

**A COMPARATIVE ANALYSIS OF AMERICAN SIGN LANGUAGE RECOGNITION
BASED ON CNN, KNN AND RANDOM FOREST CLASSIFIER**

BY

**TUSHAR KUMAR MAHATA
ID: 161-15-7252**

This Report Presented in Partial Fulfillment of the Requirements for the Degree of
Bachelor of Science in Computer Science and Engineering

Supervised By

MS. FARAH SHARMIN

Senior Lecturer
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JUNE 2021

APPROVAL

This Project/internship titled “A Comparative Analysis Of American Sign Language Recognition Based On CNN, KNN And Random Forest Classifier”, submitted by Tushar Kumar Mahata, ID No: 161-15-7252 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on June 1, 2021.

BOARD OF EXAMINERS

Chairman



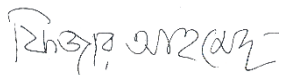
Dr. Touhid Bhuiyan

Professor and Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



Internal Examiner

Dr. Fizar Ahmed

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology



Internal Examiner

Md. Azizul Hakim

Senior Lecturer

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



External Examiner

Dr. Mohammad Shorif Uddin

Professor

Department of Computer Science and Engineering

Jahangirnagar University

DECLARATION

I hereby declare that, this project has been done by me under the supervision of **Ms. Farah Sharmin, Senior Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Farah Sharmin

Ms. Farah Sharmin
Senior Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Aniruddha Rakshit

Aniruddha Rakshit
Senior Lecturer
Department of CSE
Daffodil International University

Submitted by:

Tushar

Tushar Kumar Mahata
ID: 161-15-7252
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

Almighty God for everything. For divine blessing makes it possible for me to complete the final year project successfully I express my heartiest gratefulness and thanks to almighty God.

I wish my profound indebtedness I am very pleased to my honorable supervisor **Ms. Farah Sharmin**, Senior Lecturer, Department of CSE Daffodil International University, Dhaka. Keen interest & Deep Knowledge of my supervisor in the field of “*NLP*” to carry out this project. I could not be able to effectively complete my project without her endurance and inspiration. Her obtainment of solving sudden problem and rigorous mind unto support have aided me inventing my way to complete this project. To have her as my supervisor I am very pleased and lucky.

I would like to express our heartiest gratitude to my co-supervisor **Aniruddha Rakshit** (Senior Lecturer), and Head, Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

Finally, I must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

There is a body language based communication way for deaf and mute people called sign language. But it is very difficult to understand sign language for a normal person. For that reason, mute and deaf people are unable to express what is in their mind. There are many sign languages all over the world such as ASL, BSL etc. A part of American Sign Language(ASL) represents English alphabets by hand gesture and our dataset based on this part. This paper represents a Comparative analysis of American Sign Language Recognition. Where system uses a vast ASL dataset from the MNIST database applied three algorithms as CNN, KNN and Random Forest to create the system and analyze comparatively. This analysis is based on execution time and accuracy rate. Where CNN produces the most efficient result. It's recognized with good generalization ability. Further research and more can produce an ASL based sign language interpreter system. Then it will be the easier way to communicate with deaf and mute people.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	i
Declaration	iii
Acknowledgements	iv
Abstract	v
CHAPTER	
CHAPTER 1: INTRODUCTION	1-3
1.1 Introduction	1
1.2 Motivation	1
1.3 Rationale of the Study	2
1.4 Research Questions	2
1.5 Expected Output	2
1.6 Report Layout	2
CHAPTER 2: BACKGROUND	4-5
2.1 Preliminaries	4
2.2 Related Works	4
2.3 Comparative Analysis and Summary	4
2.4 Scope of the Problem	5
2.5 Challenges	5
CHAPTER 3: RESEARCH METHODOLOGY	6-11

3.1 Research Subject and Instrumentation	6
3.2 Dataset Utilized	6
3.3 Proposed Methodology	8
CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	12-14
4.1 Experimental Results & Analysis	12
4.2 Discussion	12
CHAPTER 5: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH	15
5.1 Summary of the Study	15
5.2 Conclusions	15
5.3 Implication for Further Study	15
REFERENCES	16

