, there is a need for further research to explore the specific factors that contribute to the market value of house improvements in Manila City and to develop accurate models for predicting these values. The objective of this study is to address the gap in existing literature by employing machine learning algorithms to the current market value of house improvements in Manila City. In particular,

we will contrast our proposed methodology with the machine learning models utilized in prior studies. Our approach includes feature reduction techniques to determine the essential features that impact house prices in Manila City, as well as genetic algorithm optimization to identify key parameters in a predictive model.

Overall, this review demonstrates the importance of understanding the factors that contribute to the market value of housing improvements and highlights the potential of machine learning techniques in predicting these values.

### 3 Methodology

The proposed methodology for predicting house prices in Manila City involves a multi-step process, shown below, done using Python, including the pandas, numpy, and scikit-learn libraries. Jupyter notebook was used as the development environment for the analysis (Fig. 1).

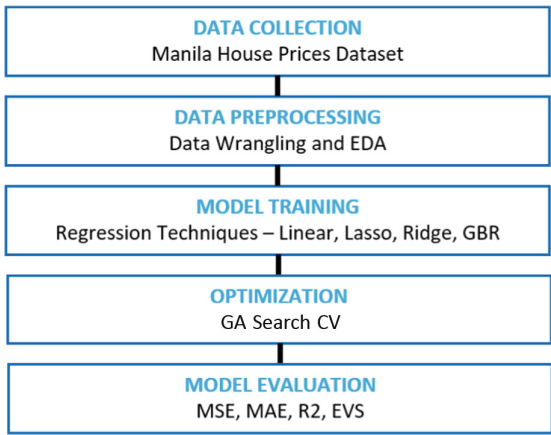


Fig. 1. Research methodology framework

#### 3.1 Data Collection

The dataset used in this study is from the Department of Public Works and Highways Public and Private Partnership Service (DPWH-PPPS) and it pertains to the current market value of the improvements affected by the NLEX-SLEX Connector Road Project right-of-way.

The location of the properties in the dataset is situated in Manila City and it includes various features such as floor area in sqm, number of storeys, classification, type of structure, additional floors or areas, and tag ID. The initial dataset prior to data preprocessing consists of 1,331 improvements (Table 1).

**Table 1.** List of features

Feature name	Data type	Description
Tag ID	object	A unique identification number assigned to each property in the dataset for tracking and referencing purposes
Floor area	float64	Refers to the total floor area of the house, usually measured in square meters
No. of storeys	int64	Refers to the number of levels in the house, including the ground floor
Additional area	object	Refers to any additional floors or areas in the house, such as a mezzanine, attic, roof deck or roof slab
Classification	object	Refers to the classification of the house, such as residential or commercial
Type of structure	object	Refers to the type of construction material used in building the house, such as concrete, semi-concrete, or wood

### 3.2 Data Preprocessing

Preprocessed the data by cleaning, deleting missing values, and transforming variables as needed. This step ensures that the dataset is ready for analysis and model building. It should be noted that there are numerous instances of missing values in the target variable - the price of the properties. Although it is possible to replace these missing values with imputed values, this approach could introduce additional bias in the input data [12]. Therefore, any observations with missing values in the 'Estimated Cost' column are eliminated from the dataset to enhance the reliability of the analysis. The final dataset used for the model was narrowed down into 1,067 improvements.

Furthermore, the Tag ID feature, which serves as a unique identifier for each house in the dataset, is not included in the regression analysis. The Tag ID feature is used exclusively for data management purposes and is therefore not pertinent to the housing market analysis. As a result, it is removed from the feature selection process and the subsequent regression analysis.

Categorical data are typically represented as text values in a dataset. For Manila City House Prices dataset, features such as additional area, classification and type of structures are needed to be converted in order to use such data in machine learning models. To transform categorical data into numerical values, two pre-processing techniques were utilized in this study: Label Encoder and One-Hot Encoder. ColumnTransformer allowed us to use pre-processing technique to different columns of a dataset while avoiding the dummy variable trap.

One-Hot Encoding produced seven columns for the additional area feature, two columns for classification, and three columns for type of structure. In order to avoid the dummy variable trap, one column for each feature has been removed. The final dataset has a shape of (1067, 11).

Feature reduction is another important step in data preprocessing that involves the selection of a subset of the most significant features from a larger feature set in a dataset.

The aim is to simplify the dataset and remove irrelevant or redundant features that may not contribute to the predictive power of a machine learning model.

The feature reduction technique used was Backward elimination. It involves iteratively removing features from a model until all the remaining features satisfy a predetermined significance level, set at 0.05. The elimination process starts with a full model that includes all features, and the significance of each feature is assessed using a statistical test such as the p-value. Features that do not meet the significance level are removed, and the model is re-fitted with the remaining significant features. This iterative process continues until all the features that remain in the model are statistically significant, and no further features can be removed without compromising the model's performance. By using this approach, eleven features generated from one hot encoder has been reduced into six statistically significant features; linear constant (column 0), with Mezzanine and Roof Slab from Additional Floor feature (column 4), Residential from classification feature (column 7), Semi-concrete from type of structure feature (column 8), floor area (column 10) and number of storeys (column 11).

### 3.3 Model Training

The performance of four distinct machine learning techniques for predicting the target variable has been evaluated: Linear Regression [2], Lasso Regression [9], Gradient Boosting Regression [10], and Bayesian Ridge Regression [6].

To ensure a fair comparison of the algorithms, we used both Hold-out validation and K-fold cross-validation with  $K = 10$  [14]. In each fold, the data was randomly partitioned; training data consisted of 80% of the dataset, while the remaining 20% was assigned to the testing data [15]. The process was repeated ten times, with each fold serving as the test set once.

### 3.4 Optimization

Genetic algorithm is a type of optimization technique inspired by the natural selection process [16]. The algorithm starts with a population generated by randomly selecting genes from a uniform distribution. An individual is the term used for each solution, and a generation is the term used for the entire set of individuals. The individuals are evaluated based on a fitness function, which measures how well they perform on the given task. In the context of hyperparameter tuning, the fitness function is typically a measure of model performance such as accuracy, precision, recall, or  $R^2$  score. The genetic algorithm then evolves the population through several generations, using three main operators: selection, crossover, and mutation. Selection involves selecting the fittest individuals from the population for the purpose of creating a new generation. Crossover is the process of joining the genetic material of two individuals at random to form a new individual. Mutation is the process of modifying an individual's genetic material at random in order to create a new one [17].

The GASearchCV class implements a genetic algorithm search for hyperparameter tuning in a scikit-learn pipeline. The estimator parameter is the estimator to tune, and the param\_grid parameter is a dictionary of hyperparameter ranges to search over. The other

parameters control the genetic algorithm search, including the number of generations, the population size, and the fitness function or scoring [18].

GAsearchCV works by creating a random initial population of hyperparameters and then evolving that population over several generations using the three genetic operators. At the end of the search, the best set of hyperparameters is returned. The hyperparameters for each algorithm and the corresponding ranges for the search were as follows:

- Multiple Linear Regression: 'normalize', True or False
- Bayesian Ridge Regression: alpha\_1 and alpha\_2, ranges from 1e-6 to 1e-2, lambda\_1 and lambda\_2, ranges from 1e-6 to 1e-2
- Gradient Boosting Regression: n\_estimators, learning\_rate, max\_depth, and subsample, with values from 50 to 200, 0.01 to 0.1, 3 to 7, and 0.5 to 1, respectively.
- Lasso Regression: alpha, ranges from 1e-6 to 1e-2.

The performance of each algorithm was assessed by computing the R-squared values for each fold and then reporting the resulting average and standard deviation.

### 3.5 Model Evaluation

To assess the effectiveness of our regression models, we used the following metrics:

- The MAE or mean absolute error, measures the average magnitude of our predictions' errors and is calculated by averaging the absolute difference between the predicted and true values. Lower MAE indicates a higher performance [19].
- The MSE or mean squared error, calculates the average squared difference between predicted and true values by taking the mean of the squared differences. Lower MSE indicates improved performance.
- The R<sup>2</sup> or R-squared, is a measure of how well the model fits the data, and is computed as the ratio of the explained variance to the total variance. The scale is between 0 and 1, where higher values indicate excellent performance.
- The EVS or Explained variance score, explains the error dispersion in a particular dataset. The formula is 1 minus the ratio of the residual variance to the total variance. The scale is between 0 and 1, where higher values indicate excellent performance.

By using all of these metrics, we aimed to get a comprehensive understanding of how well our models were performing, and to identify which model was the best for Manila city house prices task.

## 4 Results and Discussion

Table 2 and 3 present the results of our model evaluation using the MSE, MAE, R-squared, and EVS metrics for each of the four regression models that we used (Linear Regression, Bayesian Ridge, Gradient Boost, and Lasso Regression) using default parameters. Tables 2 and 3 were evaluated through Cross-fold validation and Hold-out Validation, respectively [14].

For Cross-fold validation, we can see that the Bayesian Ridge model performed the best overall, achieving the highest R-squared and EVS scores, while the Linear regression

model and Gradient Boost presented a lowest MSE and MAE, respectively. However, all four models performed relatively well, with R<sup>2</sup> scores ranging from 0.6625 to 0.7378, and EVS scores ranging from 0.6689 to 0.7378.

**Table 2.** Cross-fold validation evaluation performance through default parameters

Model	MAE	MSE	R <sup>2</sup>	EVS
Linear regression	417966.71	1104813484865.54	0.7333	0.7373
Bayesian ridge	418124.25	1104850876069.80	0.7338	0.7378
Gradient boost	400171.3	1251050671665.5	0.6625	0.6689
Lasso regression	417966.82	1104814155048.98	0.7333	0.7373

For Hold-out validation, we can see that the Gradient Boost model performed the best overall, achieving the lowest MSE, MAE, and highest R-squared and EVS scores.

**Table 3.** Hold-out validation evaluation performance through default parameters

Model	MAE	MSE	R <sup>2</sup>	EVS
Linear regression	450278.72	1392541484222.29	0.7898	0.7925
Bayesian ridge	450764.98	1405205956080.20	0.7878	0.7906
Gradient boost	419569.27	1096599623150.95	0.8344	0.8356
Lasso regression	450278.87	1392557148349.98	0.7898	0.7925

With the use of GASearchCV for optimization, the best hyperparameters for each algorithm that generated highest value of R-squared are:

- Multiple Linear Regression: ‘normalize’: False
- Bayesian Ridge Regression: ‘alpha\_1’: 0.0004915314720013189, ‘alpha\_2’: 0.0021894794843560963, ‘lambda\_1’:0.005960462584099383, ‘lambda\_2’: 0.009450054454871687
- Gradient Boosting Regression: ‘learning\_rate’: 0.033298366669962574, ‘subsample’: 0.5634666800882826, ‘n\_estimators’: 57, ‘max\_depth’: 4
- Lasso Regression: alpha of 0.0099939746971181

Tables 4 and 5 present the results of our model evaluation using the MSE, MAE, R-squared, and EVS metrics for each of the four regression models that we used (Linear Regression, Bayesian Ridge, Gradient Boost, and Lasso Regression) using optimized parameters from GASearchCV. Tables 4 and 5 were evaluated through Cross-fold validation and Hold-out Validation, respectively [14].

For Cross-fold validation, we can see that the Gradient Boost model performed the best overall, achieving the highest R-squared and EVS scores and lowest MSE Score, while the Linear regression model presented a lowest MAE. However, all four models

performed relatively well, with  $R^2$  scores ranging from 0.7333 to 0.7508, and EVS scores ranging from 0.7373 to 0.7640.

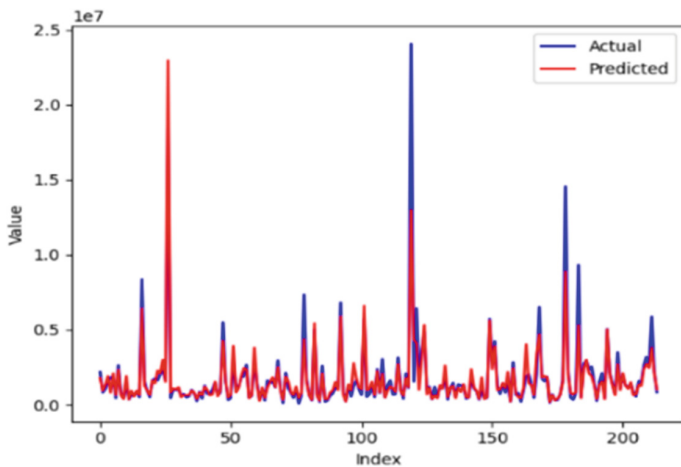
**Table 4.** Cross-fold validation evaluation performance through optimized parameters

Model	MAE	MSE	$R^2$	EVS
Linear Regression	417966.71	1104813484865.54	0.7333	0.7373
Bayesian Ridge	418124.64	1104851889106.48	0.7338	0.7378
Gradient Boost	429877.67	1093145177225.57	0.7508	0.7640
Lasso Regression	417966.71	1104813491630.05	0.7333	0.7373

For Hold-out validation, we can see that the Linear Regression model performed the best overall, achieving the lowest MSE, MAE, and highest R-squared and EVS scores.

