BANGLADESHI FRESH AND ROTTEN VEGETABLES DETECTION USING CNN

Submitted by

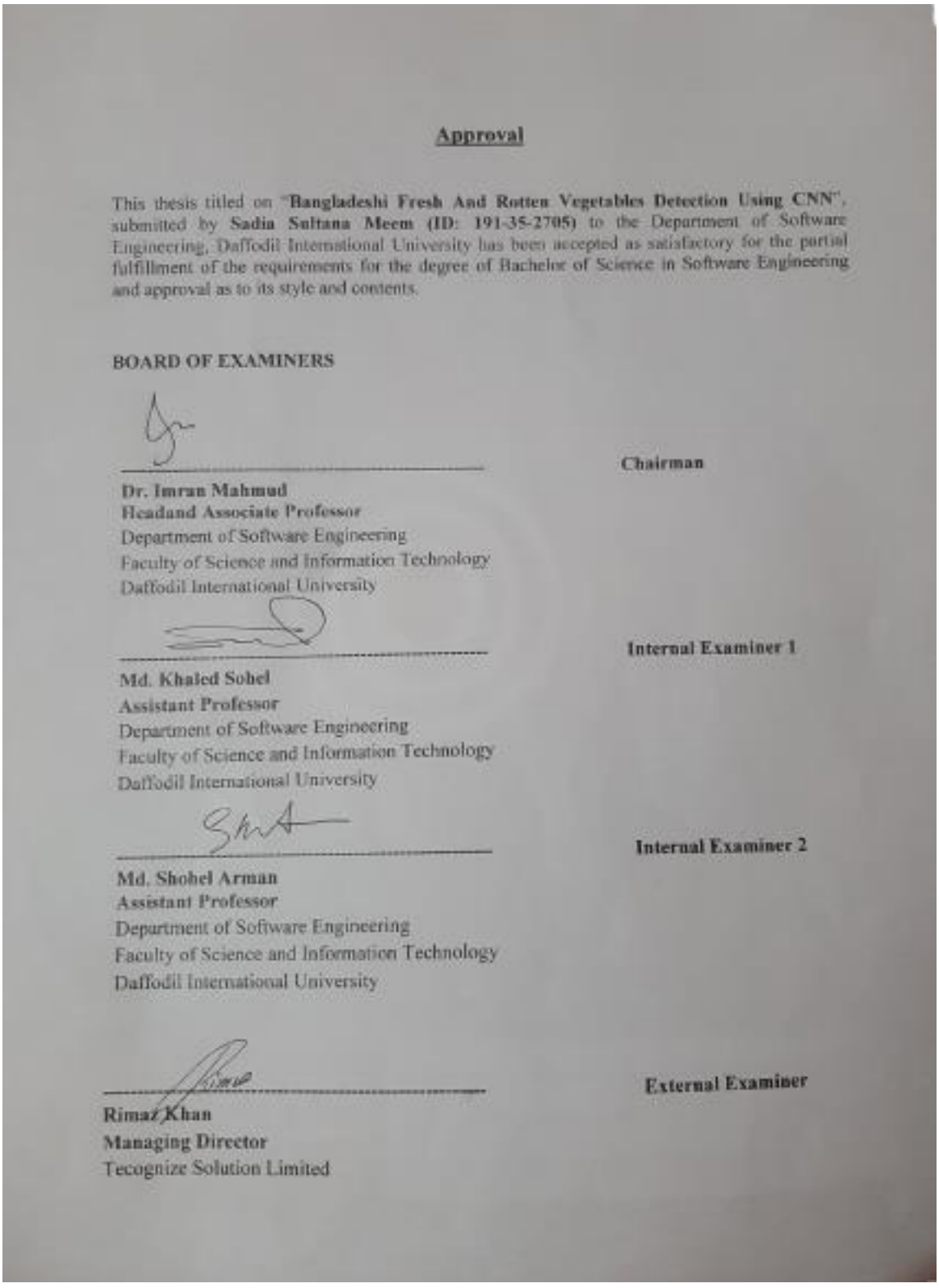
Sadia Sultana Meem
ID: 191-35-2705
Department of Software Engineering
Daffodil International University

Supervised by

Ms. Nusrat Jahan
Assistant Professor
Department of Software Engineering
Daffodil International University

This Thesis report has been submitted in fulfillment of the requirements for the Degree of Bachelor of Science in Software Engineering.

