

**DCNN-Based Vegetable Image Classification Using Transfer Learning: A Comparative
Study**

BY

**M. Israk Ahmed
ID: 171-15-8919**

AND

**Md. Shahriyar Mahmud Mamun
ID: 171-15-8600**

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

Supervised By

Mr. Asif Uz Zaman Asif
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Ms. Fahmida Afrin
Lecturer
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

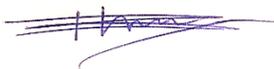
DHAKA, BANGLADESH

JUNE 2021

APPROVAL

This Project titled “**DCNN-Based Vegetable Image Classification Using Transfer Learning: A Comparative Study**”, submitted by M. Israk Ahmed (ID: 171-15-8919) and Shahriyar Mahmud Mamun (ID: 171-15-8600) 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 03rd of June, 2021.

BOARD OF EXAMINERS



Dr. Touhid Bhuiyan
Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman



Gazi Zahirul Islam
Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Raja Tariqul Hasan Tusher
Senior Lecturer

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner



Dr. Dewan Md. Farid

Associate Professor

Department of Computer Science and Engineering

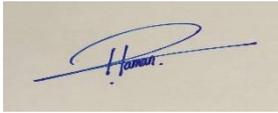
United International University

External Examiner

DECLARATION

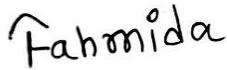
We hereby declare that, this project has been done by us under the supervision of **Mr. Asif Uz Zaman Asif, Lecturer, Department of CSE, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



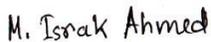
Mr. Asif Uz Zaman Asif
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:



Ms. Fahmida Afrin
Lecturer
Department of CSE
Daffodil International University

Submitted by:



M. Israk Ahmed
ID: 171-15-8919
Department of CSE
Daffodil International University



Md. Shahriyar Mahmud Mamun
ID: 171-15-8600
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to Almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Mr. Asif Uz Zaman Asif, Lecturer**, Department of CSE, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Deep learning” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr. Touhid Bhuiyan, Professor and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

We would also like to give thanks to the people who helped us regarding data collection.

Finally, we must acknowledge with due respect the constant support and patients of our parents, family members, relatives, and well-wishers.

ABSTRACT

In this paper, an attempt is addressed towards accurate vegetable image classification. A dataset consisting of 21,000 images of 15 classes is used for this classification. Convolutional neural network, a deep learning algorithm is the most efficient tool in the machine learning field for classification problems. But CNN requires large datasets so that it performs well in natural image classification problems. Here, we conduct an experiment on the performance of CNN for vegetable image classification by developing a CNN model from the ground. Additionally, several pre-trained CNN architectures using transfer learning are employed to compare the accuracy with the typical CNN. This work proposes the study between such typical CNN and its architectures (VGG16, MobileNet, InceptionV3, ResNet, etc.) to build up which technique would work best regarding accuracy and effectiveness with new image datasets. Experimental results are presented for all the proposed architectures of CNN. Besides, a comparative study is done between developed CNN models and pre-trained CNN architectures. And the study shows that by utilizing previous information gained from related large-scale work, the transfer learning technique can achieve better classification results over traditional CNN with a small dataset. And one more enrichment in this paper is that we build up a vegetable images dataset of 15 categories consisting of a total of 21,000 images.

Keywords: Vegetable image classification, deep learning, CNN, VGG16, MobileNet, Inception-V3, ResNet

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	i
Declaration	ii
Acknowledgements	iii
Abstract	iv
CHAPTER	
CHAPTER 1: INTRODUCTION	01-03
1.1 Introduction	01
1.2 Motivation	01
1.3 Rationale of the Study	02
1.4 Research Questions	02
1.5 Expected Output	02
1.6 Project Management and Finance	03
1.7 Report Layout	03
CHAPTER 2: BACKGROUND	04-06
2.1 Preliminaries	04
2.2 Related Works	04
2.3 Comparative Analysis and Summary	05
2.4 Scope of the Problem	06
2.5 Challenges	06

CHAPTER 3: RESEARCH METHODOLOGY	08-16
3.1 Research Subject and Instrumentation	08
3.2 Data Collection Procedure	08
3.3 Statistical Analysis	09
3.4 Proposed Methodology	10
3.4.1 CNN	11
3.4.2 Transfer Learning	13
3.4.3 Fine-tuning CNN Architectures for Transfer Learning	14
3.4.3.1 VGG16	14
3.4.3.2 InceptionV3	15
3.4.3.3 MobileNet	15
3.4.3.4 ResNet50	15
3.5 Implementation Requirements	16
CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSION	17-23
4.1 Experimental Results	17
4.2 Comparative Analysis	22
4.3 Discussion	23
CHAPTER 5: IMPACT AND SUSTAINABILITY	24-25
5.1 Impact on Society	24
5.2 Impact on Environment	24
5.3 Ethical Aspects	24
5.4 Sustainability Plan	24

CHAPTER 6: CONCLUSION AND FURTHER RESEARCH	26
6.1 Conclusion	26
6.2 Further Study	26
REFERENCES	27-28

LIST OF FIGURES

FIGURES	PAGE NO
Fig. 3.1: Random Examples From The Vegetable Image Dataset	09
Fig. 3.2: Number of Images Per Class	09
Fig. 3.3: Proportion of Each Observed Category	10
Fig. 3.4: CNN Layers	11
Fig. 3.5: Developed CNN Model	12
Fig. 3.6: Different Learning Process	13
Fig. 4.1: 6-Layer CNN Training-Validation Loss and Accuracy	17
Fig. 4.2: Confusion Matrix of Proposed 6-Layer CNN	17
Fig. 4.3: VGG16 Training-Validation Loss and Accuracy	18
Fig. 4.4: Confusion Matrix of Fine-Tuned VGG16	18
Fig. 4.5: InceptionV3 Training-Validation Loss and Accuracy	19
Fig. 4.6: Confusion Matrix of Fine-Tuned InceptionV3	19
Fig. 4.7: MobileNet Training-Validation Loss and Accuracy	20
Fig. 4.8: Confusion Matrix of Fine-Tuned MobileNet	20
Fig. 4.9: ResNet50 Training-Validation Loss and Accuracy	21
Fig. 4.10: Confusion Matrix of Fine-Tuned ResNet50	21
Fig. 5.1: Market Size of Automation	25

LIST OF TABLES

TABLES	PAGE NO
Table 4.1: Result Summary	22
Table 4.2: Existing Methods, Datasets and Results	22

CHAPTER 1

INTRODUCTION

1.1 Introduction

Vegetables are one of the most common food items in everyday meals worldwide. People around the world produce many kinds of vegetables. And there are similarities in many vegetable types in terms of color, texture, and shape. Even country-wise same vegetable has a different name.

From vegetable production to delivery, several common steps are operated manually. Like picking, and sorting vegetables. And recognizing a vegetable is a hard task for the customer in the market, as there are similarities between different vegetables. As several steps of vegetable production to consumption is still depended on manual operation with a huge sum of labor constrain, it truly influences the advancement of commercialization of vegetable items. To solve this issue, automation in vegetable picking, sorting, labeling is required by introducing a vegetable image classifier so that time and money can be saved. The purpose of this paper is to classify vegetable images with higher accuracy, with the help of CNN and DCNN based transfer learning. Nowadays convolutional neural network is often used for classification, segmentation, image recognition, etc. Deep network architecture is the utmost power of CNN which enables CNN to automatically learn mid to high-level considerations from new data [1], [2]. In this research, we proposed a CNN model developed from scratch as well as four fine-tuned state-of-the-art CNN architectures (InceptionV3, VGG16, Resnet, MobileNet) for vegetable image classification. Also, a comparative study is done based on the performance between CNN and its architectures.

1.2 Motivation

In modern days, in the field of agriculture, fundamental research work are classification and detection. On our planet there exists almost hundreds of thousands of species of vegetables according to some survey [3]. And vegetables are important for human beings as a result of their nutrients, minerals, phytochemical mixtures, and dietary fiber content. There are various sorts of vegetables and numerous individuals don't have any idea about them. So, the plan of a vegetable classifier will likewise carry straightforwardness to individuals lives. Also, the sorting of vegetables is manually done in super shops and distribution centers. The motivation for this work came from these two reasons. Therefore, we decided to solve this problem using deep learning

architecture, by developing a model that can detect and classify vegetables. This model can be implemented in different types of devices and can also solve other problems related to the identification of vegetables, like labeling the vegetables automatically without any need for human work.