LIST OF TABLES

TABLES	PAGE NO
Table 4.1.1: The comparison of experimental results.	12

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.2.1: Sample of dataset	6
Figure 3.2.2: Bar chart of train samples.	7
Figure 3.2.3: Bar chart of test samples.	7
Figure 3.3.1: System design	8
Figure 3.3.1.1: figure of CNN model structure.	9
Figure 3.3.2.1: figure of RF model structure.	10
Figure 3.3.3.1: figure of KNN model structure.	10
Figure 4.2.1: figure of confusion matrix.	13
Figure 4.2.2: Training accuracy vs validation accuracy.	13
Figure 4.2.3: Training loss vs validation loss	14

CHAPTER 1

Introduction

1.1 Introduction

A normal person communicates with other's using a language such as Bangla, English etc. But there are some peoples who are physically disabled to hearing or create own voice. They try to present something by their body languages. This way of communication is called sign language. There is 200+ sign languages used all over the world [1]. So, it is very tough to understand sign language for a simple person. On the other hand, the deaf and mute people can't communicate easily with normal people. Deaf and mute people cannot represent their thinking to all, and normal people also can't read them. But they can think like every normal person, and can also generate new ideas. without proper communication our society can't reach those ideas. Again, a deaf or mute people cannot lead a normal life as us. This model is a way to properly communicate. Then society will be more beneficial.

To make it easy several researches focus on sign language translation issues and here is also focused. Where the main perspective is to find the most efficient algorithm on a vast sign language dataset.

1.2 Motivation

- There are almost 6% (466 million) people who are deaf & mute [2].
- Currently 70% deaf & mute people are unemployed.
- Every 1 in 4 deaf & mute people face discrimination in their job.
- A lot of deaf & mute people don't get proper education as well as knowledge of proper sign language.
- By 2050, the number of deaf & mute will be about 900 million [2].
- Most of them can't communicate properly with ordinary people.

1.3 Rationale of the Study

To lead a good life needs proper communication. Needs a representation way of what is in mind. In most of the cases deaf and mute people are unable. They face discrimination. They are jobless and also are facing many problems in their personal life. Day by day the population of deaf and mute people is increasing. Where needs a proper solution. Creating a sign language interpreter could be a solution. Choosing perfect algorithm is the main organ for this kind interpreter. For that reason, this report focused on this topic.

1.4 Research Questions

- Have there been any finest ways to communicate with deaf and mute people?
- Will the proposed solution meet the needs with proper efficiency?
- Which algorithm's performance is the best followed by accuracy and executing time for MNIST American sign language dataset?
- How should those algorithms be applied at this dataset?

1.5 Expected Outcome

- Most efficient and easy way to communicate with mute and deaf people.
- Where proposed model's accuracy rate will be more than 90%.
- Proposed model will produce fastest executing time.

1.6 Report Layout

This report is presented in five chapters explaining "A Comparative analysis of American Sign Language Recognition based on CNN, KNN and Random Forest Classifier".

- Chapter 1: Introduction

Discussion about motivation, rationale of the study, research questions and expected output.

- Chapter 2: Background

Discussion about preliminaries, related works, comparative analysis and summary, scope of the problem and challenges.

- Chapter 3: Research Methodology

Discussion about research subject and instrumentation, dataset utilized and proposed methodology.

- Chapter 4: Experimental Results and Discussion

Discussion about experimental results & analysis and discussion.

- Chapter 5: Summary, Conclusion, Recommendation and Implication for Future Research

Discussion about summary of the Study, conclusions, implication for Further Study.

CHAPTER 2

Background

2.1 Preliminaries

The recognition of sign language has begun to appear at the end of 1990[3]. So, it's not a new computer vision problem. The primary system to be recognized was glove-based. Where used some electrochemical devices. Nowadays researchers focus on creating applications for mobile, computer etc. by using classifications as linear classifiers, neural networks, bayesian networks etc.

2.2 Related Works

Some researchers used a 3D capture element with motion tracking globes[4] or Microsoft koinet to ASL letter recognition accuracy rate is around 90%.