**Table 5.** Hold-out validation evaluation performance through optimized parameters

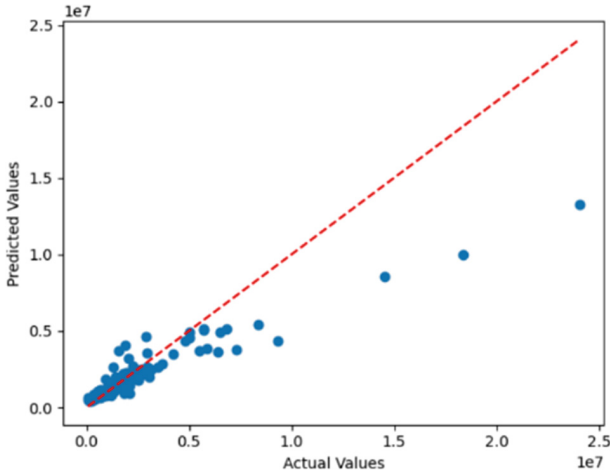
Model	MAE	MSE	$R^2$	EVS
Linear regression	450278.72	1392541484222.29	0.7898	0.7925
Bayesian ridge	450766.49	1405245581297.19	0.7878	0.7906
Gradient boost	503734.71	1528845463101.02	0.7692	0.7726
Lasso regression	450278.72	1392541640769.94	0.7898	0.7925



**Fig. 2.** Trend lines of actual vs. predicted value for optimized gradient boost model

As the optimized gradient boost model performed best overall using Cross-fold Validation, we can also see in Fig. 2 that it generated prediction results that are generally

consistent with the trend of real value since the trend moves along the same direction with each other.



**Fig. 3.** Scatter plot of actual vs. predicted value for optimized gradient boost model

We can also see that it generated a scatter plot Fig. 3, containing relatively accurate predictions. The plot is closer to the diagonal line, which suggests that the model’s predictive performance is good.

## 5 Conclusion

In this study, we employed machine learning models to develop an intelligent system that predicts house prices in Manila City. The dataset used in this study was obtained from the Department of Public Works and Highways Public-Private Partnership Service (DPWH-PPPS) and it pertains to the current market value of the improvements affected by the NLEX-SLEX Connector Road Project right-of-way. It includes various features such as floor area in sqm, number of storeys, classification, type of structure, additional floors or areas, and tag ID. After performing feature engineering and Genetic algorithm optimization, it was determined that Gradient Boosting Regressor was the top-performing model in predicting house prices in Manila City, with an R-squared score of 0.7508 and EVS value of 0.7640. This study demonstrated the usefulness of machine learning in predicting house prices and could aid in making informed decisions for real estate investment in Manila City. In future studies, enhancing the performance of the model can be achieved by integrating a more comprehensive dataset and diverse features.

**Acknowledgments.** The authors would like to acknowledge the support and contributions of all the individuals and organizations that have helped in the successful completion of this study. Special thanks to the Department of Public Works and Highways Public-Private Partnership Services for providing access to the dataset needed in this research.



## References

1. Yusof, A., Ismail, S.: Multiple regressions in analysing house price variations. *Commun. IBIMA* 1–9 (2012)
2. Zhang, Q.: Housing price prediction based on multiple linear regression. *Sci. Program.* **2021**, 1–9 (2021)
3. Trawinski, B., Telec, Z., Krasnoborski, J., Piwowarczyk, M., Talaga, M., Lasota, T., Sawilow, E.: Comparison of expert algorithms with machine learning models for real estate appraisal. In: 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications (INISTA). (2017)
4. Rosales, M.A., Magsumbol, J.-V., Palconit, M.G., Culaba, A.B., Dadios, E.P.: Artificial Intelligence: the technology adoption and impact in the Philippines. In: 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) (2020)
5. Rawool, A.G.: House price prediction using machine learning. *Iconic Res. Eng. J.* **4**(11) (2021)
6. Mu, J., Wu, F., Zhang, A.: Housing value forecasting based on machine learning methods. *Abstr. Appl. Anal.* **2014**, 1–7 (2014)
7. Fan, C., Cui, Z., Zhong, X.: House prices prediction with machine learning algorithms. In: Proceedings of the 2018 10th International Conference on Machine Learning and Computing (2018)
8. Shaikh, A.: House price prediction using multi-variate analysis. *Int. J. Creat. Res. Thoughts* **8** (2020)
9. Satish, G.: House price prediction using machine learning. *Int. J. Innov. Technol. Expl. Eng. (IJITEE)* **8**(9) (2019)
10. Mora-Garcia, R.-T., Cespedes-Lopez, M.-F., Perez-Sanchez, V.R.: Housing price prediction using machine learning algorithms in COVID-19 times. *Land* **11**, 2100 (2022)
11. Li, M., Bao, Z., Sellis, T., Yan, S., Zhang, R.: HomeSeeker: a visual analytics system of real estate data. *J. Vis. Lang. Comput.* **45**, 1–16 (2018)
12. Phan, T.D.: Housing price prediction using machine learning algorithms: the case of Melbourne City, Australia. In: 2018 International Conference on Machine Learning and Data Engineering (iCMLDE) (2018)
13. Truong, Q., Nguyen, M., Dang, H., Mei, B.: Housing price prediction via improved machine learning techniques. *Procedia Comput. Sci.* **174**, 433–442 (2020)
14. Rosales, M.A., Bandala, A.A., Vicerra, R.R., Dadios, E.P.: Physiological-based smart stress detector using machine learning algorithms. In: 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM) (2019)
15. de Luna, R.G., Dadios, E.P., Bandala, A.A.: Automated image capturing system for deep learning-based tomato plant leaf disease detection and recognition. In: TENCON 2018 - 2018 IEEE Region 10 Conference (2018)
16. Thomas, J.J., Ali, A.M.: Dispositional learning analytics structure integrated with recurrent neural networks in predicting students performance. In: Vasant, P., Zelinka, I., Weber, G.-W. (eds.) ICO 2019. AISC, vol. 1072, pp. 446–456. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-33585-4\\_44](https://doi.org/10.1007/978-3-030-33585-4_44)
17. Oyedele, A.A., Ajayi, A., Oyedele, L., Bello, S.A., Jimoh, K.O.: Performance comparison of deep learning and boosted trees for cryptocurrency closing price prediction. *SSRN Electron. J.* (2022)

18. Bujorianu, A.: Developing a Recommendation Algorithm for Patients Using the Healthentia Platform. University of Twente (2022)
19. Gao, G., Bao, Z., Cao, J., Qin, A.K., Sellis, T.: Location-centered house price prediction: a multi-task learning approach. *ACM Trans. Intell. Syst. Technol.* **13**, 1–25 (2022)



# An Analysis of the Relationship Between Temperature Rise and GDP Using Machine Learning Techniques

Iwan Halim Sahputra<sup>(✉)</sup>, Feren Nathan Wedianto, Novendra Imanuel, Daniel Jaya, and Andre

Industrial Engineering Department, Petra Christian University, Siwalankerto 121-131, 60236 Surabaya, Indonesia  
iwanh@petra.ac.id

**Abstract.** Research has been conducted using random forest, gradient boosting tree, decision tree, and linear regression algorithms, using cumulative GDP data and GDP for each agricultural, industrial, manufacturing, and service sector. The goal is to analyze the relationship between these factors on the temperature rise in nine countries. First, the decision tree and linear regression algorithms are used for preliminary relationship analysis. Then, the more complex algorithms, i.e., random forest and gradient boosting tree, are developed for further predictions. It is found that the best algorithm trained with the training data is the gradient boosting tree algorithm using 100 trees. However, when tested for accuracy with the testing data, it has a more significant error than the error using training data. This result concludes that the model produces accurate predictions only for the training data.

**Keywords:** Temperature rise · GDP · Decision tree · Random forest · Gradient boosting tree

## 1 Introduction

Global warming is one of the environmental issues that is currently being discussed until now by state leaders and scientists around the world. This increase in global temperature has occurred since 1896 and continues to experience growth in temperature to this day [1]. In 2021, there is an increase in the average temperature on the global surface by 0.84 °C. It is the sixth highest-ranked increase from 1888–2021 [2]. If this continues, it will threaten life on this earth because its surface temperature is too hot, making it uninhabitable for humans. Therefore, on December 12, 2015, as many as 196 countries reached an international agreement on climate change, commonly known as “The Paris Agreement.” This agreement limits global temperature rise to below 2 °C and preferably only 1.5 °C [3].

One of the causes of global warming is the greenhouse effect. The greenhouse effect is caused by gases such as carbon dioxide (CO<sub>2</sub>) and methane (CH<sub>4</sub>) in the atmosphere.

These gasses trap the heat that should be reflected outside the earth [4]. Most greenhouse gases arise from human activities such as using vehicles, carrying out production activities, and other activities that require fossil fuels as the primary energy.

In several previous studies regarding global temperature rise on earth, one of the researchers used a mathematical method approach to see the relationship between an increase in GDP (Gross Domestic Product), which is the sum of the values of the results of the products and services produced representing human activities towards rising temperatures on earth [5]. Other studies use the random forest, support vector regression, lasso, and linear regression algorithms to look at the relationship between global average temperature trends over the last 70 years and the concentrations of carbon dioxide (CO<sub>2</sub>), nitrous oxide (N<sub>2</sub>O), and methane (CH<sub>4</sub>) [6]. Studies are also carried out by observing temperature and humidity using several machine learning algorithms, namely Gradient Boosting Tree (GBT), Random Forest (RF), Multi-Layered Perception (MLP), and Radial Basis Function (RBF). This study found that MLP is the best algorithm for predicting daily relative humidity [7].

Back to GDP, each GDP sector may contribute differently to greenhouse gas emissions, the primary cause of global temperature rise. Understanding the relationship between the GDP sector and temperature rise can identify which industries are the most significant contributors to climate change and develop strategies to reduce their emissions [11–13]. The relationship between the GDP sector and temperature rise is also essential for policymakers responsible for developing policies to address climate change. Policymakers can identify ways to promote economic growth while reducing its negative environmental impacts by studying the relationship between the GDP sector and temperature rise. In addition, analyzing the relationship between the GDP sector and temperature rise can help policymakers understand other countries' different challenges and priorities and develop strategies for international cooperation on climate change.

In this study, random forest and gradient-boosting tree are used to study the relationship between GDP and temperature rise in several countries. Both algorithms are supervised learning algorithms in machine learning. Random forest is an algorithm where the output comes from a combination of the outcomes issued by a decision tree [8]. The advantages of random forest are they can handle various types of missing data, overcome interactions and nonlinearity, scaling to high dimensions to avoid overfitting, and produce critical variable sizes useful for variable selection.

Gradient-boosting tree or gradient boosting or gradient boosted tree is an algorithm that uses decision trees to create weak classifiers, built into the boosting using four to 8 tree levels [9]. Gradient-boosting tree can be used for regression (when the target variable is continuous) or classification (the target variable is categorical) [10]. In addition, gradient boosting has adaptable properties, such as able to fill in missing values in the data. The two algorithms will likely show the relationship and produce accurate prediction results. The data that will be used in this research is the GDP of several countries to represent human activities and their temperature data. The two algorithms will be compared to see the best prediction algorithm.

## 2 Method

The GDP data is obtained from the World Bank's website (<https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>). The data includes overall GDP and GDP for sectors: agriculture, industry, manufacturing, and services. All GDP data is normalized to minimize bias. Temperature data is obtained from NASA's website (<https://data.giss.nasa.gov/gistemp/>). The temperature and GDP data are obtained from 9 selected countries: South Africa, Brazil, Chile, India, South Korea, France, Singapore, Togo, and Turkey. They are chosen because of the completeness of the data available. The temperatures are measured at weather stations located in the capital cities of each country. Figure 1 shows an example of GDP and temperature raw data for Brazil.

In this study, random forest and gradient boosting tree algorithms are the machine learning algorithms used in developing the prediction model. The model development uses an open-source machine learning and data visualization program: Orange (<https://orangedatamining.com>). Orange is a powerful platform for data analysis and visualization, seeing data flow, and becoming more productive.

Random forest is an ensemble learning method for classification, regression, and other tasks. Random Forest builds a set of decision trees. Each tree is developed from a bootstrap sample from the training data. When creating individual trees, an arbitrary subset of attributes is drawn (hence the term "Random"), from which the best feature for the split is selected. The final model is based on the majority vote from individually developed trees in the forest.