1.3 Rationale of the Study

Farming has an enormous number of contributions to the world. Vegetables are probably the biggest piece of agrobusiness. But this sector is not digitalized like others, that's why the proper output is not available in this sector. Modern technologies like AI, Machine learning, Computer Vision, Deep learning should be part of this sector in this modern era. To digitalize, automation is the most important thing to do. Without knowing vegetables, machines can't be automated. This work takes care of these issues.

1.4 Research Questions

One of the most important parts of research work is to start from the right place. From collecting data to implementation, we faced several vital questions.

Q.1 How people will be benefited from this research?

Q.2 How the industry will be benefited from this research?

Q.3 Is it possible to get 99% accuracy in our solution?

Q.4 Which procedure is more effective and efficient?

1.5 Expected Output

The main goal of this study is to create a vegetable image classifier so that vegetable production-consumption cycle can be automated as well as reducing the human work load. Also, the dataset created for this study, will help the researchers in future work as it's the largest vegetable image dataset. However, from the research conducted we have got the following outputs also-

1. Finding viable arrangements for sorting issues and help individuals who don't have a clue about the diverse sort of vegetables.
2. Utilizing those ways to deal with identifying neighborhood vegetables.
3. Diminish Labor Intensity.
4. Exploring how AI helps to tackle the arranging issues in the horticulture area.
5. Making it simple for mass individuals to find out nearby vegetables.

6. High accuracy of perceiving and organizing.
7. Speeding the human-computer interaction.

1.6 Project Management and Finance

At present, our work is in a limited scope. Training and development of the models can be done by the systems available to us. So currently, there are no finance-related issues. As we are developing only the system, the project management is easily handled by us. But when different types of devices require our system, we have to do some device-level modification to our system which may turn out costly to maintain and upgrade. So, when we need financial help, we will ask that from responsible authorities.

1.7 Report Layout

This report is part of six separate sections to present our work. There are few segments under every section which makes the understanding of this paper straightforward.

Chapter 1: Introduction, it's the absolute first part. This will assist a reader with understanding what we will genuinely investigate all through the examination. This part also shows the motivation driving the work, research question, objectives, expected outcome.

Chapter 2: Background, shows the base of the examination. In this section, existing related researches were discussed. We additionally examine the distinction between related work and present work.

Chapter 3: Research Methodology, shows the strategy for work. This segment comprises the research subject, data assortment. How data is pre-handled and subsequent to preparing data and what is the statistical measurement.

Chapter 4: Results and Discussion, shows the model results, arrangement, and result from the work.

Chapter 5: Impact on Society, Environment and Sustainability, is about the effect of the work in the climate, how it aids society, and its maintainability.

Chapter 6: Conclusion and Future Research is the last part of this report. In this part, we talk about the closure of the work, the summary, our constraints, and how we need to manage this task in the future is examined in this section.

CHAPTER 2

BACKGROUND

2.1 Preliminaries

Deep learning algorithms are very much popular to researchers for classification and recognition problems, specially, CNN-based architectures are a blessing in the field of ML nowadays. Besides, Probabilistic Neural Networks (PNN), You Only Look Once (YOLO), Support Vector Machine (SVM), Region Proposals (R-CNN, Faster R-CNN) technology are also available. This research work is based on image classification. Here CNN and DCNN-based architectures are introduced for image classification.

In this chapter, we discussed some related work done by researchers. We also worked on identifying some scope of the problems and challenges so that classification accuracy can be improved.

2.2 Related Works

This research is about classifying vegetable images with higher accuracy and efficiency. And also finding out the efficient model along with a technique that performs well in terms of accuracy, time, and cost. Previously few works are done by researchers that are related to our work.

Almost 3 years ago, Om Patil [4] used InceptionV3(known as GoogLeNet) for vegetable classification tasks. By fine-tuning the inception network and applying the transfer learning technique, the proposed model can classify 4 types of vegetables- carrots, onions, cucumbers, and tomatoes. The accuracy of their fine-tuned inception-V3 is 99% for a comparatively smaller dataset that contains around 1200 images.

Yuki Sakai et al. [5], proposed deep neural network for classification of vegetable by extracting features and learning the object. Recognition rate of their DNN model was 97.58%. They worked with a very small dataset, having 8 types of vegetables and the dataset contains only 200 total images. For the learning process of vegetable recognition, they used 3 million iterations to get that accuracy, which is lengthy and expensive as well.

Frida Femling et al. [6], deep convolutional neural network-based transfer learning model has been used. For image collection, they used Raspberry Pi. They worked with only inception-V3 and MobileNet and got 96% accuracy with inception-V3 and 97% accuracy with MobileNet

respectively. By two properties their model is appreciated. One is propagation time, another is how much time it takes for classifying a fruit or vegetable image. But the dataset size is small as it contains 4,000 images from ImageNet and a total of 4,300 images of 10 classes.

Zhu L et al. [7], they proposed AlexNet network for vegetable image classification. And also a comparative study is done by addressing the Support Vector Machine classifier and the traditional back propagation neural network. They worked with 5 types of vegetables- pumpkin, mushrooms, broccoli, cauliflowers, and cucumber. And images were obtained from the ImageNet dataset and were expanded by adopting the data expansion method to a total of 24,000 images so that overfitting is reduced. The highest accuracy they gained from their experiment is 92.1% with the AlexNet network.

Guoxiang Zeng [8], proposed image saliency technique and VGG architecture for the classification task of fruit and vegetables. To reduce the unnecessary noise from the image, they extracted dense features from each image and filtered the complicated backgrounds of it. And for determining the significant area in an input image, they choose a bottom-up graph-based visual saliency (GBVS) model[9]. They used a total of 12,173 images spanning 26 categories, among them 13 categories were vegetables (3678 images)- broccoli, celery, cowpea, green onion, garlic, cucumber, mushroom, carrot, onion, pumpkin, chinese cabbage, tomato, and pepper. And the classification accuracy of their model is 95.6%.

He et al. [10] designed an SVM classifier for fruit and vegetable classification, which is based on appearance features extraction such as Scale-invariant feature transform and Histogram of Oriented Gradient to make contributions to automatic selling of these goods in supermarkets.

Li et al. [11] proposed an improved VGG model (VGG-M-BN) and got 96.5% accuracy. They worked with 10 categories of vegetables. Images were mostly collected from the ImageNet dataset and were expanded by adopting the data expansion method.

2.3 Comparative Analysis and Summary

Previously, few works are done that are related to our works. Though that works are good, all of them have limitations in terms of accuracy, efficiency, and effectiveness. The common problem we found in those works is the size of the dataset, and dataset source. Secondly, the training time of most of those work is so long. And in terms of cost and time, long training time is inefficient. And very few researchers work with only vegetable classification [4], [5], [7], [11] but all of that

work have limitation. The minimum type of vegetable they worked with was four [4], and the maximum was ten [11].

In our research, we started our work by creating our vegetable image dataset that is comparatively large from the dataset used in previously done research work. We select 15 types of vegetables that are used often in daily life all over the world. In this research work, we have tried CNN and CNN architecture-based transfer learning model with the fine-tuning and comparative study is done on the performance of among model with previous work for vegetable classification. And the accuracy we found is significant.

2.4 Scope of the Problem

Research in the agricultural sector has created great expectancies in educational research. That's why we choose vegetable image classification for contributing to the industry and for the better livelihood of the people.

This research proposed a model with the highest accuracy rate by using CNN architectures to classify the vegetable. From thousands of vegetable species throughout the world, different types of vegetables have similarities in looking. So, people often get confused in supermarkets most of the time in order to find out the specific vegetable. In the vegetable production to consumption cycle, vegetables are picked and sorted manually. That needs much time and labor. This work will analyze the similarities and differences between same-looking vegetables and train a model that is able to distinguish those properties and detect them with the highest accuracy. The advantages of this model are - its cost-effective and required less time.

Regardless of choice, this model can be implemented in a web application or a mobile application or a physical device, or any platform in order to reach out to most people. None of those implementations aren't discussed in this paper. Only the main steps of processing data and how the algorithm or model should be fine-tuned are discussed here.

2.5 Challenges

The first challenge is to select vegetables that are often consumed throughout the world. Then collecting images of those from the local market with proper quality is a big challenge. Because of the availability of those vegetables locally and the image quality. The quality of images specially the RGB value, and resolution has a big impact on image preprocessing to train the model. Mainly,

the complex background of images and similarities between different vegetables gives more challenge. And different vegetables are looked the same due to the angle, color, and resolution. It's hard to simply depend on a solitary element, like tone, shape, or surface to recognize the types of vegetables.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Research Subject and Instrumentation

We conducted our research to find out the highest accuracy for vegetable image classification. Initially, we experimented with 15 types of common vegetables that are found throughout the world. The vegetables we chose for experimentation are-

- I. Bean
- II. Bitter Gourd
- III. Bottle Gourd
- IV. Brinjal
- V. Broccoli
- VI. Cabbage
- VII. Capsicum
- VIII. Carrot
- IX. Cauliflower
- X. Cucumber
- XI. Papaya
- XII. Potato
- XIII. Pumpkin
- XIV. Radish
- XV. Tomato

To complete our experiment, we used Tensorflow, Keras, scikit-learn, seaborn, NumPy, matplotlib, etc. to train our intended models. And we used Google Colaboratory as our training environment.