APPROVAL



Approval

This thesis titled on "**Bangladeshi Fresh And Rotten Vegetables Detection Using CNN**", submitted by **Sadia Sultana Meem (ID: 191-35-2705)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Dr. Imran Mahmud
Headand Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Chairman

Md. Khaled Sohel
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

Md. Shohel Arman
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Rimaz Khan
Managing Director
Tecognize Solution Limited

External Examiner

DECLARATION

I hereby announce that, I have done this thesis under the supervision of **Ms. Nusrat Jahan**, **Assistant Professor**, Department of Software Engineering, Daffodil International University. It is also stated that neither this work nor any portion of this work has been submitted for any degree or graduation.

Certified By:

…….

Ms. Nusrat Jahan
Assistant Professor
Department of Software Engineering
Daffodil International University

Submitted By:

…….

Sadia Sultana Meem
ID: 191-35-2705
Batch: 28th
Department of Software Engineering
Daffodil International University

ACKNOWLEDGEMENT

At first, I want express my heartfelt thanks and gratitude to my Almighty Allah for His immense blessings that make it possible to complete my final thesis successfully.

I am really grateful and wish my profound indebtedness to **Ms. Nusrat Jahan, Assistant Professor,** Department of Software Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge & intense interest of my supervisor in the field of "*Machine Learning*" helped me to carry out my thesis. Her boundless patience, dedication, scholarly guidance, constant & energetic supervision, continual encouragement & motivation, valuable advice, constructive criticism, reading many inferior drafts and then correcting these at all stages have made it possible to complete this thesis.

I would like to express my heartiest thanks to **Dr. Imran Mahmud, Associate Professor,** and Head In-Charge, Department of Software Engineering for his cordial help to complete my thesis and also to other teachers for their support & motivation.

Finally, I want to thank my beloved family and friends who supported me from the very beginning and motivated me to complete my work successfully.

ABSTRACT

Vegetables cultivation is part of our Bangladeshi Farming Industry. Vegetables also helps human to get lot of vitamins which keep body energetic and fit. Furthermore, in our nation almost every people get benefits from agriculture. Hence, it's important to maintain vegetables freshness. Unhealthy vegetables can harm human body. Also, farmers and retailers or vendors can get losses if they buy rotten or spoiled vegetable. For this reason, I came up with an idea to distinguish all the rotten & fresh vegetables. I suggested a model that can identify the fresh-rotten veggies. It's nearly impossible for humans to do this difficult task as there can be thousands of vegetables. To distinguish all of them will be challenging. Our model will classify the veggies as Fresh & Rotten. To do this task, I used CNN model. It will classify our data into fresh- rotten after giving input of vegetables image. Then I compared our model with SVM & KNN techniques too. Our proposed model performed better than these two algorithms. CNN model obtained accuracy of 93%. I'll work with other types of vegetables & more images with other techniques to get better result in future. This research will be helpful for our country.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Background

Human body requires vegetables as it's rich in vitamins, fiber and minerals including folate, potassium etc. Fresh vegetables consumption is given a special consideration as it helps humans to have a healthy and fit body. Fresh vegetables also help body to physically develop and grow. To be free from digestive problems, fresh vegetable is really important. Bangladesh is a nation which main industry for employment is agriculture. Agriculture plays a vital role in our economic growth. And Bangladesh is one of the largest countries to produce vegetables in the world. But every year our nation faces a huge loss because of spoiled or defective vegetables. Our farmers also don't have enough money to cope up with the loss. So, it's necessary to maintain the quality of the vegetables and also early detection of defective vegetables from the fresh ones to avoid any kind of loss. To able to find difference between rotten and fresh vegetables is really a challenging task as it can be looked fresh just by looking at it but can be rotten from within. Human are capable of classifying and rating fresh and rotten vegetables but it totally wastes their time and also it is expensive, dissatisfying and troublesome. Sometimes, we can't find difference between the health and unhealthy ones of some items just by observing. And some shops or sellers can mislead by giving the rotten ones. Hence, it's really important to classify rotten or spoiled, unhealthy vegetables from the fresh vegetables.