A gradient-boosted model is an ensemble of either regression or classification tree models. Both are forward-learning ensemble methods that obtain predictive results through gradually improved estimations. Boosting is a flexible nonlinear regression procedure that helps improve the accuracy of trees. A series of decision trees are created that produce an ensemble of weak prediction models by sequentially applying weak classification algorithms to the incrementally changed data. While boosting trees increases their accuracy, it also decreases speed and human interpretability. The gradient boosting method generalizes tree boosting to minimize these issues.

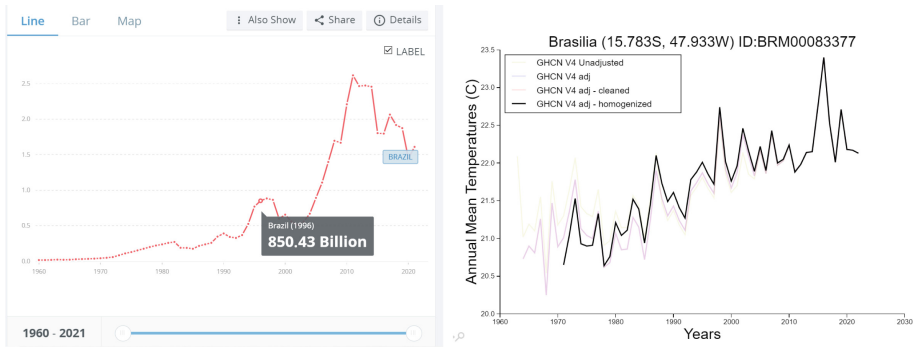
Fine-tuning the model is done using a grid search. The model evaluation uses error calculations such as MSE, RMSE, and MAE. The errors are first evaluated from the prediction results using training data to determine the best algorithm. Then, the selected algorithm is evaluated using testing data to see the final model's accuracy.

In addition to the random forest and gradient boosting tree algorithms, a decision tree model is also used to evaluate the impact of each GDP sector on the temperature. The decision tree model is also developed using the Orange program.

## 3 Results and Discussion

For the preliminary analysis, we use a linear regression approach to see the overall GDP's effect on each country's average temperature. The error and the plot are shown in Table 1 and Fig. 2, respectively. From those, it is evident that the relation between the overall GDP and the average temperature is not simply linear. Therefore, we need to use non-linear models to analyze the relation better. Thus, we use a decision tree model to

evaluate the impact of each GDP sector (i.e., agriculture, industry, manufacturing, and services) on the temperature.



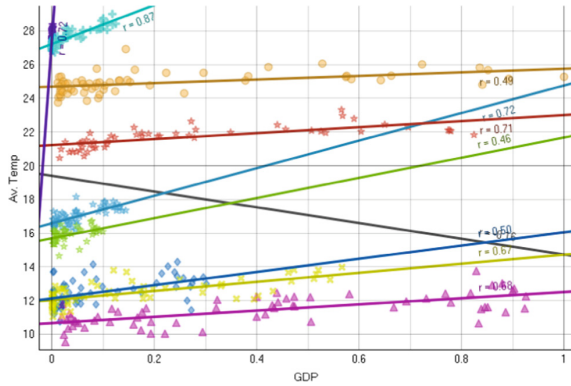
**Fig. 1.** Brazil’s GDP and temperature raw data from data.worldbank.org and data.giss.nasa.gov, respectively

**Table 1.** Errors for the linear regression model of overall GDP-average temperature

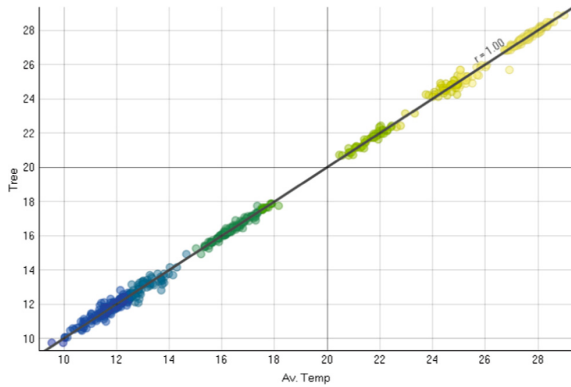
Country	MSE	RMSE	MAE	R2
South Africa	0.079	0.282	0.227	0.615
Brazil	0.100	0.317	0.246	0.732
Chile	0.076	0.275	0.219	0.250
India	0.231	0.480	0.359	0.316
South Korea	0.353	0.594	0.553	0.470
France	0.395	0.628	0.498	0.512
Singapore	0.097	0.312	0.257	0.722
Togo	0.026	0.163	0.137	0.798
Turkey	0.369	0.608	0.458	0.261

The developed decision tree model has MSE, RMSE, MAE, and R2 of 0.080, 0.284, 0.187, and 0.997, respectively. A comparison between the average temperature and the predicted temperature from the model for all countries is shown in Fig. 3. From the model, the most significant GDP sector(s) affecting the temperature of each country are shown in Table 2.

The decision tree model for India can be seen in Fig. 4 as an example of the agriculture and industry sectors dominantly affecting the temperature. On the other hand, for Singapore, like France, industry, manufacturing, and service sectors dominantly affect the temperature, as shown in Fig. 5. While, for Turkey, like Brazil, service and industry sectors dominantly affects the temperature, as shown in Fig. 6.



**Fig. 2.** Linear regression model for overall GDP-average temperature



**Fig. 3.** The average temperature (Av. Temp) vs. the predicted temperature (Tree) by the decision tree model.

The agriculture sector is a significant contributor to global warming. Here are some of the ways that agriculture can influence temperature:

1. Land use changes: Agriculture often requires the clearing of land, which can release large amounts of carbon dioxide (CO<sub>2</sub>) into the atmosphere. Deforestation for agricultural purposes is a significant contributor to greenhouse gas emissions.
2. Livestock emissions: Livestock such as cows, sheep, and goats produce large amounts of methane, a potent greenhouse gas more effective than carbon dioxide at trapping heat in the atmosphere. Livestock also produces nitrous oxide, another potent greenhouse gas.
3. Fertilizer use: Nitrous oxide is also released from using fertilizers in agriculture. The production and use of synthetic fertilizers release large amounts of nitrous oxide into the atmosphere.
4. Transportation: Agriculture is also dependent on transportation, which can result in the emissions of greenhouse gases from fuel combustion.

**Table 2.** Most significant GDP sectors affecting the temperature of each country

Country	Dominant GDP sector affecting temperature
South Africa	Service, agriculture, and manufacturing
Brazil	Service and industry
Chile	Agriculture and industry
India	Agriculture and industry
South Korea	Industry, service, agriculture
France	Manufacturing, industry, and service
Singapore	Industry, manufacturing, and service
Togo	Service and manufacturing
Turkey	Service and industry

5. Soil degradation: Agriculture can lead to soil degradation, which can reduce the ability of soil to sequester carbon. As a result, the carbon that would have been stored in the ground is released into the atmosphere.

Manufacturing and other industry sectors are also significant contributors to global warming. Here are some of the ways that manufacturing can influence climate change:

1. Energy use: Manufacturing facilities typically use a lot of energy, much of which comes from burning fossil fuels. This combustion releases greenhouse gases, such as carbon dioxide, into the atmosphere.
2. Emissions from industrial processes: Many industrial processes also produce greenhouse gases, including methane and nitrous oxide. For example, cement, steel, and aluminum production are significant sources of greenhouse gas emissions.
3. Transportation: Transporting raw materials and finished products also contributes to greenhouse gas emissions. It is particularly true for products that are transported long distances.
4. Waste: Manufacturing often produces waste products that release greenhouse gases as they decompose. For example, landfills produce methane, a potent greenhouse gas.
5. Water use: Manufacturing requires large amounts of water, leading to energy-intensive water treatment, distribution processes, and water-related greenhouse gas emissions.

The service sector can significantly negatively affect global warming, particularly energy consumption and carbon emissions. Here are some examples of how the service sector can contribute to global warming:

1. Transportation: The service sector relies heavily on transportation for the delivery of goods and services and for the movement of people. The transportation sector is a significant contributor to greenhouse gas emissions, and the service sector can contribute significantly to this through its reliance on shipping, air travel, and road transportation.
2. Energy-intensive infrastructure: The service sector relies heavily on energy-intensive infrastructure, such as data centers, servers, and communication networks. These



facilities require significant energy to operate, and much of this energy is generated from fossil fuels. In addition, the use of energy-intensive technologies and equipment can also contribute to the sector's carbon footprint.

3. Resource consumption: The service sector also consumes significant resources like water and paper. The production and disposal of these resources can also contribute to greenhouse gas emissions and other environmental impacts.
4. Increased consumption: The service sector can also contribute to global warming by promoting consumption and consumerism. For example, advertising and marketing campaigns can encourage people to buy more goods and services, increasing the demand for energy and resources and contributing to carbon emissions.

After finding the dominant sectors of GDP affecting the temperature rise using a decision tree model, we develop random forest and gradient-boosting tree models to predict the temperature rise. Fine-tuning for both models is done using the grid search method. This study only adjusts one hyperparameter: the number of trees for the random forest and the gradient-boosting tree algorithms. Each has four levels (25, 50, 75, and 100 trees) of hyperparameters to be compared using the training data.

Based on the results as shown in Table 3, it can be concluded that the best algorithm with the smallest error value is the gradient boosting tree algorithm with a hyperparameter of 100 trees. Then, the gradient-boosting tree model with 100 trees is evaluated further using data testing. As a result, the MSE, RMSE, and MAE using data testing are 0.248, 0.498, and 0.381, respectively.

**Table 3.** Grid search result for model fine-tuning

Model	Number of trees	MSE	RMSE	MAE
Random forest	25	1.541	1.241	0.825
	50	1.546	1.243	0.821
	75	1.544	1.243	0.82
	100	1.546	1.243	0.82
Gradient boosting tree	25	1.634	1.278	1.153
	50	0.254	0.504	0.402
	75	0.144	0.379	0.29
	100	0.115	0.339	0.259

Based on comparing the results, the errors evaluated using data testing are greater than in the training data. Therefore, the current model needs to be redeveloped. First, adding more data to the training data is necessary, and then the model development will be carried out again. However, this process could not be performed due to the limited data available.