3.2 Data Collection Procedure

Data is a major concern in order to train any deep learning algorithm. Dataset size and quality directly impact the accuracy of the deep learning algorithm. And there is no open dataset for the different vegetable images. So, we decide to create our dataset.

- I. **Capturing Images:** We capture images of our selected vegetables directly from fields, vegetable markets, houses.
- II. **Download Images:** Less than 1% of our image data is collected from the internet.

We used to roam different places for capturing different vegetable images. Creating a dataset consisting of 21,000 images taken 3 months, as some vegetables are produced seasonally. We took high-resolution images in a 1:1 ratio with a high-end device. And tried our most to reduce complex backgrounds while taking the pictures. Pictures are taken from different angles. Fig. 3.1 shows some random examples of our dataset.



Fig. 3.1. Random Examples From The Vegetable Image Dataset.

3.3 Statistical Analysis

Initially working with thousands of species of vegetables is hard. So we decided to conduct our research primarily with 15 types of vegetables. We collected a total of 25500 images, but for making our dataset more clean and efficient, we do some preprocessing so that it removes inconsistencies.

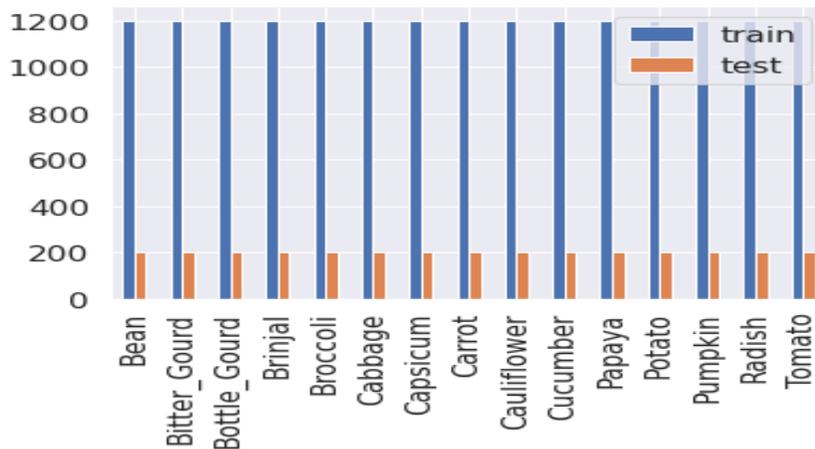


Fig. 3.2. Number of Images Per Class.

After processing we got 21000 images from 15 classes, where each class contains a total of 1400 images (Fig. 3.2). As a result, in the dataset, each class has an equal proportion (Fig. 3.3). Each image resolution is 224×224 and in *.jpg format. We split our dataset into three parts, where 70%(approx.) for training and 15%(approx.) for testing, and the rest 15%(approx.) for validation.

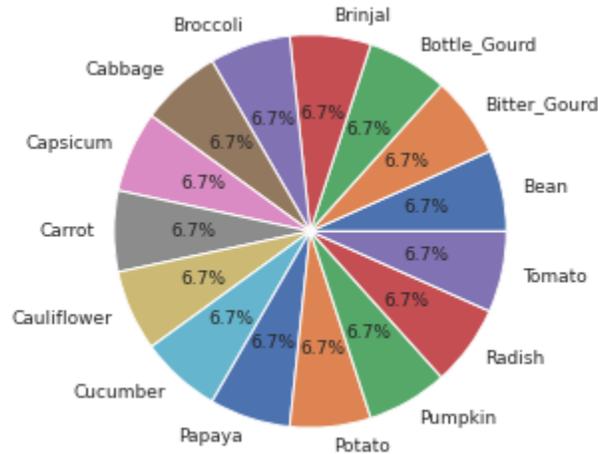


Fig.3.3. Proportion of Each Observed Category.

3.4 Proposed Methodology

In this study of vegetable image classification, we proposed a developed CNN model and state-of-the-art CNN model with transfer learning technique. In 2012, a breakthrough occurred in the image recognition area through the ILSVRC(ImageNet Large-Scale Visual Recognition Challenge) competition on the ImageNet dataset [12]. As of now different types of deep convolutional neural network architectures are introduced that are varies in numbers of layers as well as complexity and available for use by anyone. In the past few years, using transfer learning concepts like fine-tuning and layer freezing in CNN architectures beat the traditional machine learning models in terms of performance and efficiency for image classification problems. Here we used a CNN model and four state-of-the-art CNN architectures - VGG16, InceptionV3, ResNet, and MobileNet. These four models are pre-trained on the ImageNet dataset, a large-scale dataset that contains 1.2 million training data mostly animals and daily objects. By applying transfer learning techniques, learned features of these DCNN models may help to make very deep network architecture effective for our dataset.

3.4.1 CNN

In the field of computer vision, convolutional neural network basically a normalized multilayer perceptron (MLP) has been the most influential innovation. In recent years, CNN has dominated the computer vision and image processing field for large-scale image recognition, classification, and segmentation task. A typical CNN starts with an input layer, ends with an output layer, where in between them there exist multiple hidden layers. Convolutional, pooling, normalization (ReLU), fully-connected layers are part of the hidden layer [13]. Fig. 3.4 shows typical CNN layers.

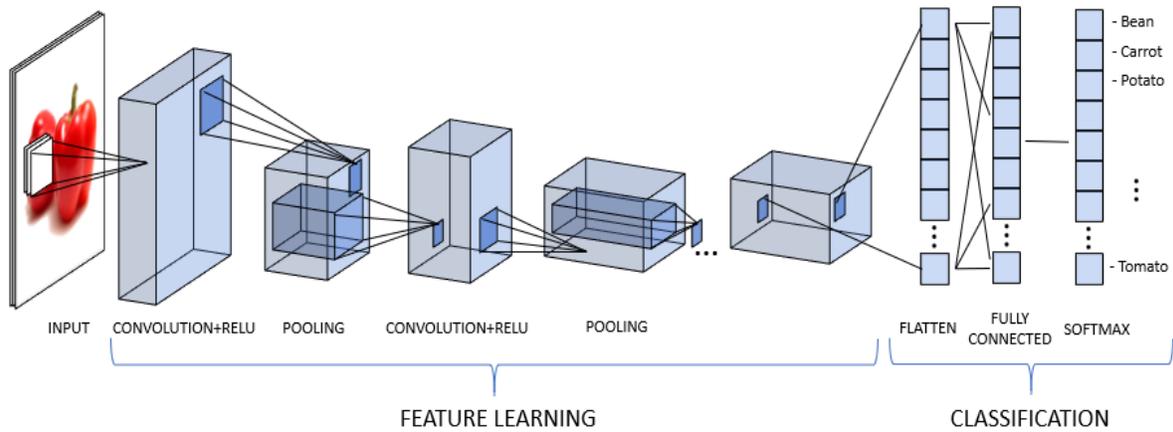


Fig. 3.4. CNN Layers [14].

Input layer takes the target image data as input. The image is then reshaped to an optimal size and forwarded to the next layer which is a convolutional layer. There exists a number of kernel or filter that actually slides over the input and performs element-wise multiplications in order to extract features. Here, through an activation function, the negative weighted input will be replaced with zero otherwise it will go to output directly. The most widely used activation function is ReLU (Rectified-Linear Unit) and for many types of neural networks, it is the default activation function. It's a non-linear function and faster than the other activation function like- Sigmoid, Scaled Exponential Linear Unit (SELU), Exponential Linear Unit (ELU), Gaussian Error Linear Unit (GELU), etc. If x is the input value in ReLU then it can be defined as:

$$R(x) = \max(0, x) \quad (1)$$

Features that are extracted from the convolutional layer then sent to the pooling layer. This layer preserves only important features from a large image by reducing parameters. Then the fully connected layer translates these highly filtered images into categories. And another non-linear

function named softmax finally gives the decimal probabilities ranged from 0 to 1 to each class. If \vec{x} is the input vector, x_i is the current real value and $\sum_k e^{x_k}$ is the normalization term ranged from 0 to 1, where k is the total number of classes, then the softmax function $s_i(\vec{x})$ can be defined as:

$$s_i(\vec{x}) = \frac{e^{x_i}}{\sum_k e^{x_k}} \quad (2)$$

In this research, we propose a 6-layer convolutional neural network that is completely built from scratch. We scaled down our image dataset from 224×224 to 32×32 for reducing overall computational time so that a good model can be created in terms of efficiency. Then we apply data augmentation techniques like rotation, rescale, shear, zoom, and horizontal flip on our 32×32 size 3-channel training image data. We used ReLU as our activation function with each convolutional layer. And for improving generalization error a dropout rate of 0.25 is used so that it can overcome overfitting issues. Finally, softmax is used in order to find the probabilities of each class in decimal numbers. Fig. 3.5 shows our developed CNN model architecture.