A ASL recognition system[5] to be classified with a HMM where this system presents with a dataset 30 words. 10.91% error rate is achieved on the database RWTH-BOSTON-50.

On different sign language datasets, [6,7,3] researchers applied a single algorithm like KNN, CNN, NNE where they got 93-97% accuracy.

According to [8] different types of algorithms are applied to a sign language dataset which is very related research, where SVM, KNN, RF, K-Means algorithms are applied.

2.3 Comparative Analysis and Summary

Normally a glove-based system for Sign Language Recognition (SLR) determines hand gesture by using parameters like hand's movement, position, angle, location etc. Where the main problem is with accuracy and efficiency. On the other hand, their dataset is not too rich. It is one of the reasons for loss. At [8], this research represents that SVM algorithm is best from others. where it's accuracy is 84%. Basically many researchers are applying different kinds of algorithms to sign language dataset. where the maximum dataset is not too rich. Another thing is which algorithm can take the solution to the next level?

2.4 Scope of the Problem

Since the end of 1990, Sign Language Recognition (SLR) has started to appear followed by different ways and datasets. Though American Sign Language (ASL)'s dataset is based on North American sign language, it is very popular all over the world. This letter gestures only a part of 200 years old ASL practice by The National Institute on Deafness and other Communications Disorders (NIDCD). This is a vast area to research in different ways. The final goal is to create an easier and efficient communication way for deaf and mute people.

2.5 Challenges

- Which sign language is most acceptable to people within 200+ languages?
- Which dataset is richer than others?
- Gesture against different users and backgrounds.
- Working with different algorithms for the same dataset is quite problematic.

CHAPTER 3

Research Methodology

3.1 Research Subject and Instrumentation

Because I am optimizing the algorithm of sign language conversion, I took the test and train dataset from MNIST[9]. Nowadays cloud computing can come handy when developing or optimizing a new ML algorithm. There are many free cloud computing services and I used google colab as it provides 30 minute free GPU usage. Which helps me a lot minimizing the data preprocessing and model training. So a google standard accelerated environment comes with 20 physical cores, variable ram and 2880 CUDA cores.

3.2 Dataset Utilized

To recognize sign language, is used a subset dataset from MNIST database. This dataset represents only a part of ASL which is alphabetical gesture. The dataset contains 27,455 samples as train data and 7172 samples as test data. This image data represents the original hand gesture of multiple users of different ages with different backgrounds. Where a single sample represents a single gesture image with shape of 28*28 pixels and grayscale value 0-255. For each alphabetic letter A-Z represents a label 0-25 as A=0, B=1, C=2....., where there is no case for J=9 or Z=25 because of gesture motion.



Figure 3.2.1: Sample of dataset

At this train dataset, there are approximately 900-1300 image data samples for each English alphabet and for test dataset, around 150-500 samples for each alphabet contains shape of every input image 28*28 with 1 color channel. This system has predicted multiple gestures (A,B,C,D.....Z) as multiclass predictions.