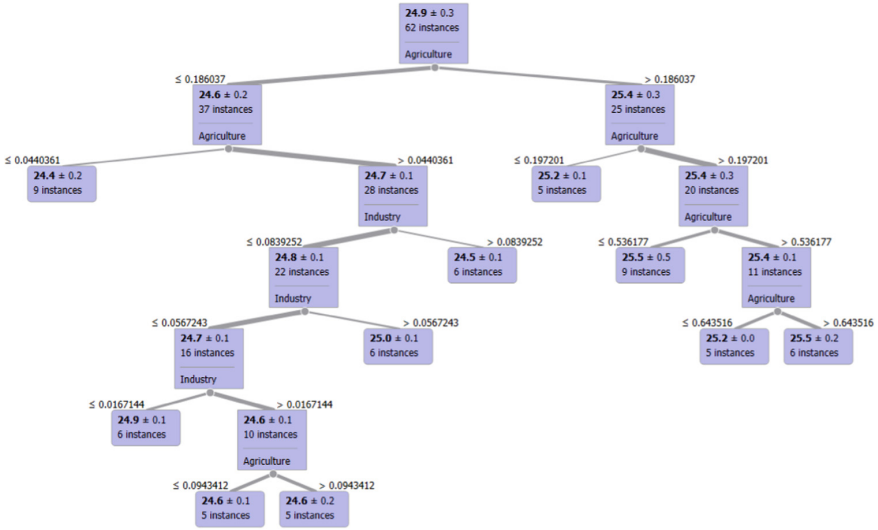


Fig. 4. Decision tree model for India

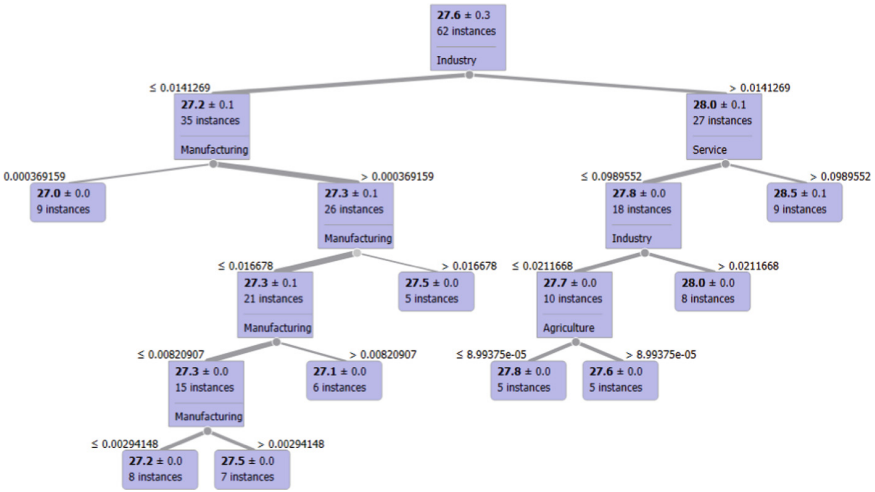


Fig. 5. Decision tree model for Singapore

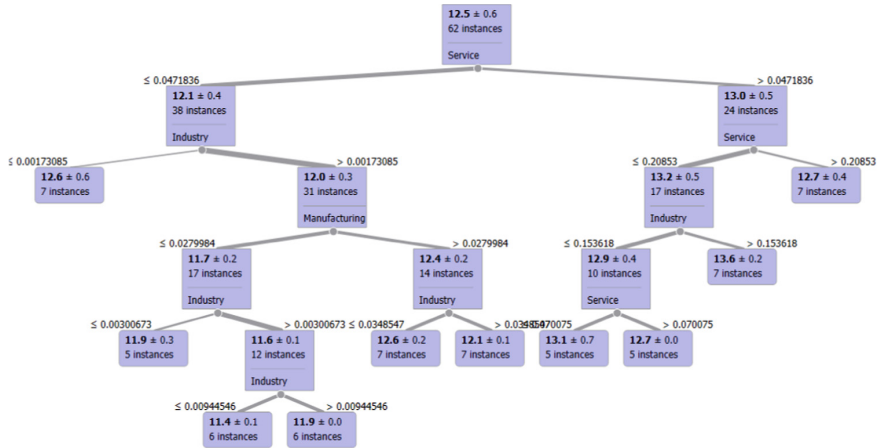


Fig. 6. Decision tree model for Turkey

## 4 Conclusion

Research has been conducted using decision tree, random forest, and gradient boosting tree algorithms, using cumulative the GDP and GDP of agricultural, industrial, manufacturing, and service sectors. Using the decision tree model, we can find the dominant GDP sectors affecting the temperature rise for each country. For example, the agriculture and industry sectors in India dominantly affect the temperature. In Singapore, like in France, industry, manufacturing, and service are the dominant sectors, while in Turkey, like in Brazil, service and industry sectors. For prediction, it is found that the best algorithm trained with the training data is the gradient-boosting tree algorithm using 100 trees. However, when tested for accuracy with the testing data, it has a more significant error than the error using training data. This result concludes that the model produces accurate predictions only for the training data. Therefore, the current model needs to be redeveloped by adding more data to the training data and fine-tuning other parameters.

## References

1. Weart, S.: Global warming: how skepticism became denial. *Bull. At. Sci.* **67**(1), 41–50 (2011)
2. National Centers for Environmental Information: Assessing the Global Climate in 2021 (2022), Retrieved from <https://www.ncei.noaa.gov/news/global-climate-2021> 12. 20 Dec 2022
3. United Nation Climate Change, The Paris Agreement, n.d., Retrieved from <https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement>. 20 Dec 2022
4. Houghton, J.: Global warming. *Rep. Prog. Phys.* **68**(6), 1343–1403 (2015)
5. Escamilla, J.B.G., Pérez, S.L., Martínez, Y.R.: Analysis on the effect of human activities in the average global temperature. *Int. J. Engl. Lit. Soc. Sci.* **2**(6), 161–170 (2017)
6. Zheng, H.: Analysis of global warming using machine learning. *Comput. Water Energy Environ. Eng.* **7**(3), 127–141 (2018)

7. Hanoon, M.S., Ahmed, A.N., Zaini, N., Razzaq, A., Kumar, P., Sherif, M., Sefelnasr, El-Shafie., A.: Developing machine learning algorithms for meteorological temperature and humidity forecasting at Terengganu State in Malaysia. *Sci. Rep.* **11** (2021)
8. Patnaik, S., Sethi, I.K., Yang, X.S.: *Advances in Machine Learning and Computational Intelligence*. Springer, Singapore (2020)
9. Walczak, B., Tauler, R., Brown, S.D.: *Comprehensive Chemometrics Chemical and Biochemical Data Analysis*. Elsevier Science (2020)
10. Semanjski, I., Gautama, S.: Smart city mobility application—gradient boosting trees for mobility prediction and analysis based on crowdsourced data. *Sensors* **15**(7), 15974–15987 (2015)
11. Fischer, C., Newell, R.G.: Environmental and technology policies for climate mitigation. *J. Environ. Econ. Manag.* **55**(2), 142–162 (2008)
12. Goulder, L.H., Stavins, R.N.: Challenges from state-federal interactions in US climate change policy. *Am. Econ. Rev.* **101**(3), 253–257 (2011)
13. Newell, R.G., Prest, B.C., Sexton, S.E.: The GDP-temperature relationship: implications for climate change damages. *J. Environ. Econ. Manag.* **108**, 102445 (2021)



# Machine Learning Techniques for Predicting Remaining Useful Life (RUL) of Machinery for Sustainable Manufacturing Lines

Lim Khai Sian<sup>1</sup> and J. Joshua Thomas<sup>2</sup>(✉)

<sup>1</sup> UOW Malaysia KDU Penang University College, 32, Jalan Anson, 10400 George Town, Penang, Malaysia

<sup>2</sup> UOW Malaysia KDU University College, Jalan Kontraktor U1/14, Glenpark U1, 40150 Shah Alam, Selangor, Malaysia  
jjoshua@uow.edu.my

**Abstract.** Maintenance is a very important activity in all industries, especially the secondary industry related to construction and manufacturing. This can affect the uptime and efficiency of the appliance. Therefore, equipment failures must be identified and corrected to avoid production downtime. With the availability of a large amount of data, Machine Learning (ML) approaches have emerged as a promising tool in maintenance management. Therefore, the aim of this research is to develop a Predictive Maintenance (PdM) strategy with a machine learning approach. We use machine learning algorithms in to develop a predictive model to predict the Remaining Useful Life (RUL) of the equipment in the manufacturing line. The predictive model is built using Gradient Boosting Regression (GBR) algorithm with Principle Component Analysis (PCA) as the feature extraction method. The test results demonstrate that the performance of our proposed model is effective in predicting the RUL of turbo engines and also significantly improves the results of predictive maintenance. Train and test sets, the trained model has RMSEs of 17.760 and 17.371, and  $R^2$ s of 0.818 and 0.825 respectively.

**Keywords:** Predictive maintenance · Remaining useful life · Gradient boosting regression

## 1 Introduction

The global competition and the demand for quick adaptation to the rapid changing market demand is driving industrial production today. In order to fulfil such re-quests, the industries are undergoing “The Fourth Industrial Revolution”. Effective and flexible maintenance management has become increasingly important serving as the primary means to achieve this goal [10]. Thanks to Industry 4.0, data analytics and machine learning techniques can be applied to collected data to support strategic decision making in maintenance management. In order to prevent production downtime due to machine failure, maintenance management has become very crucial. The two major maintenance

strategies adopted by industries are Run-To-Failure (R2F) and Preventive (PvM) maintenance [15]. R2F maintenance takes place when equipment failure occurs while PvM maintenance takes place periodically according to a schedule. These approaches are inefficient as they may result in production line failure or costly maintenance. In order to improve production efficiency and reduce production costs, Predictive Maintenance (PdM) is introduced. Many Data-driven approaches such as Neural Networks, and Random Forest have been introduced. Therefore, this research utilizes a different machine learning algorithm named Gradient Boosting Regression to build the predictive model.

## 2 Literature Review

Common Machine Learning algorithms used in predictive maintenance and the related papers are discussed. It also includes two main techniques used in this research which are Gradient Boosting Regression and Principle Component Analysis. R2F maintenance only takes place when an equipment failure occurs [2]. This is the simplest maintenance approach, but it is also the least efficient approach. The cost of downtime after failure is higher than the cost of maintenance conducted in advance. PvM maintenance takes place periodically according to a planned schedule with time or process iterations [15]. Although this approach can prevent failure, unnecessary corrective actions are often performed causing waste of resources and increased operating costs. PdM maintenance takes place when corrective actions are needed. It is based on continuous monitoring of equipment, allowing maintenance to be performed timely [2]. Besides, a predictive tool based on historical data is used to enable the early detection of failure. Since a good maintenance approach should minimize equipment failure and operating cost, while maximizing the life of equipment, Predictive Maintenance is selected in this research. This research focus on a machine learning-based data-driven approach as the model-based and hybrid approach requires expert knowledge of the equipment. From [16–23] are the authors contributions to the various research area with Machine learning and deep learning approaches.

### 2.1 Machine Learning

Machine learning has emerged as a prominent method for constructing an effective predictive model in various applications due to the ability to interpret high dimensional and multivariate data and extract the hidden correlations within data. However, the performance of machine learning is depending on the suitable choice of ML algorithms. This section discusses the main machine learning algorithms used in the PdM field. Artificial Neural Network (ANN) is one of the most common ML algorithms due to its excellent performance in many industrial applications. In [5] Multilayer Perceptron (MLP) is proposed to predict the RUL of aero-engines. The authors proposed a moving time window to consider the values in prior cycles. In this case, the input to the network is a time window, thus the dimension of the input increases as a multiple of the window size [9]. In order to achieve the same effect as the time-window based method [8], proposed a Recurrent Neural Network (RNN) based on an encoder-decoder framework to predict the Health Index (HI) of the rolling bearing. The RUL of the component is then retrieved

from a linear regression model as the HI is constructed closely related to the RUL. In [3] Random Forest (RF) is used to create a predictive model to predict the failure of wind turbines. This work proposes an improvement to the artificial immune network algorithm adopted in [16] in terms of speed, scalability and accuracy. The author is able to achieve 5–9% of accuracy improvement compared to the original solution [3, 7, 8]. In [13], RF is used to develop a predictive maintenance system named HDPass to detect the Hard Disk Drive (HDD) failure in data centers. The authors train the model on historical data and utilize data collected from end-users to perform real-time predictions [14]. Other than the papers mentioned above, many authors also selected the RF model to compare against their proposed methods. Support Vector Regression (SVR) is another algorithm used in PdM by [21]. Un-like conventional SVR, a modified regression kernel that can be used for SVR is proposed for prognostic problems. In this work, the authors modify the kernel to account for different cycles in the prognostic domain and integrate the time dependencies [10].

## 2.2 Gradient Boosting Regression

The works in the Random Forest section had proven that ensemble methods can be used in the data-driven approach of PdM. In this research, another ensemble method named Gradient Boosting Regression (GBR) is used to train the predictive model. Approaches such as Bootstrap Aggregation (Bagging) and Random Forest are built on the idea of creating base learners independently and combining their results via certain deterministic averaging processes [13]. However, boosting method is based on different strategies to combine the weak learners. The concept of boosting arose from the idea of whether a weak learner can be improved. It is articulated by Michael Kearns as the “Hypothesis Boosting Problem” [7]. Hypothesis Boosting was the concept of filtering observations, keeping those that can be ad-dressed by weak learners and focusing on building new weak learners that can ad-dress the remaining observations [1]. Gradient Boosting algorithm can be best described by first introducing the Adaptive Boosting also known as AdaBoost. The weak learners in AdaBoost are decision stumps which are decision trees with one split. AdaBoost works by increasing the weight of difficult observations and lowering the weight of easier observations [13]. New weak learners are then added sequentially to solve more difficult patterns. The final result is voted by the majority prediction of weak learners weighted by their associated accuracy. Gradient Boosting is a generalization of AdaBoost which is an optimization problem in which the goal is to minimize the loss function by adding new weak learners via a gradient descent like process [1]. This generalization allows differentiable loss functions to be used thus expanding the applicability of the algorithm to regression and multi-class classification problems.

### Algorithm

The first step of this algorithm is to create a base model to predict the data. The predicted value of the base model is the average of the dependent variables [11]. The idea of the

average of the dependent variables come from formula 1.

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, \gamma) \tag{1}$$

Here,  $L$  is the loss function and  $\gamma$  is the predicted value. The goal of this formula is to find a  $\gamma$  value where the loss function is minimum. In this research, the dependent variable is a continuous value, thus the loss function is:

$$L = \frac{1}{n} \sum_{i=0}^n (y_i - \gamma_i)^2 \tag{2}$$

Here,  $y_i$  is the dependent variable and  $\gamma_i$  is the predicted value. In order to find the minimum value of  $\gamma$  such that the loss function is minimum, the derivative of the loss function is required as shown in formula 3.

$$\frac{dL}{d\gamma} = \frac{2}{2} \left( \sum_{i=0}^n (y_i - \gamma_i) \right) (-1) \tag{3}$$

The next step is to calculate the pseudo residuals. The pseudo residual is equal to the dependent variable – predicted value. This can be proven by formula 4.

$$\gamma_{im} = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x) = F_{m-1}(x)}, \text{ for } i = 1, \dots, n \tag{4}$$

Here,  $F(x_i)$  is the previous model trained and the  $m$  is the number of decision trees made. The derivative of the loss function is calculated:

$$\frac{dL}{d\gamma} = -(y_i - \gamma_i) = -(\text{dependent variable} - \text{predicted value}) \tag{5}$$

After the residual is calculated, a decision tree is created by using the residual and the independent variables (Gupta, 2021).