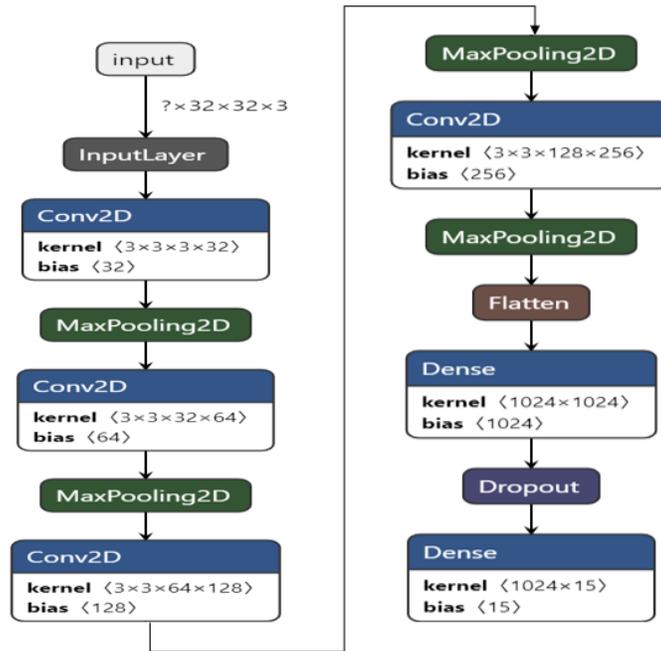


Fig. 3.5. Developed CNN Model.

3.4.2 Transfer Learning

Building and training a deep convolutional neural network from scratch is time-consuming, costly, and hard. A deep network means it contains multiple layers, where it can also be multiple convolutional layers in exact order or sequence in order to classify the exact image. To learn feature mapping, when this kind of large and deep architecture is being trained from the ground, it needs a large-scale dataset.

Transfer learning is a technique of re-using a previously developed model on a second related task. The utilization of knowledge that is learned from a previous domain for the improvement and optimization of a new domain is the core idea of it[13]. The concept of transfer learning in the field of ML was first presented in a NIPS-95 workshop named “Learning to Learn”, agenda of which was a lifelong ML technique that is able to hold and reuse already learned information [15]. Fig. 3.6 shows the difference between transfer learning and the traditional learning process. It can quickly transfer learned features from one(source) domain to the another(target) domain using a smaller dataset in the fastest manner using the easiest way [16]. Hence, the concept of transfer learning is adopted in this research and our target domain is vegetable image classification.

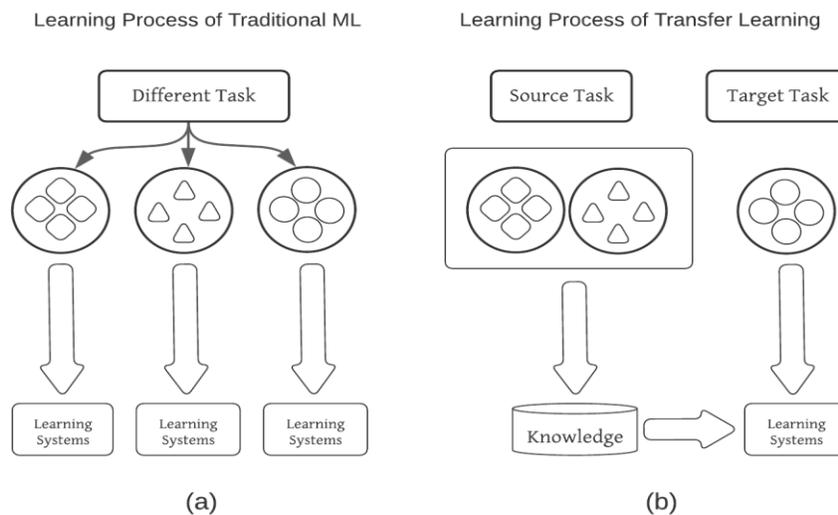


Fig. 3.6. Different Learning Process Between (a)Traditional ML and (b)Transfer Learning.

Two approaches are available for implementing transfer learning-

1. Pre-trained Model Approach
2. Develop Model Approach

In deep learning, the most commonly used approach is the “Pre-trained Model Approach” and it was selected for this research. In this approach, a pre-trained source model that is trained on large-scale data is selected, and then the whole model or parts of the model are used as the starting point for a model of another task. Where fine-tuning of the model may be required on the input-output pair of the target domain. Fine-tuning refers that, keeping the weights and biases of some layers unfrozen and used them for training so that the pre-trained model can perform well on the training data.

In this paper, we propose transfer learning techniques on four pre-trained state-of-the-art CNN architectures - VGG16, InceptionV3, ResNet50, and MobileNet with fine-tuning.

3.4.3 Fine-tuning CNN Architectures for Transfer Learning

Fine-tuning refers to the technique of using learned features or weights and biases from a pre-trained deep CNN as the initialization of a target CNN model so that the target CNN can be trained on target data in a supervised manner [1]. As the interrelation between our target dataset and the source domain dataset is notable, for each architecture we used layer-wise fine-tuning. We fine-tuned our four state-of-the-art CNN architectures by freezing the convolutional base so that previously learned weights and biases can be repurposed in our task. We used the full architecture as it is, except the output layer which is basically the last fully connected layer. Here, convolutional base is the fixed feature extractor, and the extracted feature will be used for classifying the input image. For retraining these transferred networks, we set the number of classes in the output layer to 15 referring to our multi-class classification task. Finally, we retrained the final layer.

3.4.3.1 VGG16

Visual Geometry Group (VGG) network contains VGG16 and VGG19. VGG16 was the first runner-up in ILSVRC and consists of 16 convolutional layers, three fully-connected, and five max-pooling layers [2]. It has over 138million parameters. It uses the ReLU activation function and dropout for improving generalization error with all fully connected layers, and also uses the softmax function in the output. In this model several 3*3 filters are used in order to replace the large-size filters or kernels. And these small-size kernels give the opportunity of complex feature extraction at a low cost.

3.4.3.2 InceptionV3

InceptionV3 is a state-of-the-art CNN architecture developed by google also known as GoogleNet. It has 48 layers and it replaced the last fully connected layer with average pooling right after the last convolutional layer. As a result, total number of parameters (24 million) is reduced and makes this model more computationally efficient. In InceptionV3 it has eleven inception module, each module contains convolutional layers with ReLU activation function, convolutional filter for dimension reduction, max-pooling layer, fully-connected layer, and an output layer along with a softmax activation function [12]. It uses the ReLU activation function and dropout for improving generalization error with the fully connected layer.

3.4.3.3 MobileNet

MobileNet is a DCNN architecture that uses depthwise separated convolutions to construct lightweight deep convolutional neural networks which are suitable for mobile and embedded vision applications [17]. Here, depthwise separable convolutions mean, it does the typical combining and filtering tasks in different separate layers. Except for the first layer, it uses depthwise separable convolutions in order to reduce the model size rather than typical convolutions so that computational efficiency gets increased. It has total of 28 layers where all are followed by batch normalization. It uses ReLU as an activation function and uses the softmax function in the output.

3.4.3.4 ResNet50

ResNet elaboration is Residual Network, which is a particular sort of neural network that was presented in 2015 by Microsoft [18]. ResNet50 is an example network from ResNet network architecture. ResNet introduces a revolutionary technique to overcome the degradation problem named “residual mapping”. It has 50 layers and over 23 millions of trainable parameters. It uses global average pooling instead of fully connected layers than other standard DCNN architecture. Though it’s much deeper than other used architectures in this research, it’s considerably light-weight.

3.5 Implementation Requirements

Firstly, for the implementation of our proposed CNN model and fine-tuned CNN architectures availability of good quality dataset with maximum possible classes of vegetable is required. Then preprocessing of image data is very important, as the quality (size, texture, color, resolution) of images will directly influence the model accuracy. Also, data augmentation is a good practice and necessary for the small-size dataset. Basic knowledge of ML and CNN theory, as well as programming skill required. Google Colaboratory can be used for the developing environment as it's free and for that, a computer with an internet connection is required. Otherwise, high-performance physical GPU is a must for training the model along with the computer.

