

**MACHINE LEARNING APPROACH TO CLASSIFY MANGO
BY**

**Abu Abdullah Al Asif
ID: 191-15-12721**

**Md Habibur Rahman
ID: 191-15-12884**

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

Supervised By

Ms. Nazmun Nessa Moon

Associate Professor

Department of CSE

Daffodil International University

Co-Supervised By

Asma Mariam

Lecturer

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

This research titled “Machine Learning Approach to Classify Mango” submitted by “Abu Abdullah Al Asif ” , “ Md. Habibur Rahman ” ID No: 191-15-12721, 191-15-12884 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment for the partial fulfillment of the requirements for the B.S.c in Computer science and Engineering and approved as to its style and contents. The presentation has been held on 26 January 2023

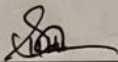
Board of Examiners

Chairman

Dr. Touhid Bhuiyan
Professor and Head

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

Internal Examiner



Subhenur Latif
Assistant Professor

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

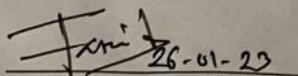
Internal Examiner



Mohammad Monirul Islam
Assistant Professor

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

External Examiner

 26-01-23

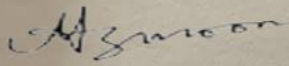
Dr. Dewan Md Farid
Professor

Department of Computer Science and Engineering
United International University

Declaration

We here by declare that this study was conducted under the supervision of **Ms. Nazmun Nessa Moon and co-supervision of Asma Mariam** Daffodil International University's Department of CSE. We further conform that no component of this project has been submitted to any other institution for the purpose of receiving a degree or diploma

Supervised by:



Ms. Nazmun Nessa Moon
Associate Professor
Department of CSE
Daffodil International University

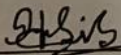
Co-Supervised by:

Asma Mariam
Sr. Lecturer
Department of CSE
Daffodil International University

Submitted by:



Abu Abdullah Al Asif
ID; 191-15-12721
Department of CSE
Daffodil International University



Md. Habibur Rahaman
ID; 191-15-12844
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 Supervisor Ms. Nazmun Nessa Moon, Associate Professor Department of CSE Daffodil International University, Dhaka and Co-Supervisor Asma Mariam, Lecturer Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Machine Learning and Data Mining” to carry out this project. Her endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to Prof. Dr. Touhid Bhuiyan, 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.

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

ABSTRACT

Our research titled "Machine Learning Approach To Classify Mango" is focusing the people not recognize mango species. Our work images in this research use deep learning, also known as machine learning. Python is used as a programming language because of how successfully it functions. Here we used raw data collect from Rajshahi. We are taken six different species of mangos. Here we have taken almost 1500 data. Data ratio 81% is train data and 19% is test data. We are using four algorithms from Transfer Learning Inception V3 95%, VGG19 74%, MobileNet 49%, and Convolutional Neural Networks (CNN) 90% provide test accuracy. Nobody else has performed this kind of mangos classification determination that I've seen.

TABLE OF CONTENTS

CONTENTS	PAGE
Approval	I
Declaration	Ii
Acknowledgments	Iv
Abstract	V
CHAPTER	
CHAPTER 1: INTRODUCTION	1-6
1.1 Introduction	1
1.2 Motivation	3
1.3 Rationale of the Study	3
1.4 Objectives	4
1.5 Research Questions	4
1.6 Expected Output	5
1.7 Report Layout	5
CHAPTER 2: BACKGROUND	7-8
2.1 Introduction	7
2.2 Related Works	7
2.3 Research Summary	8
2.4 Scope of the Problem	8
2.5 Challenges	8
CHAPTER 3: REQUIREMENT SPECIFICATION	9-17

3.1 Introduction	9
3.2 Research Subject and Instrumentation	10
3.3 Data Collection Procedure	11
3.3.1 Dataset	11
3.3.2 Varieties of mangos	12
3.3.3 Data Preprocessing	13
3.3.4 Data Organizing	14
3.3.5 Data Storing	14
3.3.6 Machine Learning Algorithms	14
3.4 Statistical Analysis	15
3.5 Implementation Requirements	17
CHAPTER 4: EXPERIMENTAL RESULT AND DISCUSSION	18-29
4.1 Experiment Setup	18
4.2 Model Summary	18
4.3 Experimental Result and Analysis	18
4.4 Discussion	29
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY	30
5.1 Impact on Society	30
5.2 Impact on Environment	30
5.3 Ethical Aspects	30
5.4 Sustainability Plan	30
CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION, IMPLICATION FOR FUTURE RESEARCH	31-33
6.1 Summary of the Study	31
6.2 Conclusion	31
6.3 Implication for Further Study	31

REFERENCES	32
APPENDIX	33

LIST OF FIGURES

vii

FIGURES	PAGE NO.
Figure 3.1: Methodology	09
Figure 3.2: Structure of the Dataset	10
Figure 3.3: Mangos data during collected	11
Figure 3.4: Datasets Percentages	11
Figure 3.5: Percentages of mango class	12
Figure 3.6: Data Preprocessing	13
Figure 3.6: Train and test data	14
Figure 3.8: Flow chart	16
Figure 4.1: InceptionV3 confusion matrix & model accuracy	20
Figure 4.2: CNN model accuracy	22
Figure 4.3: Predicting on the test dataset	23
Figure 4.4: Test dataset output	23
Figure 4.5: Vgg19 model accuracy	26
Figure 4.6: MobileNet Confusion matrix & model accuracy	28

LIST OF TABLES

TABLE NAME	PAGE NO
Table 1.1 : Species Of The Mango	1
Table 1.2 :Report Layout	5
Table 3.1: Varieties of mangoes	12
Table 4.1(a): Accuracy Table(Train Dataset)	19
Table4.1(b):AccuracyTable(TestDataset)	19

CHAPTER 1

INTRODUCTION

1.1 Introduction

Bangladesh is an agricultural country and approximately 75-80% of the population depends on agriculture. It provides a livelihood for a significant population in Bangladesh and according to the World Bank, almost half of all workers in Bangladesh and two-thirds in rural areas are directly connected with agriculture [12]. Our country has six seasons, But mango is only available in the summer season. Mango is a natural fruit also mango is the most favourite fruit. In our country mango is the king of the fruits. In a Summer Lalpur, Bagha Putia Natore of Rajshahi district. The species of the mango is shown in table 1.1.

- Species of the mango

Table 1.1: Species Of The Mango

Name Of The Mango	Description
Fozli	<ul style="list-style-type: none">• The 'Fazli' mango species is mostly farmed in West Bengal, India, and Bangladesh. It matures slowly and is only available after other types. In Indian cuisine, fazlis are frequently used during jellies and jams [6].
.Harivanga	<ul style="list-style-type: none">• In Bangladesh's northwest, particularly in the Rangpur area, the Haribhanga mango variety is grown. These round mangoes are known as Haribhanga in the local language. Haribhanga normally 200 to 400 grams and is quite meaty[7].

<p>Khirshapat</p>	<ul style="list-style-type: none"> ● Popular mango cultivar, the Khirshapat Mango, was developed in the Indian state of West Bengal and contemporary Bangladesh. It also goes by the name Himsagar. Khirshapat has a golden to orange interior and no fiber, it is widely regarded as the best mango. The fruit is medium in size and weighs between 250 and 350 grams; approximately 77% of that weight is made up of pulp[8].
<p>Langra</p>	<ul style="list-style-type: none"> ● The Banarasi Langra, commonly referred to as the "Langra" mango, is a variety of mango that is mostly grown in Bangladesh, Varanasi, and Banaras in Northern India. The term "Langra" mango is also used to refer to the West Bengal town of Malda in several regions of northern India, including Bihar[9].
<p>Lokhna</p>	<ul style="list-style-type: none"> ● A kind of mango from West Bengal, India is called Lakshmanabhog. In the Maldah area, it is grown in gardens that span about 32,000 hectares. Lakshmanabhog was chosen by the West Bengal government as an export variety to the USA [11]. Bangladeshi ordinary people called this mango name lokhna.
<p>Surjodim</p>	<ul style="list-style-type: none"> ● Mango Sun Egg or Miyazaki A Japanese mango is called mango. In the international market, it is referred to as "Red Mango," while people in Bangladesh are familiar with surjodim, which is currently being grown commercially by Noel Farms[10].

1.2 Motivation

Among the different varieties of fruits, mango is very popular in our country. We know that there are different classifications of mangoes. And based on this classification, the taste of mango can vary. Our main aim is to determine the classification of mangoes Because there are different classifications. Many times people make mistakes with the correct diagnosis. People are deceived due to unscrupulous traders and cannot identify the right mango. Our work makes it easy for the public to determine the classification of mangoes.