The third step is to apply formula 6.

$$\gamma_m = \arg \min_{\gamma} \sum_{i=1}^n L(y_i, F_{m-1}(x_i) + \gamma h_m(x_i)) \tag{6}$$

Here,  $h_m(x_i)$  is the decision tree made based on the residual and the  $m$  is the number of decision trees made. For example,  $m = 1$  refers to the first decision tree.

Formula 6 is similar to formula 1 but here the previous prediction is taken instead of the average of the dependent variable because there is no previous prediction in step 1 [11].

The last step is to update the predictive model by using formula 7

$$F_m(x) = F_{m-1}(X) + \alpha(h_m(x)) \tag{7}$$



Here, the  $F_{m-1}(x)$  is the prediction of the previous model,  $\alpha$  is the learning rate and  $h_m(x)$  is the recent decision tree made based on the residuals.

In order to speed up the training process of the predictive model, a feature extraction method named Principle Component Analysis (PCA) is used. It is a dimensionality reduction method that aims to reduce the dimension of a large dataset while retaining most of the information in the large set [6]. Since the main focus of the research is not PCA, the complex mathematical calculation is not being discussed.

### 3 Decision of Algorithm and Justification

During the data preprocessing stage, manual feature selection is also conducted to select the most relevant features by analyzing the data visualization graphs. Other than feature selection, model selection is also an important stage in building a machine learning model. After testing and evaluating several models to predict the RUL of the turbofan engine, the GBR algorithm is selected to build the predictive model. GBR is selected instead of other models because it does not have many applications in PdM. GBR is selected because it is also an ensemble method like Random Forest which is able to introduce new weak learners to improve the performance of existing trees. Although Random Forest Regression is also an ensemble method that combines all the outputs from each individual tree, there are several differences among them which make GBR a better option in this research. GBR builds trees one at a time and is added to the set to minimize the loss of the model [11]. However, RF build trees independently using a random sample. Another difference between GBR and RFG is the way the outputs are added. GBR combines the outputs from all trees along the training process while RF combines the output at the end of the training process [4].

### 4 Implementation

Exploratory Data Analysis (EDA) on NASA's turbofan engine degradation simulation dataset (C-MAPSS) to summarize their main characteristics. The EDA includes various data visualization graphs to understand the underlying pattern of the dataset. Besides, the process of building the predictive model using various ML algorithms and compare their performance. This process includes data pre-processing, model selection, training model, evaluating model, hyperparameter tuning and making predictions. Exploratory Data Analysis (EDA) Although the C-MAPSS dataset was published over a decade ago, it is still one of the most popular datasets used for PdM. C-MAPSS is a tool that simulates a realistic commercial turbofan engine [12]. The engines in the dataset operate normally in the beginning and develop a fault over time. For the training set, the engines are recorded until a failure occurs, while the test set stops recording some time prior to the failure. The objective is to predict the RUL of the turbofan engines in the test set. In other words, a predictive model is required to predict the remaining number of operation cycles after the latest cycle of the engine (Table 1).

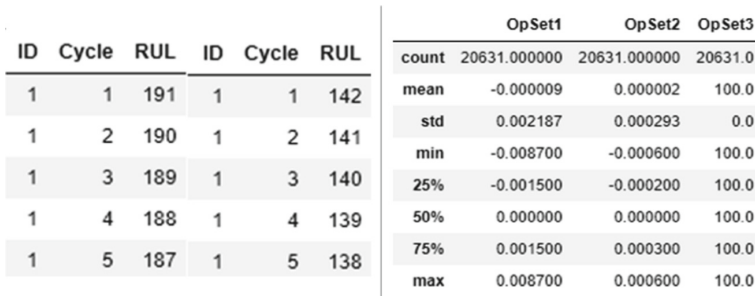
Table 2 shows the data structure of the dataset. Each row contains the engine ID, current cycle, 3 operational settings and 21 sensor values.

**Table 1.** Turbofan engine dataset

Dataset	Train size (no. of engine)	Test size (no. of engine)	Operation conditions	Fault modes
FD001	100	100	1	HPC degradation

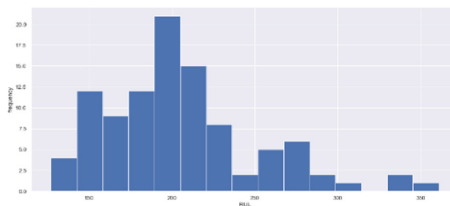
**Table 2.** Data structure of dataset

ID	Cycle	Setting1	Setting2	Setting3	Sensor1	...	Sensor21
Int	Int	Double	Double	Double	Double	Double	Double



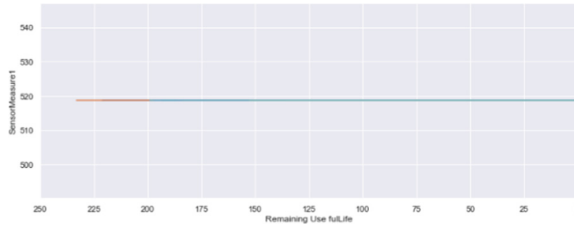
**Fig. 1.** Descriptive statistics of engine ID, cycle and engine settings

Based on the descriptive statistics of engine ID shown in Fig. 1, there is a total of 20631 rows in the dataset and engine ID from 1 to 100. When observing the descriptive statistics of the engine cycle, it shows that the earliest engine failure occurs after 128 cycles whereas the longest operation cycle is 362 cycles. The average operation cycle is between 199 and 206 cycles. However, the standard deviation of 46 cycles is quite large. It shows the descriptive statistics of the operational settings. The standard deviation of setting 3 shows that there is only one value among all the records. Although the values of setting 1 and 2 are not completely stable, the fluctuations are very small. Based on the information gathered, the description of the FD001 dataset is correct where there is only one operating condition.



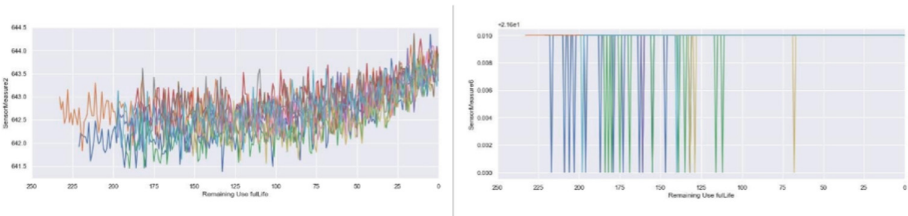
**Fig. 2.** RUL histogram

Figure 2 shows the histogram of RUL. This histogram has reconfirmed the descriptive statistics of the engine cycles above where most engine fails around 200 cycles. The distribution of this histogram is right-skewed where only a few engines can survive more than 300 cycles. Figure 5 shows the graph of sensor 1 with declining RUL as the X-axis. The graphs of sensors 1, 5, 10, 16, 18, and 19 looks similar. The flat line indicates that there are no fluctuations in the values thus these sensors do not hold any useful information.



**Fig. 3.** Graph of sensor 1

Figure 3 shows the graph of sensor 1 with declining RUL as the X-axis. The graphs of sensors 1, 5, 10, 16, 18, and 19 looks similar. The flat line indicates that there are no fluctuations in the values thus these sensors do not hold any useful information.



**Fig. 4.** Graphs of sensor 2 and sensor 6

Figure 4 shows the graph of sensor 2. The value shows an increasing trend as the RUL decreases. A similar trend can be found in sensors 3, 4, 8, 11, 13, 15, and 17. Figure 4 also shows the graph of sensor 6. The value decreases to 0 at a certain point. However, this change does not have any clear relation with the decreasing RUL.

Figure 5 shows the graph of sensor 7. The value shows a decreasing trend as the RUL decreases. The decreasing trend can also be found in sensors 12, 20, and 21. Figure 5 also shows the graph of sensor 9. In this graph, the values of certain engines show an in-creasing trend while others show a decreasing trend. A similar pattern can be found in sensor 14. Based on the EDA conducted, sensors 1, 5, 6, 10, 16, 18 and 19 are deter-mined as sensors that hold no useful information related to RUL.

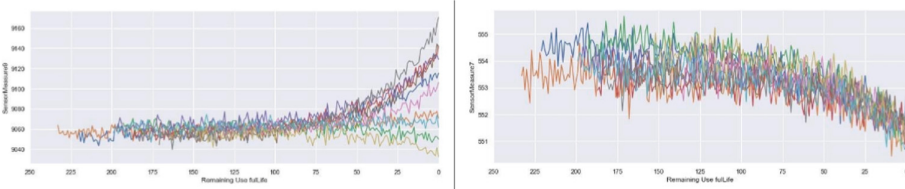


Fig. 5. Graphs of sensor 7 and sensor 9

### 4.1 Predictive Model

After performing EDA on the dataset, the predictive model is built. Other than GBR, various algorithms are used to create the model and compare their performance. Since dataset FD001 operates under 1 operating condition, only the sensor measurements are considered during the training process. Before training the model, data pre-processing is performed.

ID	Cycle	RUL	ID	Cycle	RUL
1	1	191	1	1	142
1	2	190	1	2	141
1	3	189	1	3	140
1	4	188	1	4	139
1	5	187	1	5	138

Fig. 6. Train (left) and test (right) set RUL

Figure 6 shows the calculated RUL in the train and test set. Only the last cycle of each engine in the test set is considered during the training process as they are the records that have true RUL. Both sets are then saved as CSV files to be used later in the training process.

Figure 7 shows the heatmap of the feature selection. The positive value (blue) indicates there is a positive relationship between the two features while the negative value (red) indicates the negative relationship. Sensor 1, 5, 10, 16, 18, and 19 have values 0 related to RUL indicating there is no relationship between them. Sensor 6 has a value very close to 0 thus it can be assumed that the relationship is not significant. Based on the heatmap, the assumption made in EDA is confirmed.

Figure 8 shows the graph of the PCA components. The graph shows that only 10 components are required to hold 95% of the information. The eigenvalues can be assumed as the amount of information the associated components hold. By using PCA feature extraction, the dimension of the dataset is reduced thus fewer computing resources are needed to train the predictive model.

### Gradient Boosting Regression

Gradient Boosting Regression is the main algorithm used in this research. The training process is similar to the previous algorithm but a more tedious hyperparameter tuning is performed after Randomized Search. By using the default parameters, the RMSE of

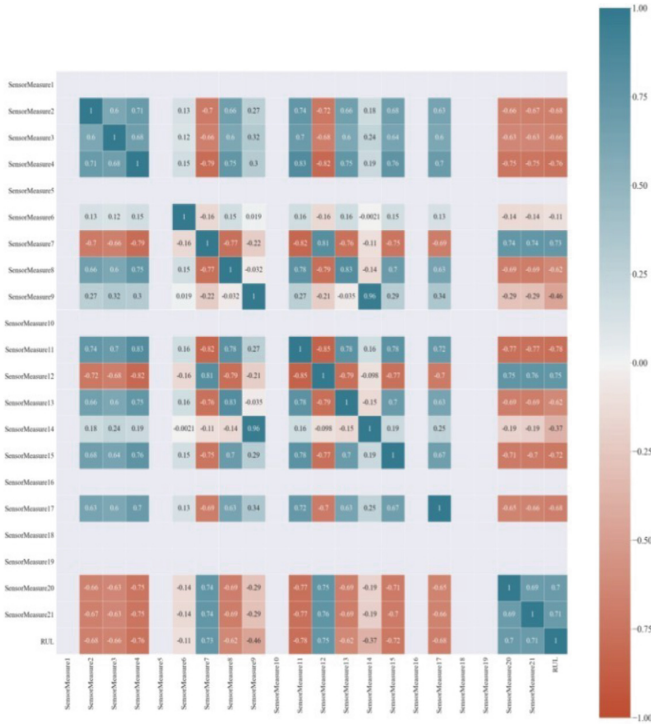


Fig. 7. Heatmap (Pearson)

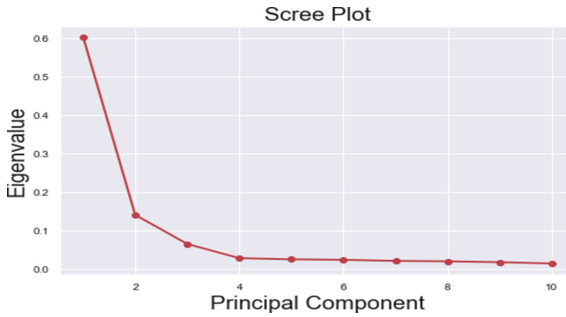


Fig. 8. Graph of PCA

the train and test set are 17.488 and 17.36 respectively and R2 are 0.824 and 0.825 respectively. Next, Randomized Search is performed to tune the parameters. The trained model produces RMSE of 17.760 and 17.371, and R2 of 0.818 and 0.825 for train and test set respectively. After the Randomized Search, Grid Search is performed to further tune the parameters. The trained model produces RMSE of 17.287 and 17.344, and R2 of 0.828 and 0.826 for the train and test set.