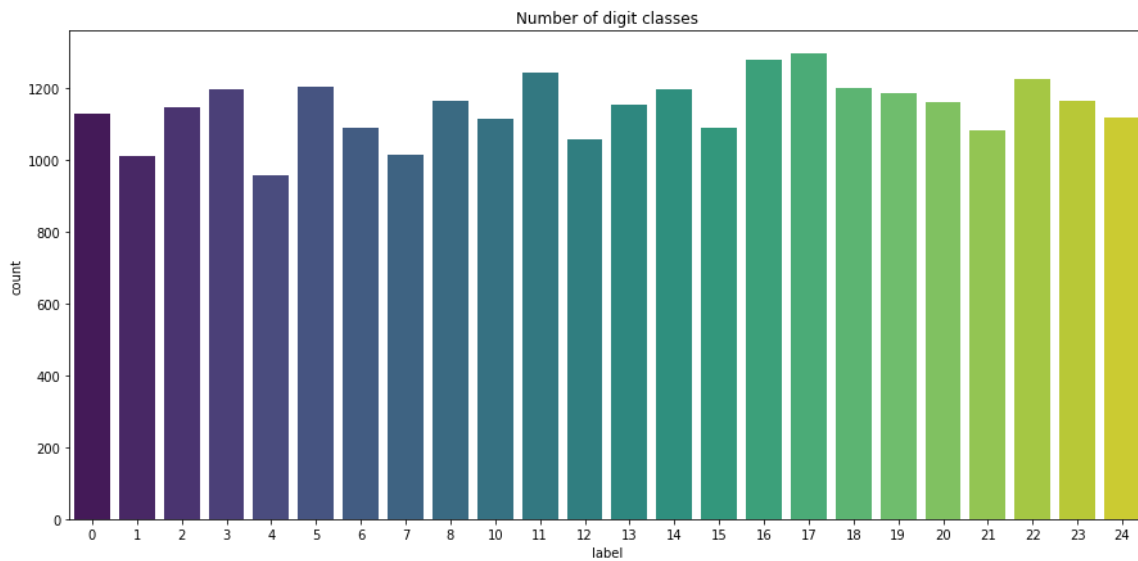


Figure 3.2.2: Bar chart of train samples.

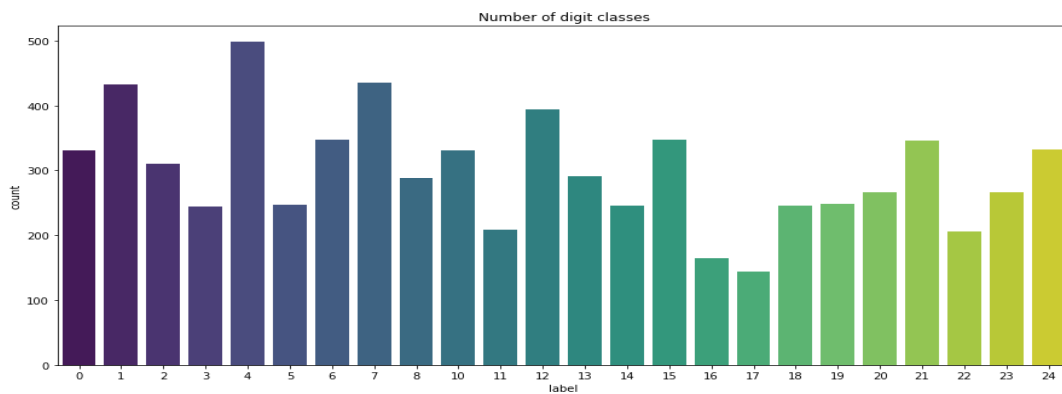


Figure 3.2.3: Bar chart of test samples.

3.3 Proposed Methodology

At American sign language dataset from MNIST, applied three different algorithms CNN, KNN and Random Forest. Analysed comparatively based on executing time and best accuracy score of test samples for each algorithm. After analysis, proposed the best possible algorithm for the dataset. As the figure below-

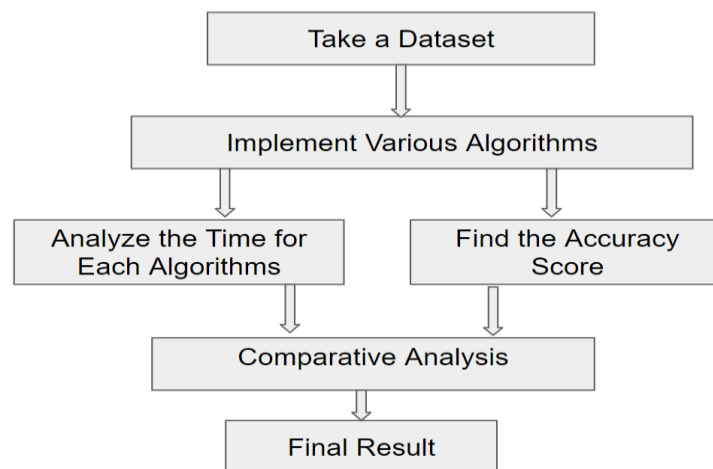


Figure 3.3.1: System design

3.3.1 Convolutional Neural Network (CNN)

To refer to the network or circuit of biological neurons was used the terms of neural network. To do image recognition and classification in neural networks one of the main categories is CNN. This model take input image data based on resolution of images and see as $h*w*d$ where, h = height, w = width and d = dimension. Normally a CNN model has two parts, Feature learning and Classification. As the structure below-