1.3 Rationale of the Study

AI (ML), which is regarded as a subset of Artificial Intelligence (AI), is a CSE field that facilitates information extraction based on example recognition. A computer learned from mistakes made in the past after redoing information analysis and expert tasks that were recently deemed too difficult for robots to quantify

The development of this framework to decode data and information in employment area despondency examination has provided crucial data to explore in a matter of collaboration, structure, and system of reasons for despondency in the occupation are both male and female individuals in Bangladesh.

Clinical neurosciences are currently implementing AI (ML) frameworks to provide imaging-based analytic and characterization frameworks for specific neurological and mental health issues. This essay examines the modern work market.

Among the different varieties of fruits, mango is very popular in our country. We know that there are different classifications of mangoes. And based on this classification, the taste of mango can vary. Our main aim is to determine the classification of mangoes Because there are different classifications. Many times people make mistakes with the correct diagnosis. People are deceived due to unscrupulous traders and cannot identify the right mango. Our work makes it easy for the public to determine the classification of mangoes.

1.4 Objective

- To measure the accuracy and classification of mango from different mango species image resources, there is six different types of mango will be analyzed using Inception V3, VGG19 , MobileNet , Convolutional Neural Networks (CNN) .
- To speed up the process of detection.
- A common development framework that developer may use to develop apps that can identify the six different kinds of mango.

1.5 Research Question

1. Can it give the correct output from the given sample data with the system?
2. What is the purpose of this thesis?
3. How was all the data collected?
4. Did each algorithm work correctly?
5. How does precision work?

Yes, it can. The dataset is properly collected and all data is properly processed.

The dataset has been properly collected and all data has been properly processed.

The main objective of the thesis is to correctly determine the classification of mangoes and protect them from fraud. Ensuring fresh mangoes of the right variety.

All the data were collected as raw data and data were collected from six mango orchards in Rajshahi.

Yes.

Four algorithms have been applied here Transfer Learning V3, VGG19, MobileNet , Convolutional Neural Networks (CNN) . We got a good output but the rest of the algorithms couldn't show more accuracy.

1.6 Expected Outcome

Using the provided dataset and our Machine Learning Approach to Classify Mango engine, we may produce an anticipated result. In this case, we used 81% of the dataset as training data and 19% as test data. Our aim is to establish the accuracy of mango classification resources, which depend on the training dataset. Finding the training dataset's correctness is our goal. Once all necessary steps have been done, our classification of mango detection will be complete. Our system achieved Transfer Learning Inception V3 95%, Convolutional Neural Networks (CNN) 95%.

1.7 Report Layout

The Report Layout is shown in table 1.2

Table 1.2: Report Layout

Chapter	Discussion
Chapter 1	This part has covered the motivation for our endeavour, our goals, and the expected outcome of our work.
Chapter 2	In this section, we've discussed the theoretical underpinnings of our research, other connected papers, comparable investigations, the scope of the problems, and challenges.

Chapter 3	We are discussing the topic of our study, the tool we used, how we collected the data, how we analyzed it statistically, and how we put it all into practice.
Chapter 4	The results of our study experiments, analysis, and findings are presented in this chapter.
Chapter 5	We discuss a summary of our findings and predictions in this chapter, and we also include some additional research methods.

CHAPTER 2

BACKGROUND

2.1 Introduction

The Mango is the king of fruit. Most people like fruit. As mango is an occasional natural fruit, it is gathered from the tree in that season and transported to different locations. Thousands of people in this country who do not know about this knowledge can get more information from us. As a result, everyone is able to get knowledge easily. Our country Bangladesh has among the 21 varieties of mango in our country 13. We take the Top 6 mangoes for our project. Top 10 varieties of mango are am rupali, Khirsha paat, lengra, lokkhon vog (Lokhna), Misri dana, Rani posondo.

2.2 Related Works

The author [1], uses a transfer learning method to determine the classification of fruits and vegetables. Convolutional Neural Network was used to produce the classification of fruits and vegetables.

According to the author [2], they are Plant identification using deep neural networks. Transfer learning was used for Plant identification.

According to the author, [3] they are classifying Iris Flower Species Identification. The major purpose of this paper is detection and classification. Their system was image processing and computer vision.

According to the author, [4] they are recognition of medicinal plant species. The major purpose of this paper recognition of medicinal plant species using convolution neural networks and transfer learning.

According to the author, [5] they are classifying citrus diseases. The major purpose of this paper is the detection and classification of plant diseases. Their system was image processing and computer vision.

According to the author, [14] they are classifying Mango Using Statistical Feature and SVM. The major purpose of this paper is the detection and classification of mango. Their system was image processing and computer vision

2.3 Research Summary

We are making predictions on mango classification. Since our project is image dataset-based, we divided the dataset into six main categories such as Fozli, Harivanga, Khishapat, Lengra, Lokna, SurjoDim. From the dataset we will be going to find out the classification of mango InceptionV3, VGG19, CNN, and MobileNet algorithms to find out the accuracy, classification and prediction of the mango dataset. We tried to get as much as high accuracy because the higher the accuracy the better for our system. The classification is a main part of our project. Classification means detecting which groups are according to their observed similarities and differences. And finally prediction and predicting class will help us to detect the mango.

2.4 Scope of the Problem

The main reason we work on mango classification is to determine the classification of mangoes. Ordinary people often misunderstand the classification of mangoes. If we don't act on it then it will be seen that people are being cheated more and more day by day. Despite mango being such a popular fruit, people misunderstand the classification. Because of this, dishonest dealers defraud regular people in several ways. It is still very difficult for the average person to diagnose me with ordinary eyes.

2.5 Challenges

It is extremely difficult to collect information about mango data. Because we had such a lot of issues, there is not a single source for the data set. It's seasonal fruits, if the summer season is gone then data collection is tough for us. There are many species of mango we can not collect data for the season. We made an effort to gather photos with similar colours and brightness. It was really difficult to take a good photo.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

We will go over the study approach and procedures in more detail in this section. Contrarily, the project's tools, data collecting, study topic, processing, and pre-processing tools. This chapter will cover statistical analysis and its application. The figure displays the entire technique 3.1.

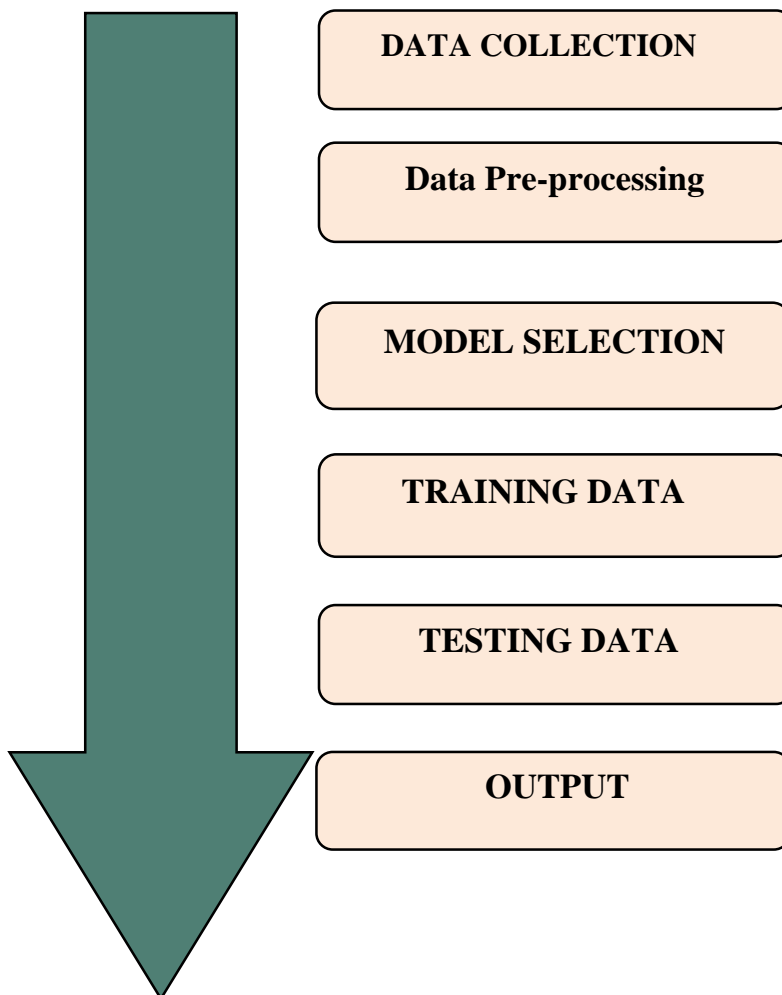


Figure 3.1: Methodology

3.2 Research Subject and Instrumentation

We discovered that categorizing mangoes is a challenging undertaking. We have to exercise caution on something.

- What dataset needs to be gathered?
- Making sure the collection was flawless.
- Properly organizing the data.
- Correctly labelling the dataset.