**Experimental Results Comparison**

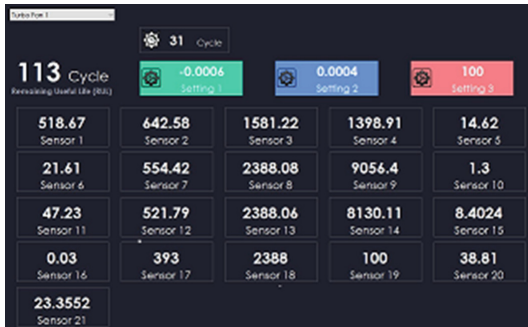
**Table 3.** Performance of various algorithms

Algorithm	RMSE (train set)	RMSE (test set)	R2 (train set)	R2 (test set)
Average model	68.88	52.62	0.0	- 0.60
Linear regression	44.66	31.95	0.58	0.41
Improved linear regression	21.51	21.90	0.73	0.72
Polynomial regression	18.06	18.42	0.81	0.80
Support vector regression	18.13	18.38	0.81	0.80
Random forest	16.51	17.68	0.84	0.82
Gradient boosting regression	17.28	17.34	0.82	0.82

Table 3 shows the result of various algorithms used to predict the RUL of the engine.

### 5 Result and Discussion

To sum up, the predictive maintenance dashboard is software that displays the sensor measurements of an engine and its RUL predicted by the trained predictive model. It combines data visualization and data analytics to ease the process of decision making. It aims to provide critical information to support flexible maintenance strategies such as predictive maintenance. A simple process of the dashboard is the users are able to view the sensor measurements and the RUL of various engines through the engine interface (Fig. 9).



**Fig. 9.** Predictive maintenance dashboard

From this interface, users can understand the current conditions of the engines and decide whether to perform maintenance based on the information provided. The optimum outcome of the research is to train a model that is able to predict the RUL with an accuracy of more than 95%. For now, the algorithm used is the Gradient Boosting Regression. This algorithm is able to provide an accuracy of 82% but a better result is possible with a more complex algorithm that considers the time cycle of the data. It is hoped that the

product developed in this research is able to provide value to businesses that wish to implement flexible maintenance strategies to improve production efficiency.

## References

1. Brownlee, J.: A gentle introduction to the gradient boosting algorithm for machine learning. *Mach. Learn. Mastery* **21** (2016)
2. Carvalho, T.P., Soares, F.A., Vita, R., Francisco, R.D.P., Basto, J.P., Alcalá, S.G.: A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* **137**, 106024 (2019)
3. Canizo, M., Onieva, E., Conde, A., Charramendieta, S., Trujillo, S.: Real-time predictive maintenance for wind turbines using Big Data frameworks. In: 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), pp. 70–77. IEEE (2017)
4. Glen, S.: Decision Tree vs. Random Forest vs. Gradient Boosting Machines: Explained Simply (2019)
5. Sundaram, R.B.: Gradient Boosting Algorithm: A Complete Guide for Beginners. *analyt-icsvidhya* (2021)
6. Jaadi, Z.: A step-by-step explanation of principal component analysis (PCA) (2021). Retrieved 7 June 2021
7. Kearns, M.: Thoughts on hypothesis boosting, ML class project (1988)
8. Chen, Y., Peng, G., Zhu, Z., Li, S.: A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. *Appl. Soft Comput.* **86**, 105919 (2020)
9. Lim, P., Goh, C.K., Tan, K.C.: A time window neural network based framework for remaining useful life estimation. In: 2016 International Joint Conference on Neural Networks (IJCNN), pp. 1746–1753. IEEE (2016)
10. Mathew, J., Luo, M., Pang, C.K.: Regression kernel for prognostics with support vector machines. In: 2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), pp. 1–5. IEEE (2017)
11. Ravanshad, A.: Gradient Boosting vs Random Forest [online]. Medium (2018), Available at: <https://medium.com/@aravanshad/gradient-boosting-versus-random-forest-cfa3fa8f0d80>. Accessed 25 Sept 2021
12. Saxena, A., Goebel, K., Simon, D., Eklund, N.: Damage propagation modeling for aircraft engine run-to-failure simulation. In: 2008 International Conference on Prognostics and Health Management, pp. 1–9. IEEE (2008)
13. Singh, H.: Understanding Gradient Boosting Machines [online]. Towards Data Science (2018), Available at: <https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab>. Accessed 30 Sept 2021
14. Su, C.J., Huang, S.F.: Real-time big data analytics for hard disk drive predictive maintenance. *Comput. Electr. Eng.* **71**, 93–101 (2018)
15. Susto, G.A., Schirru, A., Pampuri, S., McLoone, S., Beghi, A.: Machine learning for predictive maintenance: a multiple classifier approach. *IEEE Trans. Ind. Inf.* **11**(3), 812–820 (2014)
16. Tran, H.N.T., Thomas, J.J., Malim, N.H.A.H.: DeepNC: a framework for drug-target interaction prediction with graph neural networks. *PeerJ* **10**, e13163 (2022)
17. Thomas, J.J., Karagoz, P., Ahamed, B.B., Vasant, P. (eds.): *Deep Learning Techniques and Optimization Strategies in Big Data Analytics*. IGI Global (2019)
18. Murugappan, M., Thomas, J.V.J., Fiore, U., Jinila, Y.B., Radhakrishnan, S.: Covidnet: implementing parallel architecture on sound and image for high efficacy. *Future Internet* **13**(11), 269 (2021)

19. Nain, F.N.M., Malim, N.H.A.H., Thomas, J.J., Tan, M.L.: Focus web crawler on drug herbs interaction patterns. *Informatika* **46**(4) (2022)
20. Yudaev, I., Eviev, V., Sumyanova, E., Romanyuk, N., Daus, Y., Panchenko, V.: Methodology and modeling of the application of electrophysical methods for locust pest control. In: Vasant, P., Weber, G.W., Marmolejo-Saucedo, J.A., Munapo, E., Thomas, J.J. (eds.) *Intelligent Computing & Optimization. ICO 2022. Lecture Notes in Networks and Systems*, vol. 569. Springer, Cham (2023). [https://doi.org/10.1007/978-3-031-19958-5\\_74](https://doi.org/10.1007/978-3-031-19958-5_74)
21. Hai, D.S.T., Joshua Thomas, J., Jothi, J.A., Rasalingam, R.-R.: Detection of Invertebrate Virus Carriers Using Deep Learning Networks to Prevent Emerging Pandemic-Prone Disease in Tropical Regions. In: Vasant, P., Zelinka, I., Weber, G.-W. (eds.) *ICO 2021. LNNS*, vol. 371, pp. 120–131. Springer, Cham (2022). [https://doi.org/10.1007/978-3-030-93247-3\\_13](https://doi.org/10.1007/978-3-030-93247-3_13)
22. Vasant, P., Munapo, E., Thomas, J.J., Weber, G.W. (eds.): *Artificial Intelligence in Industry 4.0 and 5G Technology*. Wiley (2022)
23. Nor, N.S.M., Malim, N.H.A.H., Rostam, N.A.P., Thomas, J.J., Effendy, M.A., Hassan, Z.: Automated classification of eight different Electroencephalogram (EEG) bands using hybrid of Fast Fourier Transform (FFT) with machine learning methods. *Neurosci. Res. Notes* **5**(1), 116 (2022)





# Machine Learning-Based Predictive Modelling of Spot-Welding Process Parameters

Dinesh V. Burande<sup>1</sup>, Kanak Kalita<sup>1</sup>(✉), and Jasgurpeet Singh Chohan<sup>2</sup>

<sup>1</sup> Department of Mechanical Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, India  
drkanakkalita@veltech.edu.in

<sup>2</sup> Department of Mechanical Engineering, University Centre for Research and Development, Chandigarh University, Mohali, India

**Abstract.** This research paper presents a comparative study of machine learning (ML) algorithms for the predictive modelling of spot-welding process parameters. The focus is on predicting the nugget diameter (ND) and nugget height (NH) using four different ML algorithms (namely, linear regression, random forest regression, adaptive boosting regression, and support vector machine regression). An experimental dataset consisting of fifty data points is used in this study. Exploratory data analysis was carried out to investigate the effect of process parameters on the response variables. Performance evaluation metrics were employed to gauge the predictive accuracy of the ML models. The results indicate that the random forest regression demonstrates superior performance in predicting both ND and NH compared to the other algorithms. This study can serve as a foundation for further research in optimizing the spot-welding process through machine learning techniques.

**Keywords:** Prediction · Welding · Regression · Process parameter · Joining · Machine learning

## 1 Introduction

Welding is a process of joining two or more metals together using pressure and heat. It requires generally three things—a heat source, a filler metal, and a shielding gas of flux. It is distinct from other processes used to join metals such as soldering, braising or using glue. In welding, supplying heat to the metals creates a molten pool into which a filler is being used. Welding is a permanent joining process. After welding the resulting product is no longer a distinctive piece but a single homogeneous unit.

The advancement in welding has led to the change from hand-held tools to industrial robots that perform the process more efficiently and rapidly [1]. Welding is widely used by different industries like automotive, aerospace, shipbuilding industries, construction sites etc. [2]. There are many welding types and different variations according to the different techniques being used like friction, electron beams and many more. Each technique is tailored for particular metals, and there are methods available for welding

dissimilar metals, such as copper and iron, as well. All the welding processes induce fusion through some energy source. Here the base metals are melted using the arc being induced upon it. Processes like SMAW (shielded metal arc welding) use an electrode that melts to induce fusion on the base metal and act as a filler metal for the joint [3]. TIG welding uses a tungsten electrode and inert gas to perform the welding process [4]. Similarly, the different welding is performing similar tasks using different techniques to join two materials. Providing a high-quality weld and high efficiency in welding has always been a focus point and challenging task for researchers [2].

Friction stir welding is widely used as it helps in joining two different metals and the main advantage is there are no toxic fumes emitted from the process. In FSW tool comprise a rotating round shoulder with a threaded pin being rotated to create friction upon it causing heat which helps in joining these two materials [5]. Resistance welding is commonly used in the automotive industry as mainly the chassis material weld is being done using resistance spot welding. Resistance spot welding is the process of joining two materials using pressure and heat. Heat is generated by the resistance to the flow of current through the electrode. Increasing the welding electrode force and weld time increases the strength of the spot weld specimen and being found that the weld nugget has the highest hardness value [6].

Various researchers have carried out numerous works on the parameters, process, defects, formation mechanism, welding technique, temperature and build strength [7]. Welding process parameters are categorized in mainly two different groups i.e., controlled variables and uncontrolled variables. The controlled variable includes voltage, current, welding speed, arc length, weld metal deposition etc. and uncontrolled variables comprise heat affected zone (HAZ), weld penetration, weld strength etc. These variables affect the welding process parameter and affect the quality characteristics [8]. Obtaining the best matching parameters of controlled and uncontrolled variables are not possible with the experimental iteration or optimization process. So different computing optimization techniques and machine learning algorithms are being used to reduce the overall cost and time and get the closure result towards the best parameters. Sudhagar et al. [7] have used the SVM classification technique for predicting the type of weld i.e., good weld or bad weld using the surface image and reached an accuracy of up to 95.8%. Sefene et al. [9] have worked on the process parameter optimization of friction stir welding on 6061 aluminium alloy using a machine learning algorithm. The ML algorithm used were random forest, decision tree and gradient boosting for predicting the best fit for examining the mechanical properties like tensile strength and weld strength. It was found that random forest prediction got yielding of highest  $R^2$  value as 0.92, found to be better result in comparison to other two algorithms. Liang et al. have worked upon the weld joint penetration from weld pool surface prediction using the machine learning-based regression model. A comparison was done between Support vector regression (SVR) and multi-layer perceptron (MLP) to find out the best-fit result between the algorithms. The author found out that SVR has a better prediction than MLP for the back-side bead width even with the smaller training data [10]. Satpathy et al. [11] have carried out a study on the experimental as well as machine learning prediction technique to get the joint strength of aluminium-copper dissimilar metal using ultrasonic spot welding.

Based on the above literature study it is observed that there are very few papers on application and comparison of multiple ML models for spot welding. Thus, the current paper makes the following contribution

- It provides a comprehensive comparison of four machine-learning algorithms for the predictive modelling of spot-welding process parameters.
- This study investigates and offers insights into the effect of various process parameters on the response variables using exploratory data analysis.
- The performance evaluation metrics allow an objective assessment of the predictive accuracy of the models, which makes it easier to identify the most effective algorithm.