As the world is fast-developing, everything now-a-days is done automatically by machine without human brain. Automation in classification is doing really well as it reduces manual errors. Traditional methods are displeasing and also inconvenient. In this matter, machine can do better performance with less errors. We can get successful output by machines without doing manually anything and it saves our precious time, efforts, farmers extra work and production cost etc. Machine learning and deep learning based techniques are doing extraordinary performance in detecting or classification. It's making our monitoring

of everyday work or any kind of activities easy. Image classification or categorization is an important technique that we need in our daily life to detect and classify important things. We can evaluate many tasks of our everyday life implementing image classification through machine or deep learning models. It's benefits and importance growing day by day. And in the near future, people will use these models more widely.

There are many types of methods or models which are used to solve problems of image classification or recognizing any image. But Convolutional Neural Network which is called CNN, based on deep learning is found doing better performance in detection or image classification. It can detect images by finding valuable features of an image without human oversight. In many studies, researchers used different types of models to do this type of works and most of them used CNN method and obtained better result. As this is image-processing related work, therefore in this work, we proposed a methodology for classification of our vegetables. We are using CNN to detect our datasets of vegetables and classify them into two category which are – Fresh vegetable and Rotten vegetable. We will build an automated system for finding which is spoiled or which is fresh. This proposed model will save time and cost of farming sector because through this, dividing unhealthy veggies from fresh ones is much easier.

We are working with some vegetables like – brinjal, tomato, beans, chili, onion etc. The proposed model will categorize these vegetables into fresh and rotten. We will be able to identify fresh tomato or rotten tomato, fresh onion or rotten onion, fresh beans or rotten beans and like this many vegetables and according to this, we can buy or sell our vegetables. Thus, it will make our daily life simpler as we can rapidly find difference of the good or bad ones. We can decide what we will buy. This proposed model will work on this type of image classification like detecting any fresh and spoiled fruit or good and bad chemicals.

1.2 Motivation of the Research

Vegetables are an important item that we buy in our daily items. We need vegetables every day to have a healthy body free from health problems because vegetables have such nutrients that keeps human body active. But we need fresh vegetables for that. Nowadays, many people purchase vegetables from supermarkets or wholesalers. Those supermarkets or wholesaler keep all types of vegetables whether it can be spoiled or rotten and fresh. Sometimes, shortage of time prevents them from determining whether the fruit is fresh or decaying. Sometimes, if we have the time, we don't think twice about purchasing. Therefore, lots of vegetables are getting spoiled every day, whether on purpose or by error. These rotten vegetables are not only a financial loss, but they are also bad for our health. In addition to this, Bangladesh which is farming based country, recently became 3rd in vegetables production. Currently, 3.73M tons of vegetables are produced annually on 2.57% of our nation's total geographical area. And Bangladeshi food prices are getting increased day by day. In this worst situation, if the farming products get spoiled then it will be huge loss for people. So, if steps are taken before selling then it will be benefited for farmers or wholesalers. And also, it is nearly hard to identify difference between decay or fresh veggies. People can do rating and categorizing but it's impossible to do it correctly as there can be thousands of datasets. From this perspective, we want to build a system which will automatically detect if the vegetables are rotten or fresh. Thus, we can save money & or valuable efforts because doing these things manually will be really tough. But automated system will easily classify those vegetables. Using image classification and the proposed model, it will be easy to identify our desired results.

1.3 Problem Statement

Finding defective or unhealthy vegetables is important for every country because vegetables is our daily needs. It's become more important when a country depends on farming. Earlier, many studies have been done on finding spoiled or fresh fruits and less work on vegetables. But research on Bangladeshi vegetables classification is rare. Also,

many researchers have applied different kinds of methods. Some used SVM, KNN, some used CNN, some used segmentation etc. Some works are done with Kaggle data & some are with google data. Some have worked for their own country. But our country's environment or climate is different from others. That's why, I decided to work for our country, Bangladesh. We want to implement deep learning model, CNN on various types of vegetables which will be collected from many big Bazars of our country. We will categorize our data in 2 types which is rotten vegetable & fresh vegetable. We will also then compare with other models to get assured that our model is working better to classify the vegetables images.