3.3 Data Collection Procedure

In the realm of study, the role of data collection is crucial. In order to gain greater accuracy, it is crucial to collect adequate and trustworthy data. Six types of real-time data were gathered. We have collected data from 2022 June to 2022 July Lalpur, Bagha Putia Natore of Rajshahi district. We used Vivo Y20 for a better image. In our dataset, there are almost six types of mangos.

```
num_of_train_samples = 0
for train_dataset in os.listdir(train):
    in_folder = train + "/" + train_dataset
    in_folder_list = os.listdir(in_folder)
    num_of_train_samples = num_of_train_samples + len(in_folder_list)
print("Number of Training samples : ", num_of_train_samples)

# Total number of validation images
num_of_validation_samples = 0
for validation_dataset in os.listdir(val):
    in_folder_val = val + "/" + validation_dataset
    in_folder_val_list = os.listdir(in_folder_val)
    num_of_validation_samples = num_of_validation_samples + len(in_folder_val_list)
print("Number of Validation samples : ", num_of_validation_samples)
```

```
Number of Training samples : 1099
Number of Validation samples : 255
```

Figure 3.2: Structure of the dataset



Figure 3.3: Mangos data during collected

3.3.1 Data set

We have about 1354 data from 6 different varieties of mangoes in the dataset. Langra-118, Surjodim-191, Fozli-111, Khishapat-402, Lokna-335, and Harivanga-197 are located in our datasets, which are represented in the figure datasets percentages 3.4.

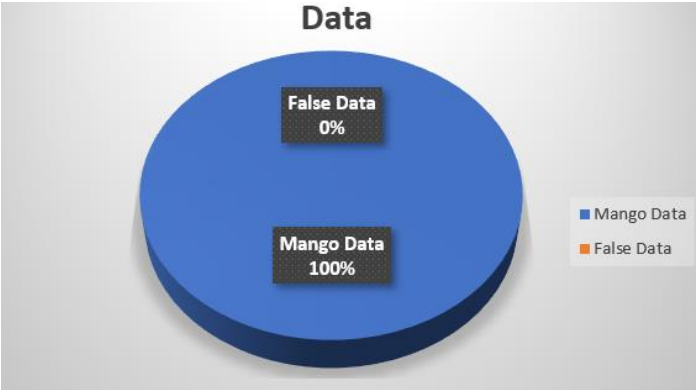


Figure 3.4: Datasets percentages

3.3.2 Varieties of mangoes

Here we have six varieties of mangoes

- Langra
- Surjodim
- Fozli
- Khirapat
- Lokna
- Harivanga

The number of datasets that are available for each class is listed below:

Table 3.1 Varieties of mangoes

Category	Count
Langra	118
Surjodim	191
Fozli	111
Khisapat	402
Lokna	335
Harivanga	197

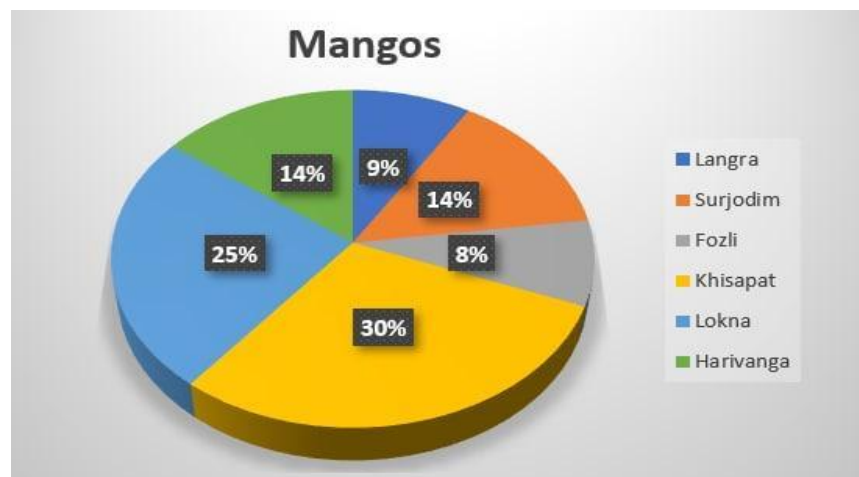


Figure 3.5: Percentages of mangos class

Each mango class's percentage in the datasets is Langra-9%, Surjodim-14%, Fozli-8%, Khishapat-30%, Lokna-25% and Harivanga-14%.

3.3.3 Data Preprocessing

We pre-processed all the data in this area. because there are several methods of image regulation in the datasets. The accuracy rate will decrease if all photos are not transformed into a single set of rules. after processing, the image has been converted to 224*244 pixels shown in figure 3.6.

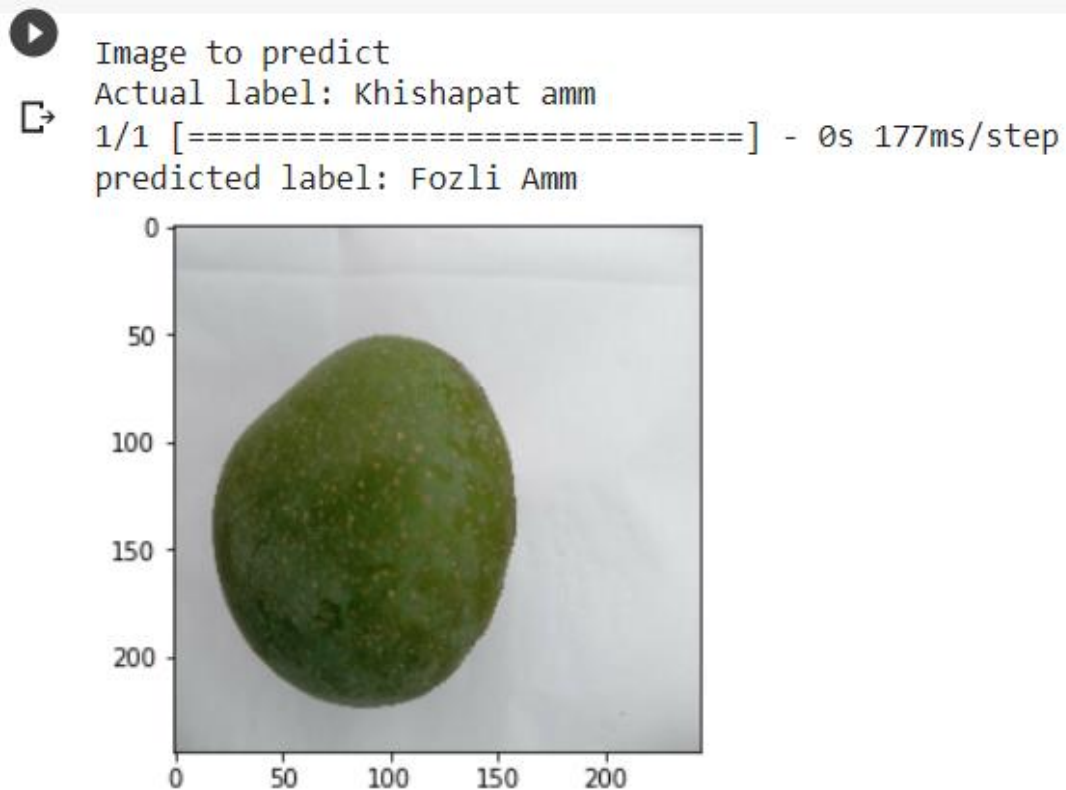


Figure 3.6 Data preprocessing

3.3.4 Data Organizing

We have two folders one for the training dataset and the other for the test dataset photos are included in the training dataset The training dataset had (1354) photos and three (6) classes. For each class, there are (1099) photos in the training dataset. On the other hand, the test dataset contains 255 photos divided into three (6) groups. Our data ratio train data 81% And test data 19%.

```
num_of_train_samples = 0
for train_dataset in os.listdir(train):
    in_folder = train + "/" + train_dataset
    in_folder_list = os.listdir(in_folder)
    num_of_train_samples = num_of_train_samples + len(in_folder_list)
print("Number of Training samples : ", num_of_train_samples)

# Total number of validation images
num_of_validation_samples = 0
for validation_dataset in os.listdir(val):
    in_folder_val = val + "/" + validation_dataset
    in_folder_val_list = os.listdir(in_folder_val)
    num_of_validation_samples = num_of_validation_samples + len(in_folder_val_list)
print("Number of Validation samples : ", num_of_validation_samples)
```

```
Number of Training samples : 1099
Number of Validation samples : 255
```

Figure 3.7: Train and test dataset

3.3.5 Data Storing

Following data organization, the process of data storing has begun. In this step, the dataset was kept on both a local PC directory and a Google Drive account. We had to utilize the local PC directory because we were using the Jupyter notebook. We also used Google Colab for security. We also uploaded the datasets to Google Drive as a result. And uploading to Google Drive was the safest option if we needed to use those datasets in the future to improve the product. All the datasets would be lost if there was a technical issue with the local computer. Because of this, we also uploaded the dataset to Google Drive.