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Experimental Results

All of the experiment in this paper is done by using own dataset, consisting of 21,000 images from 15 types of vegetables, and the dataset split 70% for training, 15% for validation, and 15% for testing purpose.

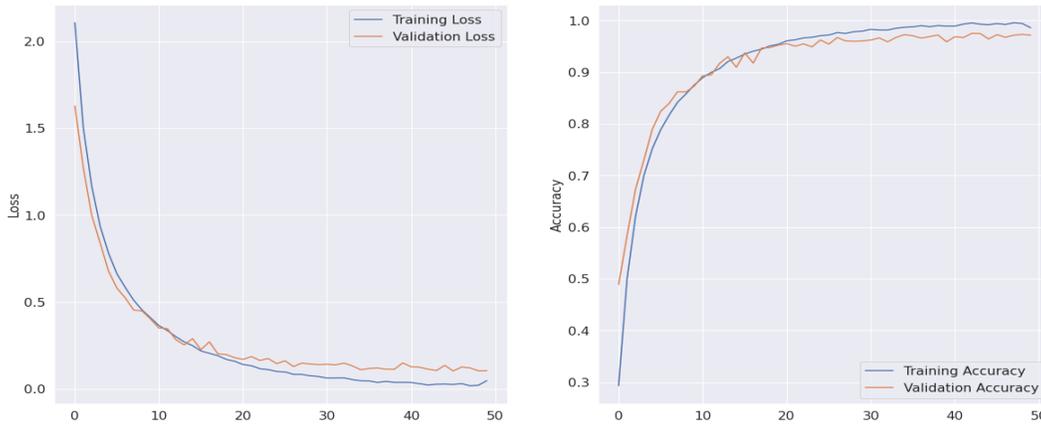


Fig. 4.1. 6-Layer CNN Training-Validation Loss and Accuracy.

In our developed 6-layers CNN model, the 3-channel input image size is 32×32 and the total number of parameters is over 1.4 million. In the 6-layer CNN model, there are four conv2D layers

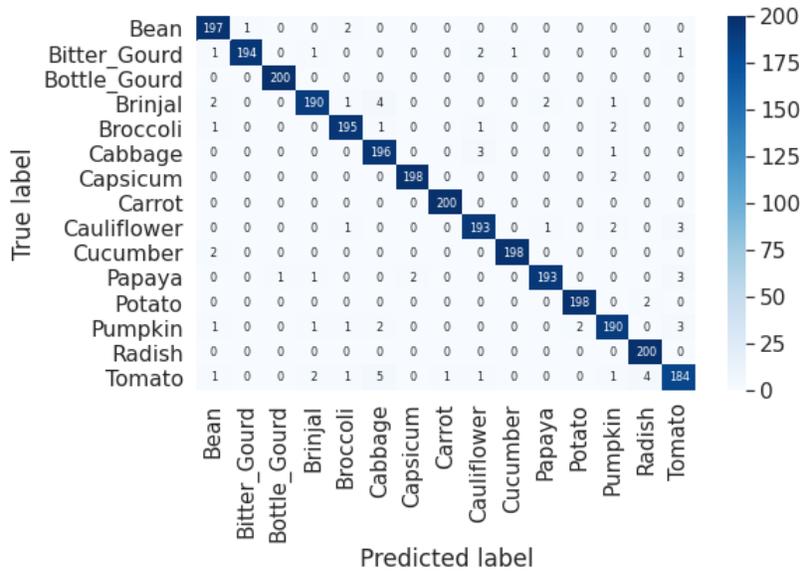


Fig. 4.2. Confusion Matrix of Proposed 6-Layer CNN.

and 2 fully connected layers. Adam optimizer is used along with a learning rate of 0.0001, and

training is done with a batch size of 64 in 50 epochs. With this combination of training, we have got the best accuracy from the developed CNN model. Fig. 4.1 shows the graph of training-validation loss and accuracy. The validation accuracy for our developed CNN model is 97.6% and testing accuracy is 97.5%. Also Fig. 4.2 shows the confusion matrix of the 6-Layer CNN model. VGG16 is a 16-layer CNN architecture that has over 138 million parameters. Input image size is

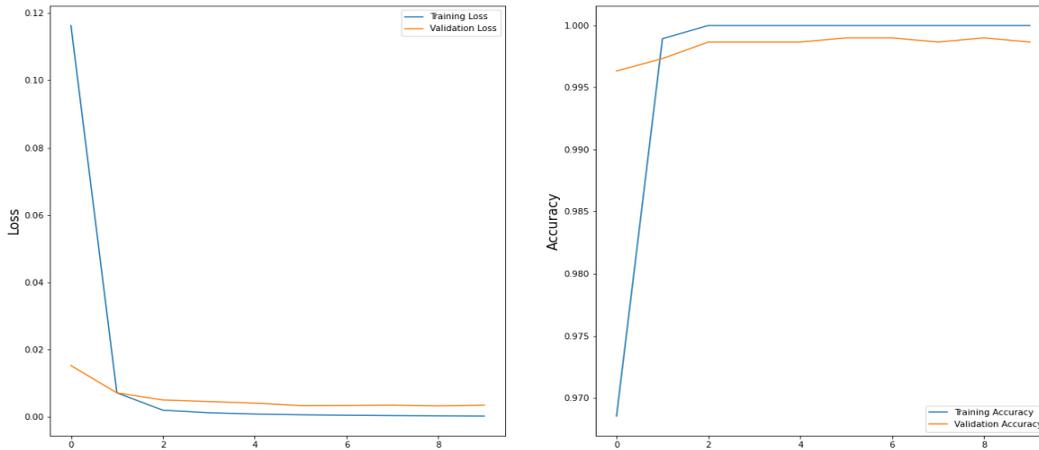


Fig. 4.3. VGG16 Training-Validation Loss and Accuracy.

224×224 and in-built image pre-processing of VGG architecture is used before they get passed to the network. Adam optimizer is used along with a learning rate of 0.0001, and training is done with a batch size of 64 in 10 epoch. Fig. 4.3 shows the training-validation loss and accuracy graph

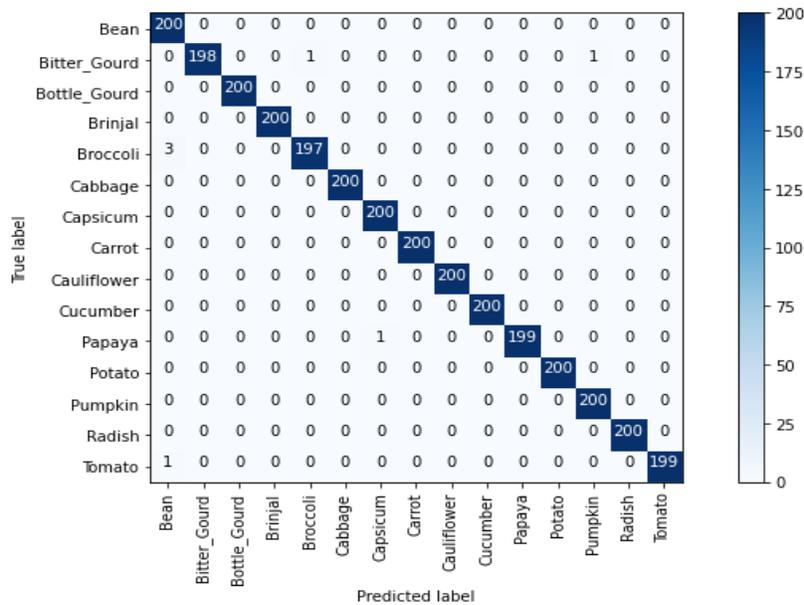


Fig. 4.4. Confusion Matrix of Fine-Tuned VGG16.

for VGG16. With fine-tuning the architecture, we have got 99.8% validation accuracy and 99.7% testing accuracy from the VGG16 model. And Fig. 4.4 shows the confusion matrix of the model.