1.4 Research Questions

The main questions for our research are given below –

- Q1: Can our model classify vegetables into fresh or rotten with the training data?
- Q3: Can our model perform better than the other models?

1.5 Research Objectives

The main goal of our work is to do classification of vegetables into fresh & rotten from vegetables image datasets. We also want to propose a model which give better accuracy. The other objectives are –

- Classifying the vegetables into fresh & rotten successfully with our proposed model
- To get better outcomes than the other models.

1.6 Research Scope

Our research scopes are given below –

- Vendors or Retailers can identity the defective vegetables.
- Consumers can check quality of the vegetables before buying.
- Supermarkets, Bazars and factories can use this automated system.
- Magistrates can perform their responsibilities in different wholesale stores around our country that sell rotting items.

1.7 Thesis Organization

In the chapter 1, we talked about vegetables classification and its necessity, the background of our work like what we're doing, motivation of our research, problem statement, objectives, research questions, scopes etc.

In the 2nd chapter, we'll talk about pervious literatures and studies related to our works where we can see many researchers work, their proposed models, methodologies.

In the chapter 3, the methodology of my work is discussed where we'll show diagram, data samples, data pre-processing steps and our proposed models etc.

In the next chapter, experimental results and discussions, model loss graph etc. will be covered.

In the last chapter, our work limitations, findings & contributions lastly recommendations for future work will be discussed.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

A researcher does a literature review by reviewing past work, research, conference papers, books, articles, and so on. With this, one may determine what kind of work has done before on the problem., summarize the entire topic, and determine what are the lacking in the study. Then according to this, they can perform models on their respective works and try to get better outcomes than other studies. Some literature related to our study are described and reviewed in this chapter.

2.2 Previous Literature

Food safety is a major issue in the modern world. Industry like – food processing uses fresh fruits & vegetables to produce many healthy food products. But it has many stages to process those products like – harvesting, sorting, grading, classification etc. Human can do these steps manually but it requires expert resources and also time-consuming. That's why, automation in this industry really saves time. Susovan Jana, et. al., 2021 proposed a method which automatically detect the rotten fruits & vegetables using a CNN architecture consisted of 4 convolutional layers and for the processing of the images, they used Computer Vision & also machine learning. But in their segmentation process, there is background color issue. But this paper showed that deep learning models outdo other approaches.

This research (Arvind Vishnubhatla, 2021) main goal was to utilize machine learning algorithms to distinguish between fresh and rotting veggies. To do this work, they had to partition their data into the training & the test set and load a number of libraries. CNN was used for feature extraction purpose. Here data augmentation is used for preventing over-fitting. But this study was done with a less variety of dataset.

As Fruits & Vegetables are really needed nutrients for humans to grow & develop physically & fruits can be rotted if they're not stored properly because of the spread of bacteria, hence Chai C., et. al., 2021 proposed a method which apply the deep learning technique-CNN which extract feature and then do classification of three types of rotten fruits (Orange, Banana, Apple). They got 98.89% of validation accuracy using this method.

To avoid food waste, a method was proposed in this research to differentiate spoiled fruits and fresh fruits. This study included a variety of fruits including apples, oranges & bananas. CNN models were used to categorize images into fresh & rotten. The effectiveness of the proposed model was assessed using a dataset from Kaggle & 97% accuracy was obtained. (T. Bharath Kumar, et. al., 2022)

Another similar kind of work is done to prevent fruit rottenness. Palakodati, et. al., 2020 suggested a model that divides the input fruit dataset into fresh and rotting categories. Three different kinds of fruits were employed in this study which were derived from Kaggle. Softmax was used to categorize the input fruits into fresh & rotten fruits and CNN model was utilized for feature extraction from images. It got accuracy of 97% The findings shown that the CNN model can successfully categorize both fresh and rotting fruits better than transfer learning models.