3.3.6 Machine Learning Algorithms

For more accuracy, we have employed INCEPTION V3, MobileNet, VGG19, and CNN. We used training and test datasets to apply these methods. Python libraries like TensorFlow, Keras, NumPy, PyTorch, Flatten, etc. are used to implement all algorithms. These libraries aid in the program's initialization.

3.4 Statistical Analysis

1354 photos are available in the dataset for mango classification and mango detection. Nearly 1099 photos make up the training dataset, while the remaining ones make up the test dataset. We used numerous algorithms, but improving accuracy was our major goal VGG19, and INCEPTION V3 provided the accuracy we needed. To determine the required precision, we continued to follow some processes. Like this

We captured real-time images of different items

- All the unwanted photos were deleted after formatting.
- Next, all of the photo's heights and widths were set to 224*224.
- Classifying each dataset with a label.
- We divided them into train and test datasets after labelling.
- Algorithms were employed following partitioning.
 - VGG19
 - INCEPTIONV3
 - CNN
 - MobileNet

• Then finally accuracy analysis.

- The Flowchart of our working process

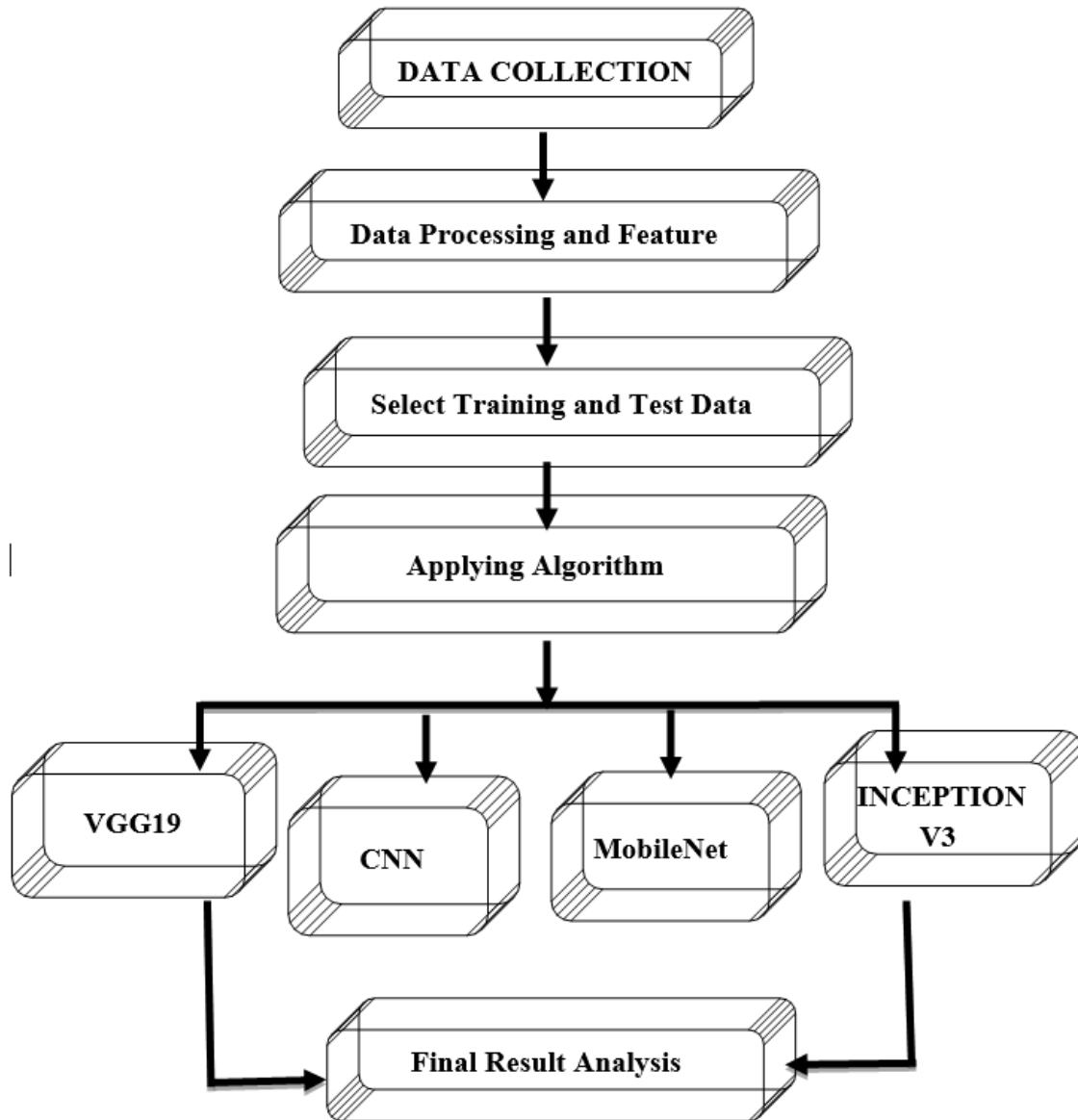


Figure 3.8: Flow chart

3.5 Implementation Requirements

- **Python 3.10.2**

Use Python 3.10.2. Python is a High-Level programming language that is open-source, simple to use, and touted as being the strongest of all computer languages. It will be very challenging to use another language because the architecture of VGG19, and inceptionV3 is so intricate. Because Python's built-in functions and commands are so simple, implementing them takes less time. It would take a long time in other programming languages that would have been utilized.

- **Google Collab/Jupyter Notebook**

Colab is a hosted Jupiter, installed so we don't have to do anything its already set up and configured such that all work is done through the Chrome browser and using cloud services, Google Colab is a virtual application. we can operate easily because it is the same as Jupiter. They are based on notebooks or notebooks, whether they be text, image, or code Jupyter Notebook also does the same thing but it works on a hardware base. It has to be installed on your system. If anyone has a high-end build computer with an external Graphics card then it's perfect for them.

Device Requirements

- Version Of Operating System (Windows 7 or more / Linux)
- Web Browser (Chrome, Firefox, or Microsoft Edge)
- Hard Disk (At least 120GB)
- Ram (More than 4 GB)
- GPU (At least 2GB)

CHAPTER 4

EXPERIMENTAL RESULT AND DISCUSSION

4.1 Experimental Setup

The initial objective was to compile datasets for our model and code. The following system process is provided:

- We gathered all the statistics from Rajshahi since our objective is to classify the six different types of mangoes.
- The hardest step was gathering all the data. Due to the time-consuming nature of collecting datasets.

4.2 Model Summary

Because this was an image analysis, the Overview of the Method was really quite difficult. The maximum, trainable, and non-trainable parameters for each algorithm can be found here. Finally, the model is applied to determine how the model is created. The dataset seems to have six dense layers and is compressed. We were using 10 to 70 epochs, each with 20 layers.

4.3 Experimental Result and Analysis

Since determining the classification of different varieties of mangoes is our primary objective. Then, using test datasets, we forecast our model. The accuracy rate is crucial, and we're happy with what we obtained in terms of accuracy. Additionally, testing our model proved successful. At InceptionV3 and CNN, we have 94% &90% accuracies. The training dataset's accuracy is shown in Table 4.1. The testing dataset's accuracy is shown in table 4.2.

Table 4.1 Accuracy Table

Model Name	Accuracy (Train datasets)
InceptionV3	94%
VGG19	93%
MobileNet	99%
CNN	99%

Table 4.2 Accuracy Table


Model Name	Accuracy (Test datasets)
InceptionV3	93%
VGG19	74%
MobileNet	49%
CNN	90%

The Model of InceptionV3 accuracy

```

Y_pred_inc = inc_model.predict_generator(validation_generator, num_of_validation_samples // batch_size+1)
y_pred_inc = np.argmax(Y_pred_inc, axis=1)
print('Confusion Matrix')
conf_matrix_inc = confusion_matrix(validation_generator.classes, y_pred_inc)
cm_inc = np.array2string(conf_matrix_inc)
print(conf_matrix_inc)
print("=====")
print('Classification Report')
target_names = ['Lokna amm', 'Surjo dim amm', 'lengra amm', 'Khishapat amm', 'Harivanga amm', 'Fozli Amm']
class_rep_inc = classification_report(validation_generator.classes, y_pred_inc, target_names=target_names)
print(class_rep_inc)

```

 <ipython-input-20-b24903156672>:1: UserWarning: `Model.predict_generator` is deprecated and will be removed in a future version. P]

```

Y_pred_inc = inc_model.predict_generator(validation_generator, num_of_validation_samples // batch_size+1)
Confusion Matrix
[[36  0  0  3  0  0]
 [ 0 39  0  0  2  0]
 [ 0  2 36  4  0  0]
 [ 0  0  0 45  0  1]
 [ 0  2  0  0 39  0]
 [ 0  0  0  0  0 46]]