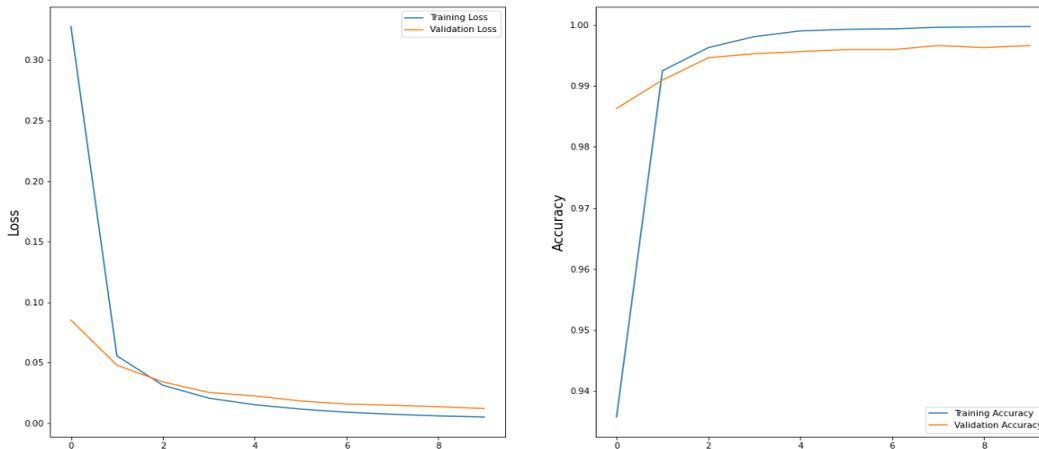


Fig. 4.5. InceptionV3 Training-Validation Loss and Accuracy.

InceptionV3 has 48 layers and 24 million parameters. The input image size is 299×299 and in-built image pre-processing of Inception architecture is used before they get passed to the network. Adam optimizer is used along with a learning rate of 0.0001, and training is done with a batch size of 64 in 10 epoch. Fig. 4.5 shows training vs validation loss and accuracy graph for InceptionV3. With fine-tuning the architecture, we have got 99.6% validation accuracy and 99.7% testing accuracy. Fig. 4.6 shows the confusion matrix for the fine-tuned InceptionV3.

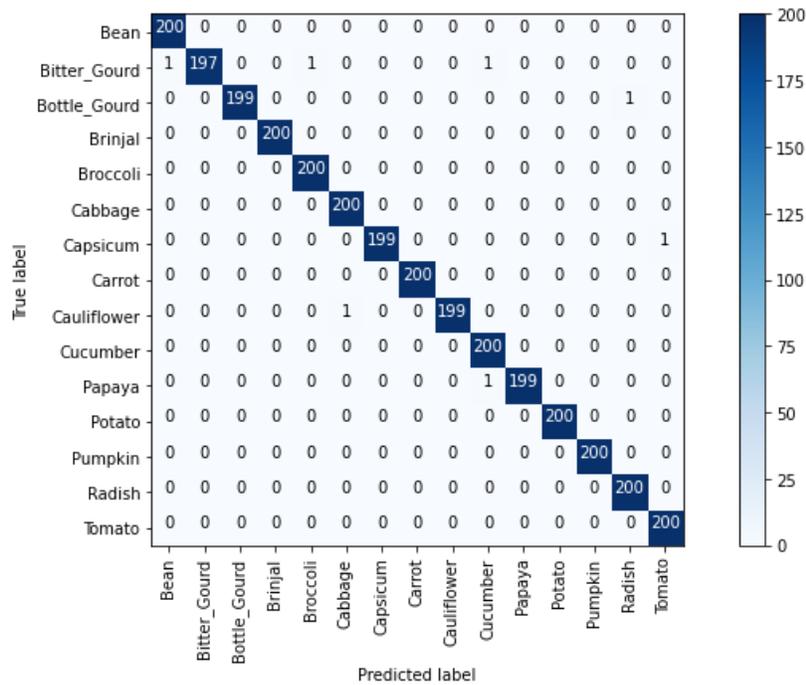


Fig. 4.6. Confusion Matrix of Fine-Tuned InceptionV3.

MobileNet is a very light-weight CNN architecture comparing to other architectures used in this research has only 28 layers and the input image size is 224×224. In-built image pre-processing of

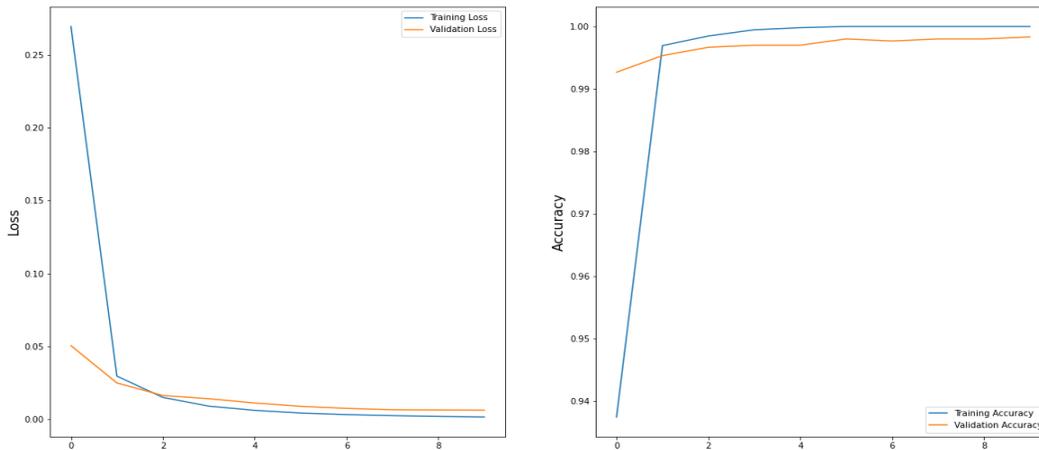


Fig. 4.7. MobileNet Training-Validation Loss and Accuracy.

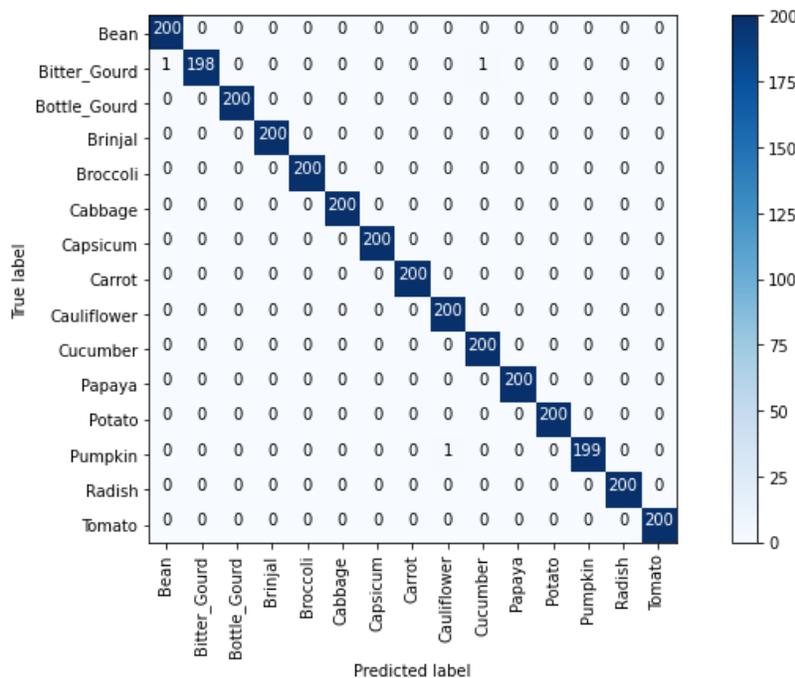


Fig. 4.8. Confusion Matrix of Fine-Tuned MobileNet.

MobileNet architecture is used before each image gets passed to the network. MobileNet V1 is used for the experiment. Adam optimizer is used along with a learning rate of 0.0001, and training is done with a batch size of 64 in 10 epoch. With fine-tuning the architecture, we have got 99.8% validation accuracy and 99.9% testing accuracy. Fig. 4.7 shows training vs validation loss and accuracy graph for MobileNet. And Fig. 4.8 shows the confusion matrix for the model.

ResNet is also used in this experiment specifically ResNet50 that has only 28 layers and the input image size is 224×224. In-built image pre-processing of ResNet architecture is used before each image gets passed to the network. Adam optimizer is used along with a learning rate of 0.0001, and training is done with a batch size of 64 in 10 epoch. Fig. 4.9 shows the training-validation loss and accuracy graph for ResNet50. With fine-tuning the architecture, we have got 99.9% validation accuracy and 99.9% testing accuracy. Fig. 4.10 shows the confusion matrix for the fine-tuned ResNet50.