Febrian Valentino, et. al., 2021 proposed a design which is based on Computer vision with CNN model and it detect freshness of fruits. This study evaluated by public datasets consists of fresh fruits and rotten fruits derived from Kaggle. Accuracy is 88%

Automation categorization is needed in fruit freshness as conventional method can be inconvenient and troublesome. This work completed by Rupali Pathak, et. al., 2021 approached CNN model to extract features from the fruit images and then category them in 'fresh' or 'rotten'. Their proposed model gives a good accuracy which is 98.23%. This research also done comparison with transfer learning model and show that CNN models are comparatively more effective than the transfer learning models.

The goal of the study is to develop a model for fruit identification and categorization. They

used deep learning model – CNN for extraction. It may be used to determine if the fruits are fresh or not by examine their condition. ImageNet dataset was used as it's quality is high. Five types of fruits were used. The achieved accuracy is 92.23%.ML algorithms give less accuracy than deep learning. (Deepali M Bongulwar, 2021)

Diclehan Karakaya, et. al., 2019 done a study on classifying fruits using Support Vector Machine (SVM) as the traditional methods are time consuming. In this paper they analyzed image dataset of 3 kinds of fruits to differentiate the fresh ones from the rotten ones. They've done segmentation of the images and then extracted feature using histograms, CNN, co-occurrence matrix. This paper shows that better success rate is obtained from CNN based feature extractors. Concatenation of many of proper features can induce better success.

In fruit sector image processing & machine vision systems are found to be helpful to enhance product quality. This study discusses numerous image processing algorithms for fruit categorization. In this methodology seven types of fruits classification were done. CNN based models were used for the classification. Accuracy was 98%. It showed CNN gives better performance in this field. (Mehenag Khatun, et.al., 2020)

CNN models have shown extraordinary performance in image classification. And also humans can do error sometimes because of distraction in ensuring quality of fruits. That's why, Shawon Ashraf, et. al., 2019 have done a study which uses CNN models (VGG16, Inceptionv3) to identify fruits whether as fresh or as rotten. For training the datasets they used transfer learning. Training set accuracy is good for VGG16 but not for multiclass classification (Inceptionv3) due to dataset low dimensionality.

The techniques shown in this research (Ann Nosseir, et.al., 2019) distinguishes between four different types of fruits and tells the fresh from the decaying. The algorithms are based on colour and textural characteristics of the photos of the fruits. They used KNN & SVM respectively for this purpose. They worked with less fruits.

Automation improves nation's economic growth, quality and production in agriculture

science. Fruits & vegetables selection has an impact on the export market and quality assessment. Even though human can do grading & sorting but it's not always appropriate also time is wasted & readily impacted by their surroundings. Consequently, a proper technique is required for fruit grading. The quality of fruits and vegetables was addressed using a number of approaches, including pre-processing, feature extraction, segmentation, and classification, in this work. (Anuja Bhargava, et.al., 2021) This study makes an effort to examine and contrast the various approaches and algorithms put forth by scholars at each stage. They used computer vision for this.

In order to increase the precision of automatic vegetable detection and classification, this research of Zhenbo Li, et. al., 2020 proposes a deep learning-based approach for doing so. They used VGG, VGG-M, VGG-M-BN, AlexNet models to train the datasets. VGG-M-BN had accuracy of 96.55% compared to VGG (92.1%) and AlexNet (86.3%). This research also demonstrates that an increase in batch size enhances the VGG-M-BN network's classification accuracy.