```

```

print("=====")
print('Classification Report')
target_names = ['Lokna amm', 'Surjo dim amm', 'lengra amm', 'Khishapat amm', 'Harivanga amm', 'Fozli Amm']
class_rep_inc = classification_report(validation_generator.classes, y_pred_inc, target_names=target_names)
print(class_rep_inc)

```

↳ <ipython-input-20-b24903156672>:1: UserWarning: `Model.predict_generator` is deprecated and will be removed in a fut
Y_pred_inc = inc_model.predict_generator(validation_generator, num_of_validation_samples // batch_size+1)

Confusion Matrix

```

[[36  0  0  3  0  0]
 [ 0 39  0  0  2  0]
 [ 0  2 36  4  0  0]
 [ 0  0  0 45  0  1]
 [ 0  2  0  0 39  0]
 [ 0  0  0  0  0 46]]

```

=====
Classification Report

	precision	recall	f1-score	support
Lokna amm	1.00	0.92	0.96	39
Surjo dim amm	0.91	0.95	0.93	41
lengra amm	1.00	0.86	0.92	42
Khishapat amm	0.87	0.98	0.92	46
Harivanga amm	0.95	0.95	0.95	41
Fozli Amm	0.98	1.00	0.99	46
accuracy			0.95	255
macro avg	0.95	0.94	0.95	255
weighted avg	0.95	0.95	0.95	255

```

training_accuracy_inc = inc_training.history['accuracy'][-1]
training_loss_inc     = inc_training.history['loss'][-1]
validation_accuracy_inc = inc_training.history['val_accuracy'][-1]
validation_loss_inc   = inc_training.history['val_loss'][-1]
print("Training Accuracy Inception :", training_accuracy_inc )
print("Training Loss Inception      :", training_loss_inc)
print("Validation Accuracy Inception :", validation_accuracy_inc)
print("Validation Loss Inception     :", validation_loss_inc)

```

```

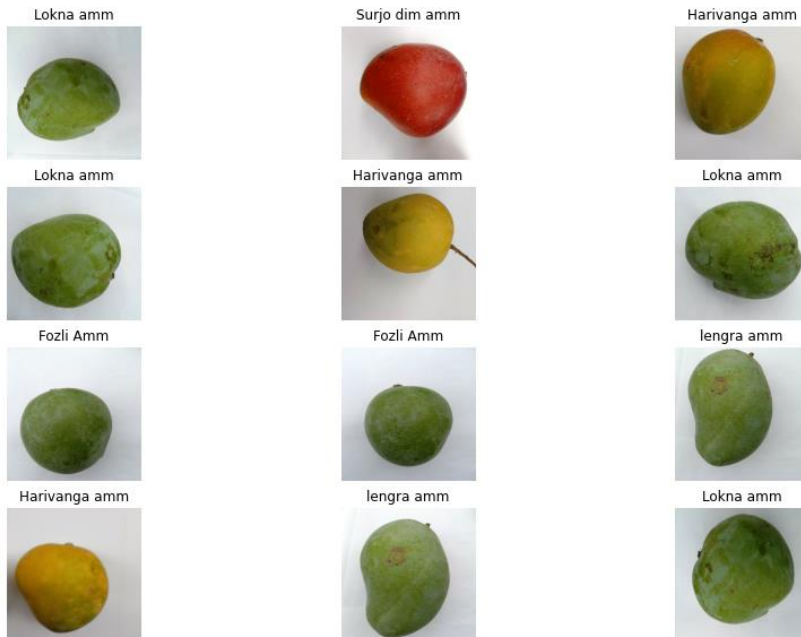
↳ Training Accuracy Inception : 0.9437675476074219
Training Loss Inception       : 0.20261017978191376
Validation Accuracy Inception : 0.9375
Validation Loss Inception     : 0.18989339470863342

```

Figure 4.1: InceptionV3 confusion matrix & model accuracy

The Model of CNN Accuracy

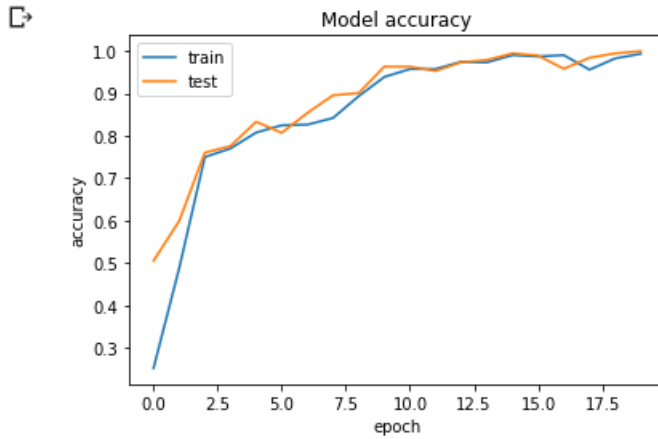
```
import matplotlib.pyplot as plt
plt.figure(figsize=(15,10))
for image, label in dataset.take(1):
    for count in range(12):
        ax = plt.subplot(4,3,count+1)
        plt.imshow(image[count].numpy().astype("uint8"))
        plt.title(class_names[label[count]])
        plt.axis("off")
plt.show()
```



```

▶ plt.plot(model_history.history["accuracy"])
  plt.plot(model_history.history["val_accuracy"])
  plt.title('Model accuracy')
  plt.ylabel("accuracy")
  plt.xlabel("epoch")
  plt.legend(['train', 'test'], loc='upper left')
  plt.show()

```



```

▶ plt.plot(model_history.history["loss"])
  plt.plot(model_history.history["val_loss"])
  plt.title('Model accuracy')
  plt.ylabel("accuracy")
  plt.xlabel("epoch")
  plt.legend(['train', 'test'], loc='upper left')
  plt.show()

```

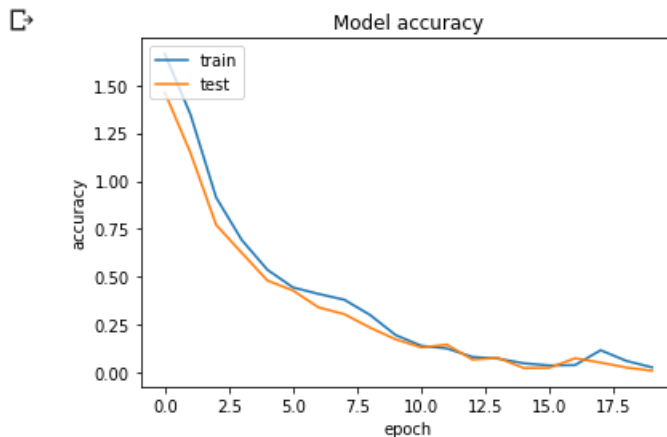


Figure 4.2: Model accuracy

Identification: Using our method, we want to determine using testing data. Based on the results of the evaluation, we will choose the mango. The next image shows the necessary approach. 4.3

```
[44] plt.figure(figsize=(15,15))
17a for images, labels in test_ds.take(1):
    for i in range(9):
        ax = plt.subplot(3, 3, i+1)
        plt.imshow(images[i].numpy().astype("uint8"))

        predicted_class, confidence = predict_diseases(model_1, images[i].numpy())
        actual_class = class_names[labels[i]]

        plt.title(f"Actual class: {actual_class},\n Predicted: {predicted_class}.\n Confidence: {confidence}%")
        plt.axis("off")

1/1 [=====] - 0s 45ms/step
1/1 [=====] - 0s 15ms/step
1/1 [=====] - 0s 14ms/step
1/1 [=====] - 0s 15ms/step
1/1 [=====] - 0s 15ms/step
1/1 [=====] - 0s 16ms/step
1/1 [=====] - 0s 17ms/step
1/1 [=====] - 0s 14ms/step
1/1 [=====] - 0s 16ms/step
```