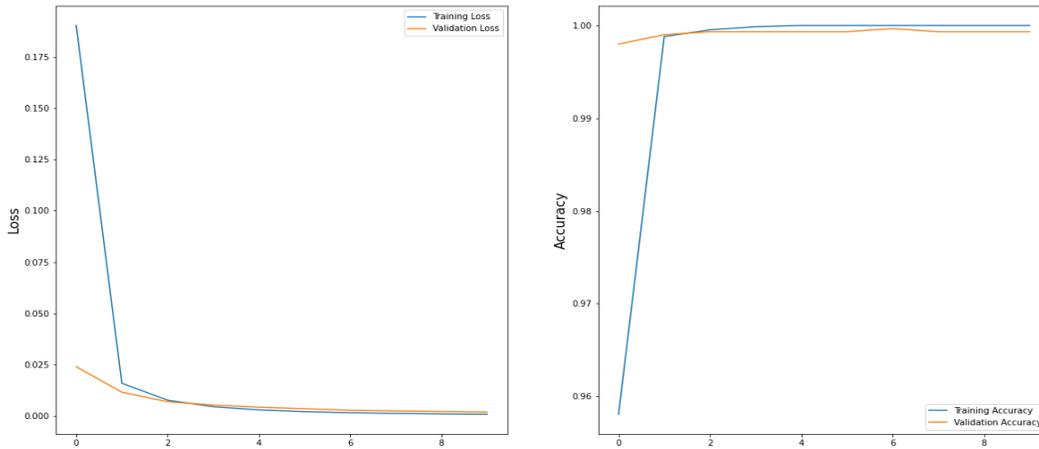


Fig. 4.9. ResNet50 Training-Validation Loss and Accuracy.

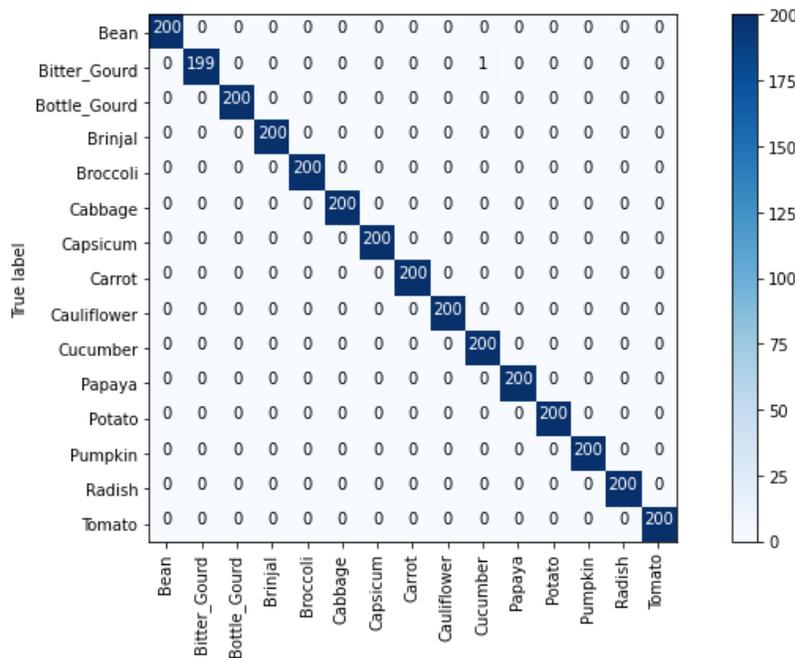


Fig. 4.10. Confusion Matrix of Fine-Tuned ResNet50.

4.2 Comparative Analysis

Building a CNN model from scratch is not a very easy task and with a small dataset it's harder to find out the best accuracy from a developed CNN model. For bringing out the best possible accuracy, tweaking the CNN models such as adding more layers, dropout, changing activation function, trying with different optimizer along with learning rate is important. Proposed 6-layer CNN is tweaked and optimized for the working vegetable dataset and gives an accuracy of 97.5%, which is the highest compared with all the previous work conducted by building a model from scratch. Table 4.1 shows the result summary along with the applied technique. All the previous works done by using state-of-the-art CNN model has no major impact, the accuracy of those models is not significant as most of the experiment was done on a smaller dataset or collected from ImageNet. Table 4.2 shows previous methods, dataset size and results. With the proposed fine-tuning approach in state-of-the-art CNN architecture, the output is significantly impressive. All the four DCNN architecture used with transfer learning technique gives the accuracy over 99% each. Maximum accuracy is achieved from MobileNet and ResNet which is 99.9%.

Table 4.1. Result Summary

Technique	Method/Algorithm	Dataset size	Epochs	Accuracy	Training Time
Build from scratch	CNN	21000	50	97.5%	≈ 1 hour
Transfer learning	VGG 16		10	99.7%	≤ 40 min.
	Inception V3		10	99.7%	< 1 hour
	MobileNet		10	99.9%	< 30 min.
	ResNet		10	99.9%	< 30 min.

Table 4.2. Existing Methods, Datasets and Results

No	Author	Method/Algorithm	Dataset Size	Dataset Source	Classes	Accuracy
1	Om Patil et al. [4]	Inception V3	1200	Self-collected	4	99%
2	Yuki Sakai et al. [5]	DNN	200	Self-collected	8	97.38%
3	Frida Femling et al. [6]	MobileNet	4300	ImageNet	10	96%
		Inception V3				97%
4	Zhu L et al. [7]	AlexNet	24000	ImageNet	5	92%
5	Guoxiang Zeng [8]	VGG	3678	Self-collected	13	95.6%

4.3 Discussion

Typical CNN and DCNN based architecture have the major difference in terms of depth and complexity of the network. Building CNN from scratch requires prior knowledge on the related field. But it is not so hard to make a standard CNN model. The problem is CNN requires a large-scale dataset for good results. Small size dataset or unoptimized CNN model often causes an overfitting issue. Good accuracy may be achieved by increasing the depth of the network. Increasing depth means more layers, more filtering, more trainable parameters that may give better accuracy, and also vice-versa. But increasing the layer most of the time is not a suitable way cause it increases the training parameters as a result model gets heavier and training time increased. While training the deep network, an increase in layer sometimes causes accuracy to overfill, then a rapid degradation occurs. So, its needs optimization also. Making an optimized deep CNN is very long-term work and isn't possible to make without specific expertise. This research proposed a 6-layer CNN model which is optimized and lightweight. Several techniques also applied during the experiment like increasing depth means adding more layers for an increase in the accuracy but in that case, overfitting issue arise.

DCNN based architectures are very deep and highly optimized networks compare to the typical CNN model. These architectures are complex and often trained on large-scale datasets like ImageNet. But transfer learning technique creates the opportunity to use these models even with a small dataset. It's less time-consuming and easy to implement. Though these models are heavy compared to typical CNN this research shows that they can give higher accuracy.

So, from the result, for vegetable image classification tasks transfer learning technique works better with the used dataset consist of 21000 images. And as of now highest possible accuracy is gained in this research by using the proposed way of fine-tuning those architecture.

CHAPTER 5

IMPACT AND SUSTAINABILITY

5.1 Impact on Society

The work we have done is ordinarily utilized in various sorts of systems to computerize the work process. If it is being utilized appropriately, it tends to be a decent effect on society. Assume one doesn't know which vegetables he is looking at, in that point he can utilize our tech to distinguish the vegetables. Our tech can likewise robotize the sorting and labeling of distribution centers and super shops. It will be quicker better and it will save bunches of significant human resources. In this way, our work will be a decent effect on society.

5.2 Impact on Environment

Our work is on the software level. This is not a physical product so it won't be hurtful to nature or climate. In some cases, it can help us positively. As our work will help in faster labeling and sorting, as a result, the percentage of rotten and wastage will be significantly reduced.

5.3 Ethical Aspects

Our work isn't doing cheat or damage to other people, so there are no issues from moral viewpoints. It does not have a bad impact on the environment, human life. We are just making human life easier by detecting vegetables through deep learning.

5.4 Sustainability Plan

The world is going towards automation. We are also going in the same direction. The AI-powered automation market is growing day by day. So the Sustainability of this work is remarkably acceptable.

We can see from the report (Fig. 5.1) the market size of computerization is expanding very fast and in 2024 it will be a market of USD 90 Billion with a compound annual growth rate of 12%.



Fig. 5.1. Market Size of Automation [19]

CHAPTER 6

CONCLUSION AND FURTHER RESEARCH

6.1 Conclusion

Agriculture is the most important sector but this sector is less focused on digitalization than others. There were some works done previously on vegetable classification but those are in limited scope with a very small dataset and less accuracy. By taking account of those problems this research is conducted to resolve those issues. In this study, vegetable image classification is done using a typical CNN model and CNN-based pre-trained VGG16, InceptionV3, MobileNet, and Resnet50 with two techniques. Proposed typical CNN model has six-layer and was build from scratch. On the other hand, pre-trained state-of-the-art CNN architectures are fine-tuned and applied using transfer learning techniques. From various species of vegetables, only 15 types of vegetables are selected for primary research of vegetable image classification task. A dataset consist of 21000 images from 15 classes is created locally and used for training and testing. A comparative study is also done on the performance between typical CNN and pre-trained CNN, to check which one is better, efficient, less time-consuming. Also, experimental results for different models and techniques are discussed and the overall accuracy achieved 99.9%. From the experimental result, it's clear that pre-trained CNN architectures are the future of machine vision. And as of now, it's the highest possible accuracy for the vegetable classification task, which seems quite promising.