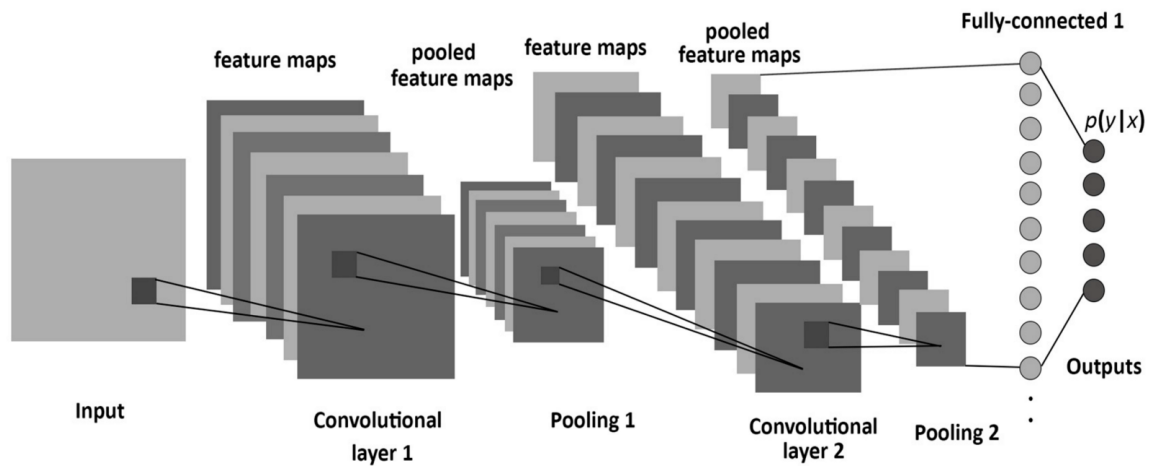


Figure 3.3.1.1: figure of CNN model structure.

The first layer of the applied model is convolutional layer 1 to extract features from an input image data by using filter. where image matrix 28×28 and filter matrix 5×5 . For next convolutional layer, convolutional layer 2 mechanism with map size 24×24 matrix and filter matrix 3×3 . Pooling layer has played an important role in reducing the number of parameters of large images. here applied max-pooling (2×2 max-pool) to reduce it's dimensionality, drawscale an input representation. Then the final outcome from feature learning be flatten input of classification where the classification activation function is softmax to classify the output.

3.3.2 Random Forest (RF)

RF classifier is an ensemble method that trains multiple decision trees in parallel with bootstrapping followed by aggregation. Bootstrapping indicates that several individual decision trees are trained in parallel on various subsets of the training dataset using different subsets of available features. Bootstrapping ensures that each individual decision

tree in the random forest is unique, which reduces the overall variance of the RF classifier. For the final decision, RF classifier aggregates the decisions of individual trees; consequently, RF classifier exhibits good generalization.

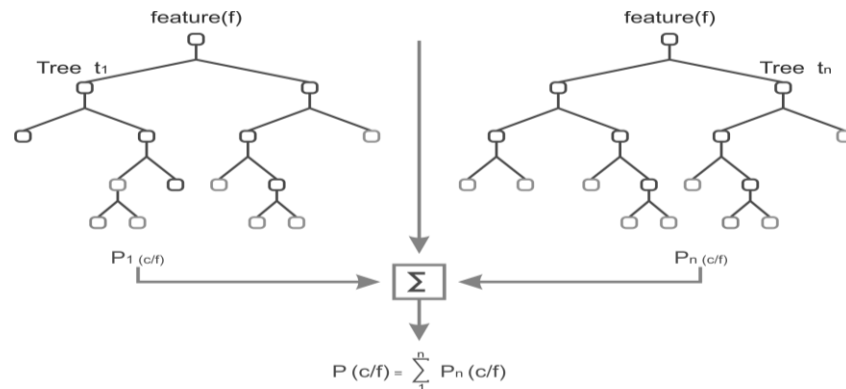


Figure 3.3.2.1: figure of RF model structure.