Figure 4.3: Predicting on the test dataset

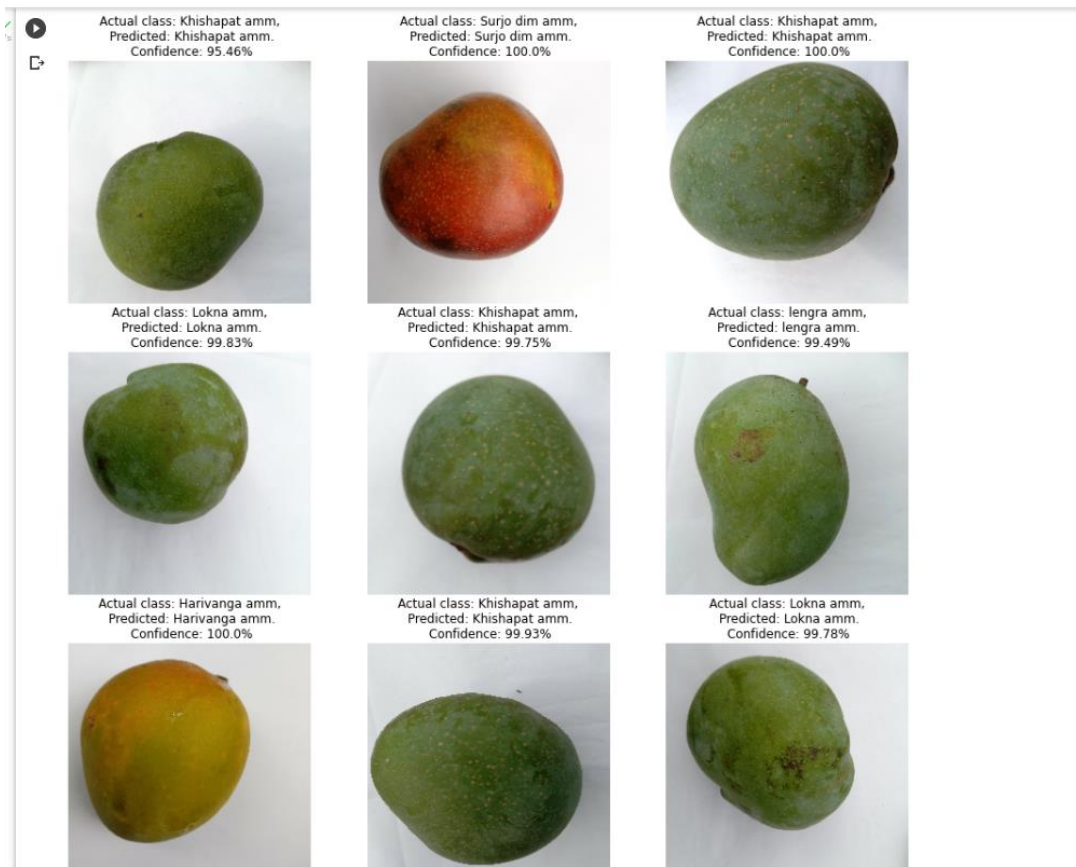


Figure 4.3: Test dataset output

The Model of VGG19 Accuracy

```

init_time = datetime.datetime.now()
learning_rate_reduction = ReduceLRonPlateau(monitor='val_accuracy', patience = 1, verbose=1, factor=0.3, min_lr=0.000001)

trained = model.fit(
    train_images, train_labels,
    validation_data = (val_images, val_labels),
    batch_size = BATCH_SIZE,
    epochs=EPOCHS,
    callbacks=[learning_rate_reduction],
)

required_time = datetime.datetime.now() - init_time
print(f'\nRequired time: {str(required_time)}\n')

```

```

Epoch 1/10 [=====] - 21s 101ms/step - loss: 0.6144 - accuracy: 0.7528 - val_loss: 1.0108 - val_accuracy: 0.4375 - lr: 0.00
Epoch 2/10 [=====] - ETA: 0s - loss: 0.3082 - accuracy: 0.8446
Epoch 2: ReduceLRonPlateau reducing learning rate to 0.0003000000142492354.
Epoch 3/10 [=====] - 11s 100ms/step - loss: 0.3082 - accuracy: 0.8446 - val_loss: 0.8743 - val_accuracy: 0.4375 - lr: 0.00
Epoch 4/10 [=====] - 11s 102ms/step - loss: 0.2574 - accuracy: 0.8639 - val_loss: 0.8617 - val_accuracy: 0.7422 - lr: 3.00
Epoch 5/10 [=====] - ETA: 0s - loss: 0.2503 - accuracy: 0.9331

```

```

Epoch 6: ReduceLRonPlateau reducing learning rate to 2.700000040931627e-05.
Epoch 6/10 [=====] - 11s 98ms/step - loss: 0.2407 - accuracy: 0.9328 - val_loss: 0.8404 - val_accuracy: 0.7422 - lr: 9.000e-06
Epoch 7/10 [=====] - ETA: 0s - loss: 0.2383 - accuracy: 0.9334
Epoch 7: ReduceLRonPlateau reducing learning rate to 8.100000013655517e-06.
Epoch 8/10 [=====] - 11s 98ms/step - loss: 0.2383 - accuracy: 0.9334 - val_loss: 0.8390 - val_accuracy: 0.7422 - lr: 2.700e-06
Epoch 9/10 [=====] - ETA: 0s - loss: 0.2331 - accuracy: 0.9375
Epoch 9: ReduceLRonPlateau reducing learning rate to 2.429999949526973e-06.
Epoch 10/10 [=====] - 11s 100ms/step - loss: 0.2331 - accuracy: 0.9375 - val_loss: 0.8372 - val_accuracy: 0.7422 - lr: 8.100e-07
Epoch 11/10 [=====] - ETA: 0s - loss: 0.2372 - accuracy: 0.9320
Epoch 11: ReduceLRonPlateau reducing learning rate to 1e-06.
Epoch 12/10 [=====] - 11s 100ms/step - loss: 0.2372 - accuracy: 0.9320 - val_loss: 0.8370 - val_accuracy: 0.7422 - lr: 2.43e-07
Epoch 13/10 [=====] - 11s 100ms/step - loss: 0.2408 - accuracy: 0.9328 - val_loss: 0.8380 - val_accuracy: 0.7422 - lr: 1.00e-07
Epoch 14/10 [=====] - 11s 98ms/step - loss: 0.2373 - accuracy: 0.9303 - val_loss: 0.8376 - val_accuracy: 0.7500 - lr: 1.000e-07

Required time: 0:02:02.834030

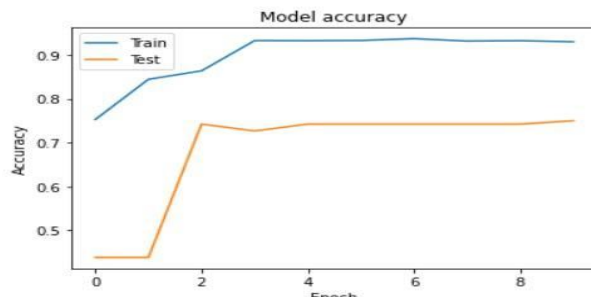
```

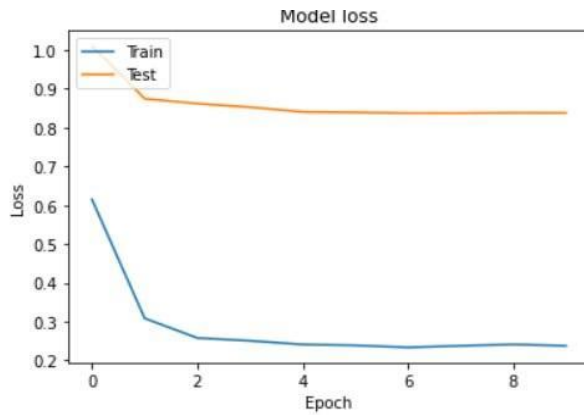
```

[ ] plt.plot(trained.history['accuracy'])
plt.plot(trained.history['val_accuracy'])
plt.title('Model accuracy')
plt.ylabel('Accuracy')
plt.xlabel('Epoch')
plt.legend(['Train', 'Test'], loc='upper left')
plt.show()

plt.plot(trained.history['loss'])
plt.plot(trained.history['val_loss'])
plt.title('Model loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(['Train', 'Test'], loc='upper left')
plt.show()

```



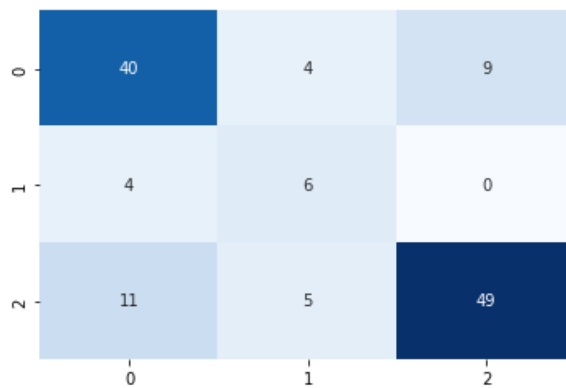


```
[ ] train_loss, train_acc = model.evaluate(train_images, train_labels, verbose=0)
print('\naccuracy:', train_acc, ' loss: ', train_loss)
```

accuracy: 0.9452433586120605 loss: 0.22229105234146118

```
print(classification_report(test_labels, predict), '\n')
cm = confusion_matrix(test_labels, predict)
sns.heatmap(cm, annot=True, cmap="Blues", fmt='.0f', cbar=False)
plt.show()
```

	precision	recall	f1-score	support
0	0.73	0.75	0.74	53
1	0.40	0.60	0.48	10
2	0.84	0.75	0.80	65
accuracy			0.74	128
macro avg	0.66	0.70	0.67	128
weighted avg	0.76	0.74	0.75	128