6.2 Further Study

There are some options for future implications. Various types of actual devices can be made by utilizing this work. The sorting and labeling process of vegetables can be automated to save both time and human resources in super shops, warehouses. Extending this work by continuing the study with more classes and types by contributing to the existing dataset, to make it more robust.

REFERENCES

- [1] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang, and T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification," in *Proceedings - 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2017*, 2018, vol. 2018-Janua, doi: 10.1109/CISP-BMEI.2017.8301998.
- [2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015.
- [3] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *32nd International Conference on Machine Learning, ICML 2015*, 2015, vol. 1.
- [4] O. Patil, "Classification of Vegetables using TensorFlow," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 6, no. 4, 2018, doi: 10.22214/ijraset.2018.4488.
- [5] Y. Sakai, T. Oda, M. Ikeda, and L. Barolli, "A vegetable category recognition system using deep neural network," 2016, doi: 10.1109/IMIS.2016.84.
- [6] F. Femling, A. Olsson, and F. Alonso-Fernandez, "Fruit and Vegetable Identification Using Machine Learning for Retail Applications," 2018, doi: 10.1109/SITIS.2018.00013.
- [7] L. Zhu, Z. Li, C. Li, J. Wu, and J. Yue, "High performance vegetable classification from images based on AlexNet deep learning model," *Int. J. Agric. Biol. Eng.*, vol. 11, no. 4, 2018, doi: 10.25165/ijabe.v11i4.2690.
- [8] G. Zeng, "Fruit and vegetables classification system using image saliency and convolutional neural network," in *Proceedings of 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference, ITOEC 2017*, 2017, vol. 2017-Janua, doi: 10.1109/ITOEC.2017.8122370.
- [9] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," 2007, doi: 10.7551/mitpress/7503.003.0073.
- [10] J. He, Y. Ma, Q. L.-J. of C. N. University, and undefined 2016, "Fruit and vegetable automatic classification based on appearance feature.," *cabdirect.org*, Accessed: Mar. 27, 2021. [Online]. Available: <https://www.cabdirect.org/cabdirect/abstract/20163275132>.
- [11] Z. Li, F. Li, L. Zhu, and J. Yue, "Vegetable recognition and classification based on improved VGG deep learning network model," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, 2020, doi: 10.2991/ijcis.d.200425.001.
- [12] A. Rehman, S. Naz, M. I. Razzak, F. Akram, and M. Imran, "A Deep Learning-Based Framework for Automatic Brain Tumors Classification Using Transfer Learning," *Circuits, Syst. Signal Process.*, vol. 39, no. 2, 2020, doi: 10.1007/s00034-019-01246-3.
- [13] M. Hussain, J. J. Bird, and D. R. Faria, "A study on CNN transfer learning for image classification," in *Advances in Intelligent Systems and Computing*, 2019, vol. 840, doi: 10.1007/978-3-319-97982-3_16.
- [14] "Understanding of Convolutional Neural Network (CNN) — Deep Learning | by Prabhu | Medium." <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148> (accessed Mar. 26, 2021).
- [15] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data*

Engineering, vol. 22, no. 10. 2010, doi: 10.1109/TKDE.2009.191.

- [16] V. Chauhan, K. D. Joshi, and B. Surgenor, "Image classification using deep neural networks: transfer learning and the handling of unknown images," in *Communications in Computer and Information Science*, 2019, vol. 1000, doi: 10.1007/978-3-030-20257-6_23.
- [17] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv*. 2017.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, doi: 10.1109/CVPR.2016.90.
- [19] "Logistics Automation Market by Application, Size, Growth and Overview – 2027 | MRFR." <https://www.marketresearchfuture.com/reports/logistics-automation-market-8019> (accessed Mar. 26, 2021).

Similarity Index

DCNN-Based Vegetable Image Classification Using Transfer Learning: A Comparative Study

ORIGINALITY REPORT

15% SIMILARITY INDEX	11% INTERNET SOURCES	6% PUBLICATIONS	8% STUDENT PAPERS
--------------------------------	--------------------------------	---------------------------	-----------------------------

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	6%
2	Submitted to Daffodil International University Student Paper	3%
3	dokumen.pub Internet Source	1%
4	Xiaogang Li, Tiantian Pang, Biao Xiong, Weixiang Liu, Ping Liang, Tianfu Wang. "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification", 2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017 Publication	1%
5	Amit Doegar, Srinidhi Hiriyannaiah, G. M. Siddesh, K. G. Srinivasa, Maitreyee Dutta. "Cloud-Based Fusion of Residual Exploitation-Based Convolutional Neural Network Models for Image Tampering Detection in	<1%

Bioinformatics", BioMed Research International, 2021

Publication

6	arxiv.org Internet Source	<1 %
7	link.springer.com Internet Source	<1 %
8	Submitted to Landmark University Student Paper	<1 %
9	www.mdpi.com Internet Source	<1 %
10	Frank Y. Shih, Himanshu Patel. "Deep Learning Classification on Optical Coherence Tomography Retina Images", International Journal of Pattern Recognition and Artificial Intelligence, 2019 Publication	<1 %
11	Submitted to Birkbeck College Student Paper	<1 %
12	"MultiMedia Modeling", Springer Science and Business Media LLC, 2017 Publication	<1 %
13	"Proceedings of 6th International Conference on Recent Trends in Computing", Springer Science and Business Media LLC, 2021 Publication	<1 %

14	Andrea Gasparetto, Dalila Ressi, Filippo Bergamasco, Mara Pistellato et al. "Cross-Dataset Data Augmentation for Convolutional Neural Networks Training", 2018 24th International Conference on Pattern Recognition (ICPR), 2018 Publication	<1 %
15	d-nb.info Internet Source	<1 %
16	dspace.cc.tut.fi Internet Source	<1 %
17	Guoxiang Zeng. "Fruit and vegetables classification system using image saliency and convolutional neural network", 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC), 2017 Publication	<1 %
18	Li, Zhang, Wu. "Efficient Object Detection Framework and Hardware Architecture for Remote Sensing Images", Remote Sensing, 2019 Publication	<1 %
19	"Computational Intelligence in Pattern Recognition", Springer Science and Business Media LLC, 2020 Publication	<1 %

20	<p>"Innovative Data Communication Technologies and Application", Springer Science and Business Media LLC, 2020</p> <p>Publication</p>	<1 %
21	<p>"International Conference on Communication, Computing and Electronics Systems", Springer Science and Business Media LLC, 2020</p> <p>Publication</p>	<1 %
22	<p>"Machine Learning and Information Processing", Springer Science and Business Media LLC, 2020</p> <p>Publication</p>	<1 %
23	<p>Frida Femling, Adam Olsson, Fernando Alonso-Fernandez. "Fruit and Vegetable Identification Using Machine Learning for Retail Applications", 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2018</p> <p>Publication</p>	<1 %
24	<p>Md. Ferdous Wahid, Md. Jahid Hasan, Md. Shahin Alom, Shamim Mahbub. "Performance Analysis of Machine Learning Techniques for Microscopic Bacteria Image Classification", 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019</p> <p>Publication</p>	<1 %

25	Shahram Taheri, Önsen Toygar. "Animal classification using facial images with score-level fusion", IET Computer Vision, 2018 Publication	<1 %
26	Xiaoxia Li, Wei Li, Qiang Yang, Wenjun Yan, Albert Y. Zomaya. "Edge-Computing-Enabled Unmanned Module Defect Detection and Diagnosis System for Large-Scale Photovoltaic Plants", IEEE Internet of Things Journal, 2020 Publication	<1 %
27	Yunlong Yu, Fuxian Liu. "Dense Connectivity Based Two-Stream Deep Feature Fusion Framework for Aerial Scene Classification", Remote Sensing, 2018 Publication	<1 %
28	pdfs.semanticscholar.org Internet Source	<1 %
29	"Computer Vision and Machine Learning in Agriculture", Springer Science and Business Media LLC, 2021 Publication	<1 %
30	Mariappan Sakkan, Sanat Sarangi, Swagatam Bose Choudhury, Prachin Jain, Hemavathy B, Rajesh Urkude, Srinivasu Pappula. "Disease Assessment in Rice with Pervasive Stationary and Mobile Sensing - Learnings from Rabi 2019-20", 2021 International Conference on	<1 %

COMmunication Systems & NETworkS (COMSNETS), 2021

Publication

Exclude quotes Off

Exclude bibliography On

Exclude matches Off