## 2 Machine Learning Algorithms

### 2.1 Linear Regression

Linear regression, a supervised learning approach, offers a statistical means for assessing the connection between variables. This predictive method establishes relationships between dependent and independent variables. When a model has multiple input variables, it is known as a multivariate regression model. The primary goal of the regression model is to determine a statistically significant association between two variables. Simple linear regression can be expressed as

$$y = c + mx \quad (1)$$

$c$  is the  $y$  intercept and  $m$  is the slope.

However, for  $n$  number of predictors ( $x_1, x_2, \dots, x_n$ ), a regression can be expressed as

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_n \quad (2)$$

### 2.2 Random Forest Regression

Random forest is an ensemble learning method. It's a classic method for regression and classification problems. It works on the base learning of decision trees. The problem with the decision tree is that they tend to have a high variance. That leads to the overfit of the training set and they are unable to generalize well. Random Forest attempts to correct that problem by ensembling many different trees. The steps are:

1. It creates a bootstrapped dataset with a subset of available variable.
2. It fits a decision tree based on that dataset.
3. Tree like any other provides the rules for predicting the outcome.
4. Repeat the process and tally up the votes for the individual trees predicting the outcome.
5. Finally, in the end, the outcome with the most votes is the random forest prediction.

$$\text{Prediction through random forest} = \frac{1}{K} \sum_{k=1}^K h_k(x) \quad (3)$$

$$\text{Mean Square error} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_{OOB})^2 \quad (4)$$

$$R^2_{OOB} = 1 - \text{MSE}_{OOB} / \text{Var}_y \quad (5)$$

### 2.3 Adaptive Boosting Regression

Adaptive Boosting Regression or AdaBoost regression is an ensemble regression technique that works based on weak learners. In this several weak learners are combined to form a strong learner. For every sample data set, weight is assigned to each observation and further, it is being used in hypothesis learning. Next, the false prediction is recognized and assigned to the next base learner with high weight. The process is repeated several times to reach the target.

### 2.4 SVM Regression

As we know Support vector regression is the part of support vector machine created by Vladimir Vapnik. It is being used for both regressions as well as classification. It is one of the most effective classifiers among those, which are sort of linear. The objective of the SVM algorithm is to find a hyperplane in an  $N$  ( $N$  is no. Feature or independent variables) dimensional space. The main idea behind support vector regression is to choose a margin line which can cover all data points and the boundary between the margin lines is the regression model. The upper margin line drawn above the prediction line is called the max positive error margin and similarly, the below margin line is the negative max error margin. The points covering inside the margin line are not considered in calculating the error but the points outside the upper and lower margin called vectors as a slack variable help in giving the percentage of error in the model. These slack variables' error is calculated by getting the deviation between the actual data point and the max error margin. The data falling on the margin or above the margin line as called as support vector helps in calculating the error. It also gives the flexibility of defining the level of error that can be acceptable in the model. SVR is quite similar to the linear regression model but much more sensitive to outliers in comparison to linear regression. The benefit of SVR is it can be incorporated with nonlinear problems also. It helps in preventing overfitting issues.

## 3 Predictive Model Performance Evaluation Metrics

The difference between  $i$ th actual response parameter ( $y_a^i$ ) and the predicted response parameter ( $y_p^i$ ) is termed residual ( $\varepsilon^i$ ).

$$\varepsilon^i = y_a^i - y_p^i \quad (6)$$

$R^2$  which is a measure of the goodness of fit i.e., the coefficient of determination is calculated as,

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_a^i - y_p^i)^2}{\sum_{i=1}^n (y_a^i - y_m)^2} \quad (7)$$

where,  $y_m$  is the mean value of the actual response parameters.

To negate the effect of the direction of errors, often their absolute values are considered. By finding the average of the absolute errors for all the samples in the  $n$  sample space, the mean absolute error, MAE is calculated as

$$\text{MAE} = \frac{\sum_{i=1}^n |y_a^i - y_p^i|}{n} \quad (8)$$

The squares of the errors are also often used to quantify errors. By finding the average of the squared errors for all the samples in the  $n$  sample space, the mean-squared-error, MSE, is calculated as

$$\text{MSE} = \frac{\sum_{i=1}^n (y_a^i - y_p^i)^2}{n} \quad (9)$$

Often despite having moderate or low MSE, the presence of outliers may result in abnormally high errors at certain sample points. This can be identified by using

$$\text{Max. Error}(y_a, y_p^i) = \max.(|y_a^i - y_p^i|) \quad (10)$$

The median error is resilient to the presence of outliers and may be calculated as,

$$\text{Med. Error}(y_a, y_p^i) = \text{median}(|y_a^1 - y_p^1|, \dots, |y_a^i - y_p^i|, \dots, |y_a^n - y_p^n|) \quad (11)$$

## 4 Results and Discussion

### 4.1 Case Study: Spot Welding

The experimental dataset is from the works of Pashazadeh et al. [12]. They used a neural network and multi-objective genetic algorithm in their work. In their experiment, AISI 1008 steel sheet material was used. The sheet thickness was 0.7 mm. The specimen was sliced into sheets and cleaned with acetone. Before starting the welding process, it dried up. The centring was done by marking a small circle to get the alignment accuracy of the electrode tip. Three process parameters namely welding pressure ( $P$ ), current ( $I$ ) and time ( $t$ ) were considered. The nugget diameter ( $ND$ ) and nugget height ( $NH$ ) were the output response. In total 50 distinct combinations of welding pressure ( $P$ ), current ( $I$ ) and time ( $t$ ) were used to conduct the experiments. A statistical summary of the experimental dataset is presented in Table 1.

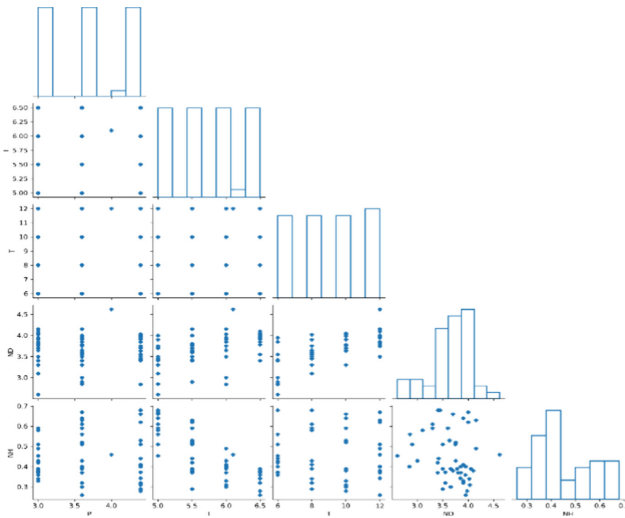
### 4.2 Effect of Process Parameters

In this study, initially, the effect of various process parameters on the responses is investigated by using exploratory data analysis. Figure 1 shows a pair plot of the various process parameters and the responses.

Figure 2 shows the correlation heat map of the various process parameters and the responses. It is observed that the process parameters do not have any significant

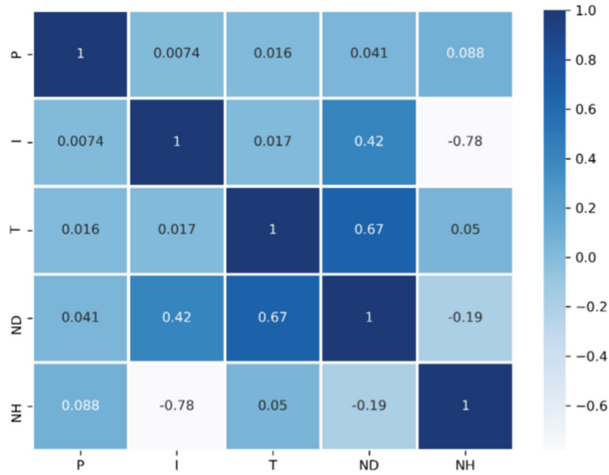
**Table 1.** Statistical features of the dataset

Feature	Welding pressure ( $P$ )	Welding current ( $I$ )	Welding time ( $t$ )	Nugget diameter ( $ND$ )	Nugget height ( $NH$ )
Count	50	50	50	50	50
Max	4.40	6.50	12.00	4.62	0.68
Min	3.00	5.00	6.00	2.60	0.26
Mean	3.69	5.77	9.12	3.67	0.45
SD	0.58	0.57	2.29	0.39	0.12
Quartile 1	3.00	5.50	8.00	3.46	0.37
Quartile 2	3.60	6.00	10.00	3.75	0.41
Quartile 3	4.40	6.40	12.00	3.94	0.53



**Fig. 1.** Pair plot of welding pressure ( $P$ ), current ( $I$ ) and time ( $t$ ) with nugget diameter ( $ND$ ) and nugget height ( $NH$ )

correlation among them. This indicates that the problem of multicollinearity does not exist for this dataset. The responses are also very weakly correlated to each other. The welding pressure ( $P$ ), current ( $I$ ) and time ( $t$ ) correlate 0.041, 0.42 and 0.67 with the nugget diameter ( $ND$ ) respectively. This indicates that the welding current ( $I$ ) and time ( $t$ ) have a moderate correlation with the nugget diameter ( $ND$ ). On the other hand, the welding pressure ( $P$ ), current ( $I$ ) and time ( $t$ ) correlate 0.088, -0.78 and 0.05 with the nugget height ( $NH$ ). This indicates that barring welding current ( $I$ ) the other two process parameters do not share an appreciable direct correlation with the nugget height ( $NH$ ).

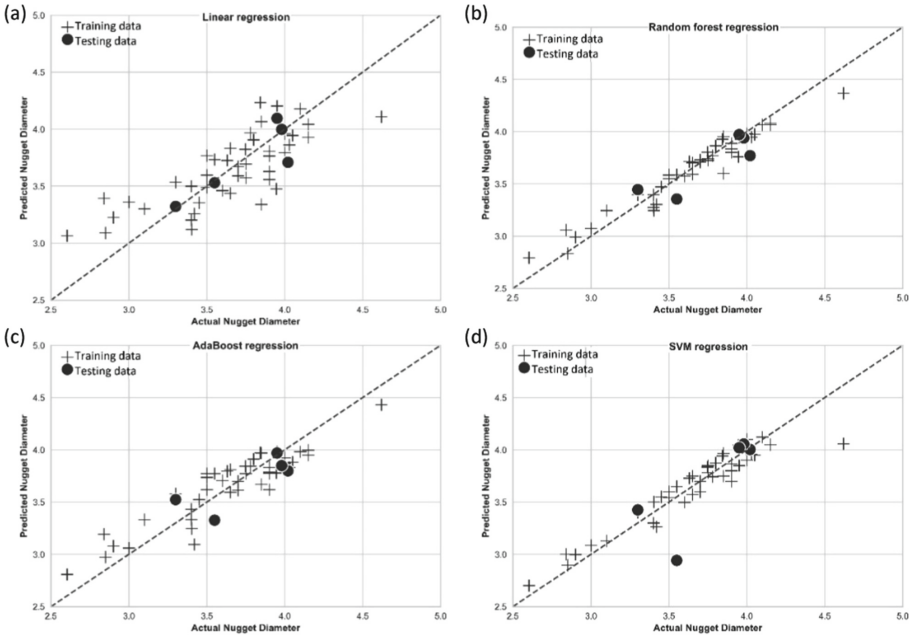


**Fig. 2.** Correlation heatmap of welding pressure ( $P$ ), current ( $I$ ) and time ( $t$ ) with nugget diameter ( $ND$ ) and nugget height ( $NH$ )

### 4.3 ML Modelling of Nugget Diameter ( $ND$ )

The dataset containing the 50 experimental data points is split into training and testing sets consisting of 45 and 5 samples respectively. This split is needed to ascertain that the developed ML models are accurate. Figure 3 shows the comparison of the various ML models in predicting the nugget diameter ( $ND$ ). The data points above the diagonal line mean that the ML model was overpredicted, whereas the data points below the diagonal line mean that the ML model was underpredicted. It is observed from Fig. 3a that linear regression can predict the nugget diameter ( $ND$ ) with reasonable accuracy. However, several data points are seen to be quite away from the diagonal line, indicating a high residue. In contrast, the random forest regression-based prediction of nugget diameter ( $ND$ ) is much more accurate. The AdaBoost regression predictions show that there is a good improvement over the linear regression but poorer than the random forest regression. The SVM regression shows a tight cluster of data points hugging the diagonal line. However, a couple of outliers data points in both the training and testing phase are observed.