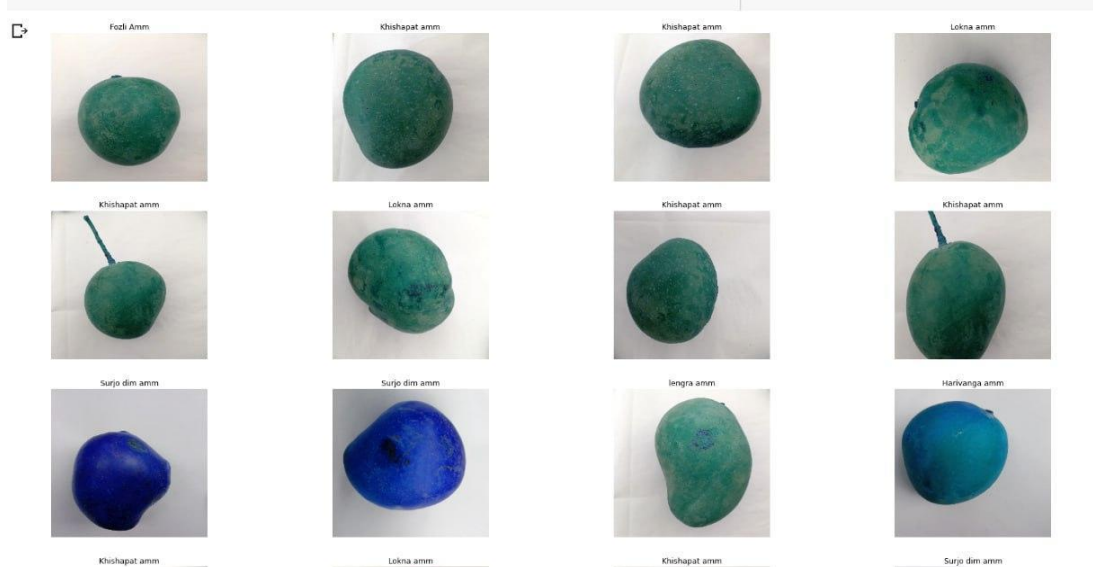


Figure 4.3: Vgg19 model accuracy

The Model of Mobilenet Accuracy

```

EPOCHS = 20
hist = model.fit(xtrain, ytrain, epochs=EPOCHS, callbacks=[lr_callbacks], validation_data=(xtest,ytest))

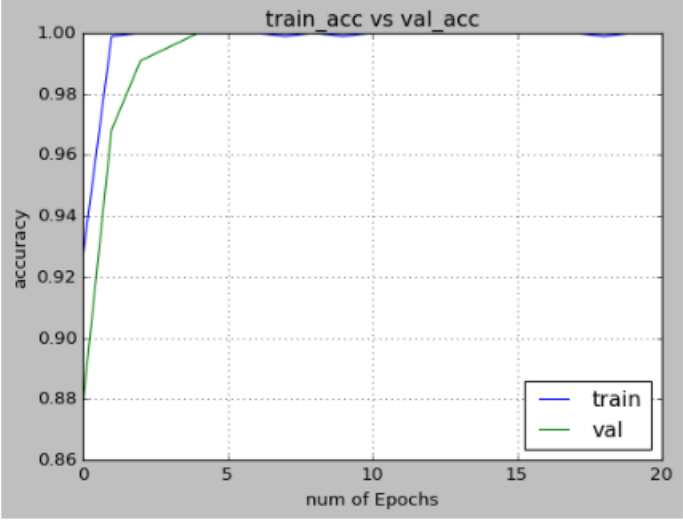
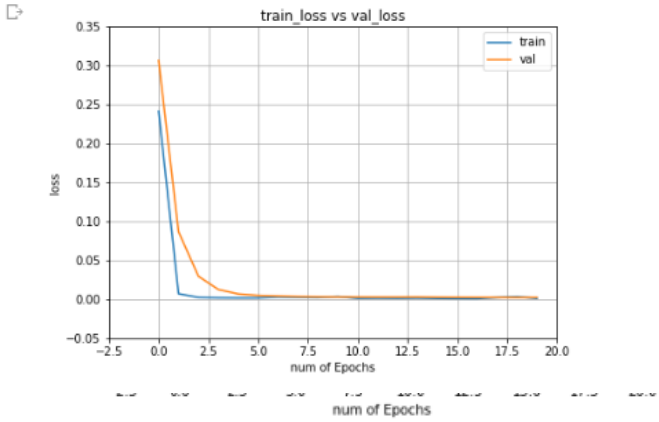
Epoch 1/20
28/28 [=====] - 13s 110ms/step - loss: 0.2413 - accuracy: 0.9261 - val_loss: 0.3064 - val_accuracy: 0.8773 - lr: 0.0010
Epoch 2/20
28/28 [=====] - 2s 55ms/step - loss: 0.0069 - accuracy: 0.9989 - val_loss: 0.0865 - val_accuracy: 0.9682 - lr: 0.0010
Epoch 3/20
28/28 [=====] - 2s 56ms/step - loss: 0.0022 - accuracy: 1.0000 - val_loss: 0.0293 - val_accuracy: 0.9909 - lr: 0.0010
Epoch 4/20
28/28 [=====] - 2s 56ms/step - loss: 0.0018 - accuracy: 1.0000 - val_loss: 0.0123 - val_accuracy: 0.9955 - lr: 1.0000e-04
Epoch 5/20
28/28 [=====] - 2s 56ms/step - loss: 0.0017 - accuracy: 1.0000 - val_loss: 0.0066 - val_accuracy: 1.0000 - lr: 1.0000e-04
Epoch 6/20
28/28 [=====] - 2s 56ms/step - loss: 0.0016 - accuracy: 1.0000 - val_loss: 0.0047 - val_accuracy: 1.0000 - lr: 1.0000e-04
Epoch 7/20
28/28 [=====] - 2s 56ms/step - loss: 0.0015 - accuracy: 1.0000 - val_loss: 0.0028 - val_accuracy: 1.0000 - lr: 1.0000e-04
Epoch 8/20
28/28 [=====] - 2s 56ms/step - loss: 0.0014 - accuracy: 1.0000 - val_loss: 0.0023 - val_accuracy: 1.0000 - lr: 1.0000e-04
Epoch 9/20
28/28 [=====] - 2s 55ms/step - loss: 0.0020 - accuracy: 1.0000 - val_loss: 0.0022 - val_accuracy: 1.0000 - lr: 1.0000e-05
Epoch 10/20
28/28 [=====] - 2s 56ms/step - loss: 0.0031 - accuracy: 0.9989 - val_loss: 0.0021 - val_accuracy: 1.0000 - lr: 1.0000e-05
Epoch 11/20
28/28 [=====] - 2s 56ms/step - loss: 9.9288e-04 - accuracy: 1.0000 - val_loss: 0.0022 - val_accuracy: 1.0000 - lr: 1.0000e-05

```

```
epochs = 20
train_loss = hist.history['loss']
val_loss = hist.history['val_loss']
train_acc = hist.history['accuracy']
val_acc = hist.history['val_accuracy']
xc = range(epochs)

plt.figure(1,figsize=(7,5))
plt.plot(xc,train_loss)
plt.plot(xc,val_loss)
plt.xlabel('num of Epochs')
plt.ylabel('loss')
plt.title('train_loss vs val_loss')
plt.grid(True)
plt.legend(['train','val'])
#print plt.style.available # use bmh, classic,ggplot for big pictures
plt.style.use(['classic'])

plt.figure(2,figsize=(7,5))
plt.plot(xc,train_acc)
plt.plot(xc,val_acc)
plt.xlabel('num of Epochs')
plt.ylabel('accuracy')
plt.title('train_acc vs val_acc')
plt.grid(True)
plt.legend(['train','val'],loc=4)
#print plt.style.available # use bmh, classic,ggplot for big pictures
plt.style.use(['classic'])
```



```

) pred_ls = np.argmax(pred_labels, axis=1)
  test_ls = np.argmax(test_image_label, axis=1)

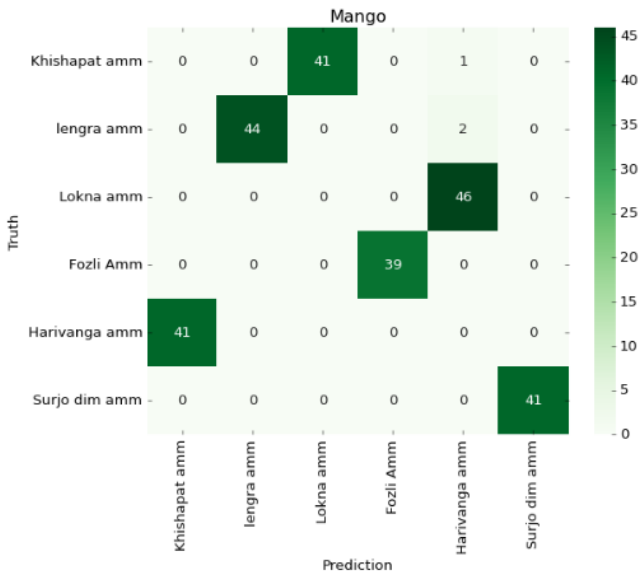
  conf_arr = confusion_matrix(test_ls, pred_ls)

  plt.figure(figsize=(8, 6), dpi=80, facecolor='w', edgecolor='k')

  ax = sns.heatmap(conf_arr, cmap='Greens', annot=True, fmt='d', xticklabels= test_class,
                  yticklabels=test_class)

  plt.title('Mango')
  plt.xlabel('Prediction')
  plt.ylabel('Truth')
  plt.show(ax)