RF classifier outperforms most other classification methods at accuracy. Unlike DT classifier, RF classifier doesn't need feature scaling. RF classifier is more robust to the selection of training samples and noise in training dataset. RF classifier is hard to interpret but easy to tune the hyperparameter.

3.3.3 K-Nearest Neighbor (KNN)

Using 'feature similarity' to predict the values in its own way for the K-nearest neighbors (KNN) algorithm. Structure

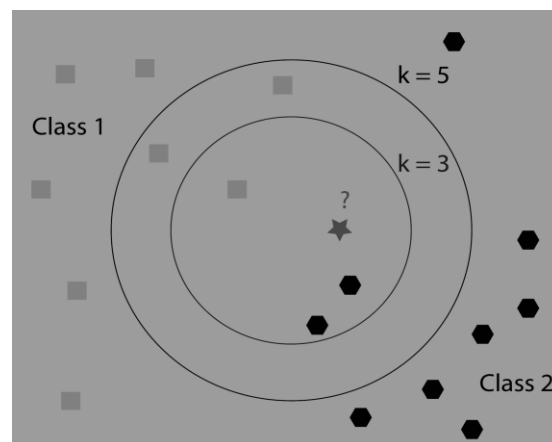


Figure 3.3.3.1: figure of KNN model structure.

In the applied model in this dataset it has 25 different classes. This model working procedure is- implementing the dataset needs to choose the nearest data point value of K. Here, $K = 5$ is used for this model. then using a method called Euclidean to find out the distance between each row of training samples and test sample. Next, sort them in an ascending order based on distance value and choose the top K rows based on the most frequent class of those rows. It will implement a class to the test point.

CHAPTER 4

Experimental Results and Discussion

4.1 Experimental Results & Analysis

Table 4.1.1: The comparison of experimental results.

Algorithms	Accuracy	Executing Time
CNN	96.82%	2.518 s
KNN	82.29%	559.6 s
RF	80.73%	0.604 s

The above table represents that CNN algorithm gives the highest accuracy rate 96.82% for testing samples. Where in KNN the accuracy rate is 82.29% and for RF accuracy rate is 80.73%. On the other hand, The lowest executing time for testing samples is gotten from the RF algorithm which is 0.604 seconds. Where 2.518 seconds for CNN and 559.6 seconds for KNN algorithm.

4.2 Discussion

From the analysis I proposed that CNN is better then KNN and RF for this dataset. Though CNN's executing time is more than RF, it represents an excellent accuracy rate.

This model trained and tested with MNIST dataset where it produces a good accuracy rate For Convolutional Neural Network. But it has given false positive results for some gestures, 6(G) predicted 19 times as 4(E) again 20 times given false positive for 24(Y). Here is the confusion matrix of prediction and truth labels-