To quantitatively gauge the prediction performance of the four ML models, the various statistical metrics are calculated for training and testing datasets. It is observed from Table 2 that in terms of  $R^2$ (training) the models can be ranked as RFR > SVR > ABR > LR whereas for  $R^2$ (testing), the ranking is RFR > LR > ABR > SVR. It should be noted that despite the overall high accuracy of SVM in training the presence of a single outlier in the testing set cost it dearly since the test data set is of small volume. Similarly, in terms of MSE, the ranking is RFR > SVR > ABR > LR for training and LR > RFR > ABR > SVR for the testing set.



**Fig. 3.** Actual versus predicted nugget diameter (*ND*) for various regressions (a) Linear (b) Random forest (c) AdaBoost (d) SVM

**Table 2.** Predictive power of various ML models for nugget diameter (*ND*)

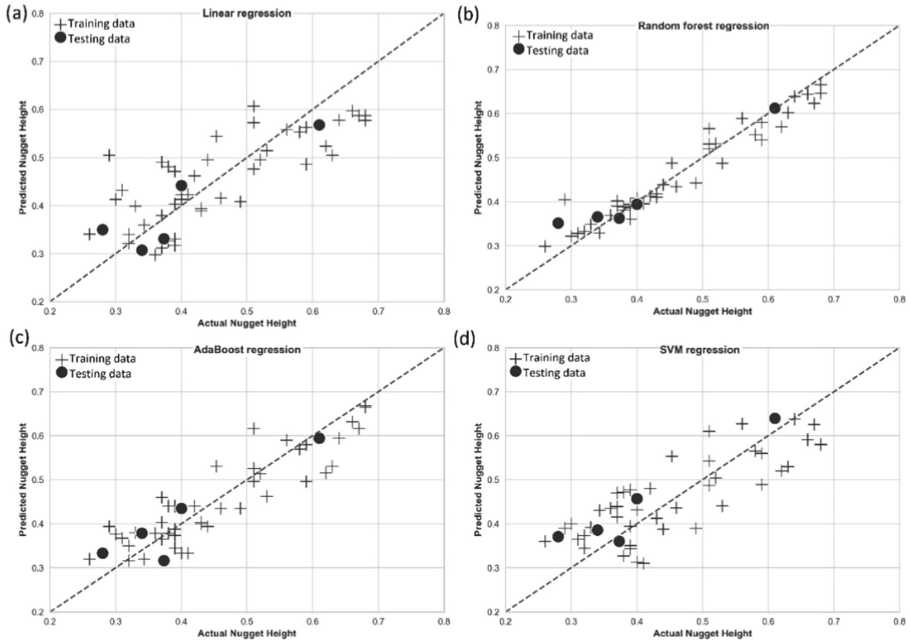
Metric	LR	RFR	ABR	SVR	LR	RFR	ABR	SVR
	Training				Testing			
$R^2$	0.6024	0.9354	0.8284	0.8966	0.7093	0.6985	0.5948	0.0307
EVS	0.6024	0.9356	0.8296	0.8968	0.7183	0.7461	0.6471	0.0904
MSE	0.0630	0.0102	0.0272	0.0164	0.0237	0.0246	0.0331	0.0791
MAE	0.2105	0.0792	0.1428	0.1009	0.1040	0.1301	0.1633	0.1791
Max. error	0.5547	0.2514	0.3533	0.5620	0.3089	0.2487	0.2250	0.6068
MSLE	0.0031	0.0005	0.0013	0.0007	0.0010	0.0011	0.0016	0.0044
MedAE	0.1866	0.0677	0.1267	0.0999	0.0248	0.1475	0.2200	0.0770

#### 4.4 ML Modelling of Nugget Height (*NH*)

The ML-based modelling of nugget height (*NH*) is shown in Fig. 4. The linear regression is seen to have scattered data points all around the diagonal (Fig. 4a), which indicates the presence of significant residues. However, the residue of the testing data points for the linear regression is smaller as compared to its training data points. In contrast, the random forest regression has much better prediction performance as observed in Fig. 4b.



The scatter of data points of the AdaBoost regression (Fig. 4c) is seen to be much improved over the linear regression. The predictive power of the ML models in terms of statistical metrics is presented in Table 3.



**Fig. 4.** Actual versus predicted nugget height ( $NH$ ) for various regressions (a) Linear (b) Random forest (c) AdaBoost (d) SVM

**Table 3.** Predictive power of various ML models for nugget height ( $NH$ )

Metric	LR	RFR	ABR	SVR	LR	RFR	ABR	SVR
	Training				Testing			
$R^2$	0.5925	0.9328	0.7966	0.6362	0.8200	0.9055	0.8569	0.7673
EVS	0.5925	0.9328	0.7968	0.6368	0.8201	0.9273	0.8664	0.9079
MSE	0.0056	0.0009	0.0028	0.0050	0.0023	0.0012	0.0018	0.0029
MAE	0.0620	0.0226	0.0431	0.0630	0.0458	0.0232	0.0398	0.0472
Max. error	0.2150	0.1150	0.1067	0.1005	0.0702	0.0717	0.0567	0.0907
MSLE	0.0027	0.0004	0.0013	0.0024	0.0012	0.0007	0.0010	0.0016
MedAE	0.0606	0.0173	0.0342	0.0646	0.0420	0.0109	0.0383	0.0460

## 5 Conclusions

This research paper provided an in-depth comparison of four machine learning algorithms for the predictive modelling of spot-welding process parameters. Through a comprehensive analysis of the performance metrics, it was determined that random forest regression outperformed linear regression, adaptive boosting regression, and support vector machine regression in predicting both nugget diameter and nugget height. This finding suggests that employing random forest regression for predicting spot-welding process parameters can lead to more accurate results. This will eventually improve the overall efficiency and safety of the spot-welding process. Future research can explore other machine learning algorithms, as well as deep learning techniques, to further enhance the predictive performance of the models.

## References

1. Rout, A., Deepak, B.B.V.L., Biswal, B.B.: Advances in weld seam tracking techniques for robotic welding: a review. *Robotics and Computer-Integrated Manufacturing* **56**, 12–37 (2019)
2. Zhao, D., Ivanov, M., Wang, Y., Du, W.: Welding quality evaluation of resistance spot welding based on a hybrid approach. *J. Intell. Manuf.* **32**, 1819–1832 (2021)
3. Arifin, A.: Dissimilar metal welding using shielded metal arc welding: a review. *Technology Reports of Kansai University* **62** (2020)
4. Tanaka, K., Shigeta, M., Komen, H., Tanaka, M.: Electrode contamination caused by metal vapour transport during tungsten inert gas welding. *Sci. Technol. Weld. Join.* **26**, 258–263 (2021)
5. Rai, R., De, A., Bhadeshia, H.K.D.H., DebRoy, T.: Friction stir welding tools. *Sci. Technol. Weld. Joining* **16**, 325–342 (2011)
6. Kahraman, N.: The influence of welding parameters on the joint strength of resistance spot-welded titanium sheets. *Mater. Des.* **28**, 420–427 (2007)
7. Sudhagar, S., Sakthivel, M., Ganeshkumar, P.: Monitoring of friction stir welding based on vision system coupled with machine learning algorithm. *Measurement* **144**, 135–143 (2019)
8. Kumar, B.P., Vijayakumar, Y.: Optimization of shielded metal arc welding parameters for welding of pipes by using Taguchi approach. *Int. J. Eng. Sci. Technol.* **4**, 2083–2088 (2012)
9. Mishra, A., Sefene, E.M., Tsegaw, A.A.: Process Parameter Optimization of Friction Stir Welding on 6061AA Using Supervised Machine Learning Regression-based Algorithms. arXiv preprint [arXiv:2109.00570](https://arxiv.org/abs/2109.00570) (2021)
10. Liang, R., Yu, R., Luo, Y., Zhang, Y.: Machine learning of weld joint penetration from weld pool surface using support vector regression. *J. Manuf. Process.* **41**, 23–28 (2019)
11. Satpathy, M.P., Mishra, S.B., Sahoo, S.K.: Ultrasonic spot welding of aluminum-copper dissimilar metals: a study on joint strength by experimentation and machine learning techniques. *J. Manuf. Process.* **33**, 96–110 (2018)
12. Pashazadeh, H., Gheisari, Y., Hamed, M.: Statistical modeling and optimization of resistance spot welding process parameters using neural networks and multi-objective genetic algorithm. *J. Intell. Manuf.* **27**, 549–559 (2016)

# Author Index

## A

- Akash, Fardin Rahman 189  
Al Munem, Abdullah 231  
Alam, Farhana 127, 189  
Andre, 315  
Anem, Mohammed Sharafullah 217  
Anilkumar, Abhishek 294  
Arefin, Mohammad Shamsul 113, 127, 153,  
177, 189, 217  
Asif, Md. Asif Uzzaman 153

## B

- Banmongkol, Channarong 284  
Bansal, Aarti 209  
Bebeshko, B. 104  
Bektaş, Gebrail 243  
Bhagat, Rahul 167  
Boojhawon, Ravindra 273  
Borzin, R. 17  
Boucher, Sasha 199  
Burande, Dinesh V. 337  
Burynin, D. 56

## C

- Chanza, Martin 199  
Chaudhary, M. P. 66  
Chohan, Jasgurpeet Singh 337  
Chowdhury, Tasnia Afrin 127

## D

- da Costa Silva de Paula, Elizete 83  
Dabeedoal, Janvee 273  
Daus, Yu 47  
De Koker, Louise 199  
de Luna, Robert G. 304  
Deb, Kaushik 252

## F

- Faiyaz, Nafis 217  
Faruki, Tasnim 113  
Formoso, Leonardo Pitança 3

## G

- Girish, G. P. 167  
Grishin, A. 140  
Grishin, V. 140  
Gukhool, Oomesh 273

## H

- Haider, Sakib 177  
Haque, Rafid Mahmud 231  
Hareesh, V. 73, 294  
Harjoko, Agus 36  
Hassan, Mahmud 217  
Hossain, Md Jakir 189  
Hossain, Md. Sagor 153

## I

- Imam, Omar Tawhid 127  
Immanuel, Novendra 315

## J

- Jaya, Daniel 315  
Joshua Thomas, J. 325  
Juliano, Suzana Franco 83

## K

- Kalita, Kanak 337  
Kamil, Dea Angelia 36  
Karthik, Vudhya Muni Sai 73  
Khai Sian, Lim 325  
Khan, Mohammad Ashik Iqbal 231  
Khan, Mozammel H. A. 231  
Kochetkov, V. 17  
Kolesenkov, A. 47  
Kostin, V. 17  
Kostrov, B. 47  
Kumar, Vijay 209  
Kumari, Nanda 284

## L

- Lakhno, V. 104  
Lebed, N. 263

**M**

Mabud, Malyha Bintha 113  
 Mabuza, Gugulethu 199  
 Madhu, M. Nimal 73  
 Maha, Lamyca Tasneem 231  
 Malyukov, V. 104  
 Malyukova, I. 104  
 Manshahia, Mukhdeep Singh 66  
 Mohylnyi, H. 104  
 Munapo, Elias 199

**N**

Namburu, Sai Sylesh Gupta 73  
 Naresh, M. 95  
 Nigdeli, Sinan Melih 243  
 Nimal Madhu, M. 294

**O**

Oliveira, Selma Regina Martins 3, 83

**P**

Panchenko, V. 47, 56, 140  
 Patel, Amrutbhai N. 27  
 Patel, Priyankaben K. 27  
 Peddakrishna, Samineni 95  
 Perdio, Lejan Daniel I. 304  
 Pinho, Gabriel Souza 3  
 Preeti, S. H. 167  
 Premjith, B. 294  
 Promy, Fahmida Nusrat 127  
 Proshkin, Y. 56

**R**

Rahman, Md. Ashiqur 252  
 Rahman, S. M. Arafat 177  
 Reza, Ahmed Wasif 113, 127, 153, 177,  
 189, 217  
 Reza, Nahid 177  
 Ripa, Tanjina Akter 217  
 Rosales, Marife A. 304

**S**

Safa, Noor Fabi Shah 231  
 Sahputra, Iwan Halim 315  
 Saiful, Md. 177  
 Salam Aonty, Shuhena 252  
 Semenova, N. 140  
 Shrestha, Deepanjali 273  
 Silaev, A. 17  
 Singh, Sweta 167  
 Singla, Sandeep 209  
 Smirnov, A. 56  
 Smirnov, O. 104  
 Sokolova, N. 17  
 Sumona, Rehnuma Naher 217

**T**

Tamim, Hasnath Ahmed 153  
 Tasnim, Rafa 113  
 Thirupathi, M. 95  
 Tokarev, K. 263  
 Tuhin, Rashedul Amin 113

**U**

Uddin, Borhan 153

**V**

Vasudevan, Nandu 73

**W**

Wahyono, 36  
 Wedianto, Feren Nathan 315

**Y**

Yadav, Ramsagar 66  
 Yadukrishnan, V. 294  
 Yücel, Melda 243  
 Yudaev, I. 17, 263

**Z**

Zaman, Sarah Samiha 189