```



```

▶ from sklearn.metrics import classification_report, log_loss, accuracy_score
  y_pred=model.predict(test_images)
  pred=np.argmax(y_pred,axis=1)
  ground = np.argmax(test_image_label,axis=1)
  print(classification_report(ground,pred))

```

```

↳ 8/8 [=====] - 0s 49ms/step
      precision    recall  f1-score   support

0         0.00      0.00      0.00         42
1         1.00      0.96      0.98         46
2         0.00      0.00      0.00         46
3         1.00      1.00      1.00         39
4         0.00      0.00      0.00         41
5         1.00      1.00      1.00         41

   accuracy          0.49         255
  macro avg          0.50          255
 weighted avg          0.49          255

```

Figure 4.4: MobileNet model accuracy

4.4 Discussion

We have developed a model that may be used to identify the sources. But the accuracy rate is what we mostly learn. The success of our future efforts depends on this accuracy rate. Furthermore, accuracy primarily depends on the dataset. Since the research is image-based, the picture quality and clarity may be quite high. And that caused us a great deal of worry. the algorithms, second. A solid algorithm is also essential to accuracy. Since accuracy increases with more parameters. As a result, our accuracy ratio for the training set was up to 97%, while our accuracy ratio for the test set was 90%.

CHAPTER 5

Impact on Society, Environment and Sustainability

5.1 Impact on Society

Despite being the most popular fruit in Bangladesh, the mango has received very little study. Mangos are a traditional fruit. It has a lot of attributes. I learned a lot of useful information about mangoes for my research. It is our society's population. Knowing more about mangoes would be beneficial. They will be aware of the benefits of mangoes. I've dealt with six distinct varieties of mangoes, each of which has unique qualities. The youthful generation of today is ignorant about mango categorization. This research provides a lot of help to future students. I also talked about the nutritional value and quality of my mangoes.

5.2 Impact on Environment

Nowdays, We can see most of urban people device not recognized mango. In earlier times people were all associated with agriculture they can easily recognize. If we don't help them to know then our future generations will does not recognized mangos. Anyone will cheat them.

5.3 Ethical Aspects

India is the country of origin for this delicious, juicy, sweet fruit, which has garnered lovers all over the world. The puja ceremony, one of the most important Hindu ceremonies, employs water laced with mango leaves to generate the right resonance for revered deities. Mangoes are often used by Indian poets to express feelings of passion and love. Additionally, contemporary Indian authors like Arundhati Roy and Anita Desai use mangoes as a metaphor for prosperity, sweetness, and opportunity.

5.4 Sustainability Plan

We are using a machine learning algorithm on Google colab we got the best accuracy 95% from InceptionV3. We are connected to python 3 Goggle compute Engine here we got 108 GB storage and 12 GB ram it is a little bit slow in processing speed. But it is working properly in future we run on a good system it will get the result more faster.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION, IMPLICATION FOR FUTURE RESEARCH

6.1 Summary of the Study

We have built up a model that gives us a proper idea about the Machine Learning Approach to Classify Mango. The main thing was to get ideas about class definitions. For this reason, accuracy was a very important part because If the accuracy becomes low then there will be no benefit as our test depends on the dataset. Our test and prediction depend on the test dataset. So we had to be careful about test dataset accuracy along with train accuracy.

From our four data models, we can easily say InceptionV3 or CNN both give us almost the 94% & 90% accuracy in terms of test and train sets. Almost 90% accuracy we got. For this reason, our test and prediction will be more accurate as well.

6.2 Conclusion

Mango is a popular fruit in our country and is a seasonal fruit that has many classifications. Ordinary people have little knowledge about mango varieties. As a result, most of the time people misunderstand the mango variety. Hustler traders are harming the common man for their benefit. The main purpose of making this find is the mango classification. We are planning to work this side so that ordinary people would not be deceived by these unscrupulous traders and people can recognize the right variety of mango.

6.3 Implication for Further Study

We will work with more data to make our examination more proficient.

- We will collect data as much as possible
- Also will try to add mangos classification.
- Our main goal is to create a web application for the prediction dataset. If any images of mangos can detect the classification.

REFERENCES

- [1]Tomar, Hanshu. Multi-class image classification of fruits and vegetables using transfer learning techniques. Diss. Dublin Business School, 2020.
- [2] Ghazi, Mostafa Mehdipour, Berrin Yanikoglu, and Erchan Aptoula. "Plant identification using deep neural networks via optimization of transfer learning parameters." *Neurocomputing* 235 (2017): 228-235.
- [3]Pinto, Joylin Priya, Soumya Kelur, and Jyothi Shetty. "Iris flower species identification using machine learning approach." 2018 4th International Conference for Convergence in Technology (I2CT). IEEE, 2018.
- [4]Gogul, I., and V. Sathiesh Kumar. "Flower species recognition system using convolution neural networks and transfer learning." 2017 fourth international conference on signal processing, communication and networking (ICSCN). IEEE, 2017.
- [5]Rauf, Hafiz Tayyab, et al. "A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning." *Data in brief* 26 (2019): 104340.
- [6]Wikipedia,[https://en.wikipedia.org/wiki/Fazli_\(mango\)#:~:text=The%20Fazli'%20mango%20is%20a,going%20up%20to%20a%20kilo](https://en.wikipedia.org/wiki/Fazli_(mango)#:~:text=The%20Fazli'%20mango%20is%20a,going%20up%20to%20a%20kilo). Last visit 01-01-2023
- [7]Wikipedia,[https://en.wikipedia.org/wiki/Haribhanga_\(mango\)#:~:text=Haribhanga%20is%20highly%20fleshy%20and,between%20the%20two%20neighbouring%20countries](https://en.wikipedia.org/wiki/Haribhanga_(mango)#:~:text=Haribhanga%20is%20highly%20fleshy%20and,between%20the%20two%20neighbouring%20countries). Last visit 01-01-2023
- [8]Wikipedia,<https://en.wikipedia.org/wiki/Himsagar>. Last visit 01-01-2023
- [9]Wikipedia,<https://en.wikipedia.org/wiki/Langra>. Last visit 01-01-2023
- [10]Wikipedia,https://bn.wikipedia.org/wiki/%E0%A6%B8%E0%A7%82%E0%A6%B0%E0%A7%8D%E0%A6%AF%E0%A6%A1%E0%A6%BF%E0%A6%AE_%E0%A6%86%E0%A6%AE.Last visit 01-01-2023
- [11]<https://en.wikipedia.org/wiki/Lakshmanbhog>.Last visit 01-01-2023
- [12]worldbank,<https://www.worldbank.org/en/results/2016/10/07/bangladesh-growing-economy-through-advances-in-agriculture>.Last visit 01-01-2023
- [13]Unb,Top Mango Varieties in Bangladesh (unb.com.bd).Last Visit 01-05-2023
- [14]<https://bn.quora.com/%E0%A6%AB%E0%A6%B2%E0%A7%87%E0%A6%B0-%E0%A6%B0%E0%A6%BE%E0%A6%9C%E0%A6%BE-%E0%A6%86%E0%A6%AE-%E0%A6%95%E0%A7%87%E0%A6%A8>. Last Visit 20-01-2023

APPENDIX

Building up this project was a difficult task. During the project, we faced a few problems. The first problem was collecting the proper dataset. At first, we collected some data online. But for better accuracy, those images were not appropriate. So we started collecting raw data which was very difficult. We have collected data from 2022 June to 2022 July Lalpur, Bagha Putia Nature of Rajshahi district. The second problem we faced, finding a proper algorithm. We found a CNN algorithm that gave only 50% accuracy. But after some research, we found that the parameters were not appropriate so again we modified the algorithm and then got a proper accuracy.

Machine Learning Approach to Classify Mango

ORIGINALITY REPORT

7 %	7 %	1 %	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	7 %
2	cupdf.com Internet Source	<1 %
3	hdl.handle.net Internet Source	<1 %

Exclude quotes On

Exclude matches Off

Exclude bibliography On