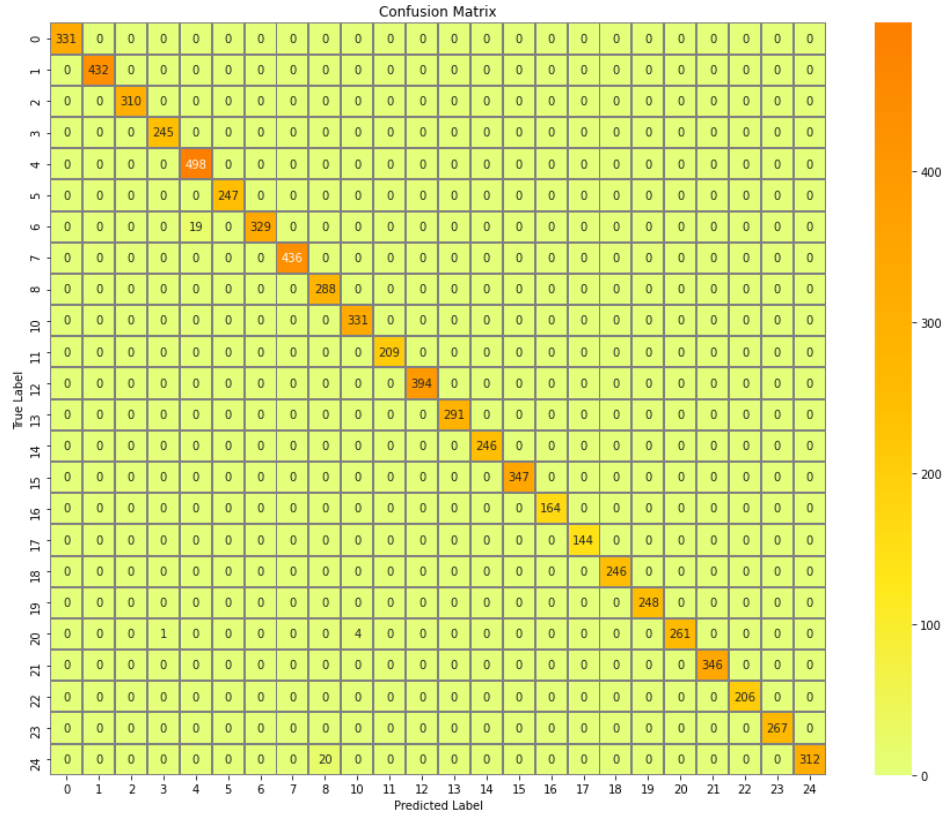


Figure 4.2.1: figure of confusion matrix.

This CNN model has been fitted with 10 epoches. Here, accuracy of train and validation has increased from every epoch. After 10 epoches the train accuracy is 83.53% and validation accuracy is 96.82%. As the figure below shows-

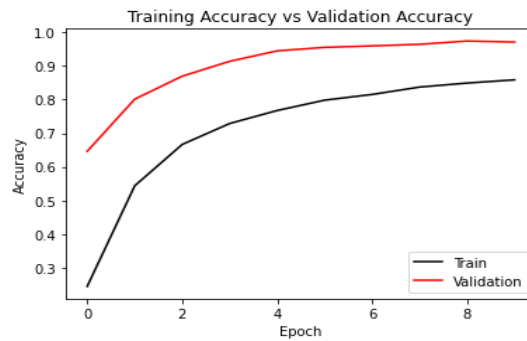


Figure 4.2.2: Training accuracy vs validation accuracy.

On the other hand, After 10 epoches the train loss is 0.4786 and validation loss is 0.1359.
As the figure below shows-

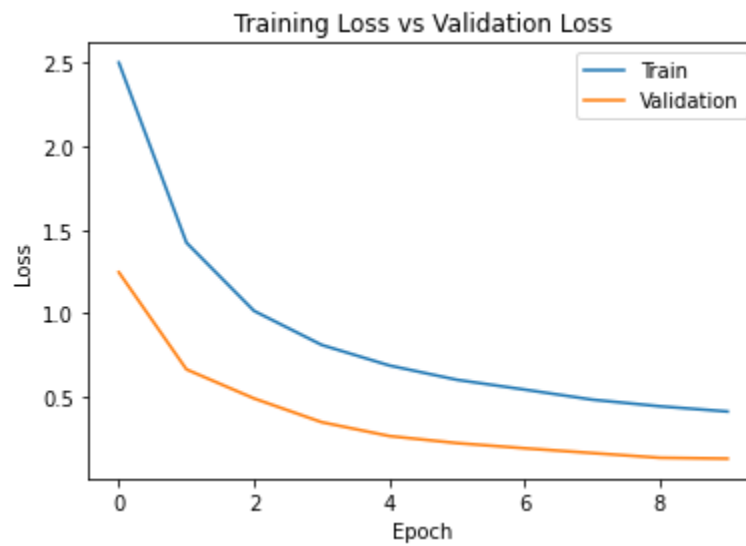


Figure 4.2.3: Training loss vs validation loss.

This model has been tested by a test dataset (7172 image data samples) where overall accuracy of the model is 96.82%.

CHAPTER 5

Summary, Conclusion, Recommendation and Implication for Future Research

5.1 Summary of the Study

Deaf and mute people are disable to hear and talk. So, the way to communicate with them is body language or sign language. But there are many sign languages all over the world. So, it is very problematic to recognize a simple person. It creates an area for researchers to research. In this age people are very comfortable with smart phones and computers. It is a matter of efficiency. For this perspective, here is a model developed with deep learning. For maximum positive output, it used a vast dataset and applied the most powerful classification CNN with 10 epochs. As a result it produces 96.82% accuracy. KNN and Random Forest also developed for this perspective, where KNN produces 82.29% accuracy and Random Forest produces 80.73% accuracy.

5.2 Conclusions

This is the age of digitalization depending on computers and smartphones. So, creating a recognizer of sign language is very user friendly. From this thinking applied different algorithms to the dataset. Where CNN produces a good accuracy. Though the validation loss rate is 0.1359. where some of gestures conflict with others. After comparative analysis, it is proposed that CNN is better then KNN and Random Forest for MNIST sign language dataset.

5.3 Implication for Further Study

- Dataset should be more rich by custom creating.
- Different sign languages (ASL, BSL, JSL etc) recognizer based models with vast datasets.
- Combine those models into one system and make it a real time interpreter.
- Implementing this system into computer or mobile applications.

References:

- [1] Sign language - Wikipedia, available at <<https://en.wikipedia.org/wiki/Sign_language/>>, last accessed on 28-05-2021 at 12:20 PM.
- [2] WHO | Estimates, available at <<<https://www.who.int/deafness/estimates/en/>>>, last accessed on 27-05-2021 at 05:20 PM.
- [3] Bikash Chandra Karmokar, Kazi Md. Rokibul Alam* , and Md. Kibria Siddiquee, “Bangladeshi Sign Language Recognition Employing Neural Network Ensemble,”
- [4] H. Brashear, K.-H. Park, S. Lee, V. Henderson, H. Hamilton, and T. Starner. “American Sign Language Recognition in Game Development for Deaf Children” Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility, Portland, Oregon, 2006
- [5] Thad Starner and Alex Pentland, “Real-Time American Sign Language Recognition from Video Using Hidden Markov Models,” AAAI Technical Report FS-96-05.
- [6] Thongpan Pariwat , Pusadee Seresangtakul, “Thai Finger-Spelling Sign Language Recognition Employing PHOG and Local Features with KNN,” Int. J. Advance Soft Compu. Appl, Vol. 11, No. 3, November 2019.
- [7] N.Priyadharsini , N.Rajeswari, “Sign Language Recognition Using Convolutional Neural Networks,”International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 5 Issue: 6, June 2017.
- [8] Proddutur Shruthi, Dr. Jitendra Jaiswal, “A Comparative Analysis On Sign Language Prediction Using Machine Learning Algorithms,” JETIR, Volume 7, Issue 6, June 2020.
- [9] Sign Language MNIST | Kaggle , available at <<<https://www.kaggle.com/datamunge/sign-language-mnist/>>>, last accessed on 26-05-2021 at 09:20 AM.

A COMPARATIVE ANALYSIS OF AMERICAN SIGN LANGUAGE RECOGNITION BASED ON CNN, KNN AND RANDOM FOREST CLASSIFIER

ORIGINALITY REPORT

14%

SIMILARITY INDEX

11%

INTERNET SOURCES

7%

PUBLICATIONS

10%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	4%
2	Siddharth Misra, Hao Li. "Noninvasive fracture characterization based on the classification of sonic wave travel times", Elsevier BV, 2020 Publication	3%
3	dokumen.pub Internet Source	2%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
5	Tilottama Goswami, Shashidhar Reddy Javaji. "Chapter 6 CNN Model for American Sign Language Recognition", Springer Science and Business Media LLC, 2021 Publication	1%
6	www.javatpoint.com Internet Source	1%
7	Submitted to University College London	

8

www.bartleby.com

Internet Source

1 %

9

Md. Zahid Hasan, Md. Sazzadur Ahamed, Aniruddha Rakshit, K. M. Zubair Hasan. "Recognition of Jute Diseases by Leaf Image Classification using Convolutional Neural Network", 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019

Publication

<1 %

10

Graur, Dan. "Molecular and Genome Evolution", Oxford University Press

Publication

<1 %

11

digitalcommons.calpoly.edu

Internet Source

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off