This study propose method for rotten vegetables segmentation. The segmentation techniques like - Color Based Segmentation, Edge Detection & Marker Based Segmentation were used successfully to distinguish between from healthy & unhealthy veggies.8 sets of images were used for the evaluation of the methods. (Kyamelia Roy, et. al., 2019)

This paper gives a description on several techniques including – data pre-processing, feature extraction, segmentation & classification to do quality check of Tomato. To extract tomato properties including color, shape, image processing etc procedures have been done using MATLAB software. ANN, SVM & Decision tree models were compared in this research. SVM gave better accuracy than the other two algorithms. (Haichun Zuo, 2022)

This study done by Jianxin Xue, et. al., 2022 shows that fresh-cut cauliflower samples were divided into several groups and checked for levels of rotting using machine vision. First, segmentation of the fresh-cut samples & extraction of the single samples was done using the modified watershed algorithm. Then using two feature extraction techniques pictures

feature parameters were extracted. ELM and PLS-DA discriminant models were developed concurrently. PLS-DA and ELM discriminant models had identification accuracy for decaying samples of 95 and 90.9%, respectively. The findings demonstrated the capability of machine vision to separate coherent fresh-cut cauliflower samples and to identify fresh and rotting cauliflower samples both qualitatively and quantitatively.

The fruit/vegetable crop okra is sometimes referred to as lady's finger. Okra is graded according to its freshness, softness, shape, color, decay, scarring etc. Four categories are made in this research. To classify the okra, machine vision technology is investigated. The researchers Meenaxi M, et. al., 2020 used AlexNet, ResNet50 and GoogLeNet for Okra grading. Dataset size of Okra which includes all sizes, is 3200. The accuracy for ResNet50 is 99%, which is higher than the accuracy for AlexNet, which is 63.45%, and GoogleNet, which is 68.99%.

In this paper (Yanlei Xu, et. al., 2022) the researchers proposed an autonomous classification of lettuce freshness based on the Improved ResNet (Im-ResNet). They created an image capturing method to increase the classification accuracy. The compared Im-ResNet with 4 other models ResNet50, GoogleNet, AlexNet, VGG16. The findings demonstrated that their proposed approach gave better result which has accuracy of 96.50 percent.

Labiba Gillani, et. al., 2021 suggested automatic categorization of fruits & vegetables. They used dataset consists of 11 vegetables & fruits divided into 3 freshness groups. VGG16 & YOLO-deep learning models were used for recognition. They also developed android based app. They got 84% mean average precision.

Kausar, et. al., 2018 suggested PCNN model with the fewest parameters to distinguish numerous fruits more reliably. To prevent overfitting, they used gap layer. They showed classification results using PCNN on the freshly released fruit-360 dataset. The PCNN achieves an accuracy of 98.88% based on the trial findings of 55244 color fruit pictures.

This paper (Mukhiddinov, et. al., 2022) describes a deep-learning technique which is updated YOLOv4 model that detects item type in images before categorizing it into fresh or rotten.

The suggested system works with collecting a picture dataset of fruit & vegetable, developing an efficient YOLOv4 model. Furthermore, the suggested model's backbone was improved with the Mish activate function and quick detection. 12000 images & 5 types of fruits & veggies were used. The result shows that, improved YOLOv4 performed better accuracy than YOLOv4 with 73.5% and 72.6% accuracy on the dataset.

2.3 Conclusion

In the above mentioned literatures, we see that there are various types of machine and deep learning techniques to do this kind of work. To get their desired outcomes, they used segmentation, KNN, SVM, CNN various models, Feature extraction etc. In our work, we will used CNN model to classify our datasets and finding better result.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

The goal of our study is to detect the rotten and the fresh vegetables. We are mainly working with Bangladeshi vegetables. As ours is a farming nation so it's really important to secure the safety of products to avoid losses. This proposed model will differentiate the defective or rotten fruits from the fresh or healthy ones. The detection model is built using data consists of Bangladeshi fresh and rotten vegetables. We have used deep learning model – Convolutional Neural Networks (CNN) to implement this. We've also used K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) for comparison with our model. This chapter explains the details of our proposed CNN model and also the SVM and KNN model.

3.2 The Methodology Diagram

To do our study, we have so many different steps to follow. If an architecture is done of the system, we can easily go with the flow to do the work without any confusion. Else the proposed methodology will not be easy to complete. For this, we need a diagram which will show all the working steps to continue the study. Through the methodology diagram, we can see the whole working process of what steps we've followed to complete this research. We've outlined every step of our workflow in the below Fig 3.2. This diagram shows components that makes the system function. This is our proposed methodology diagram.

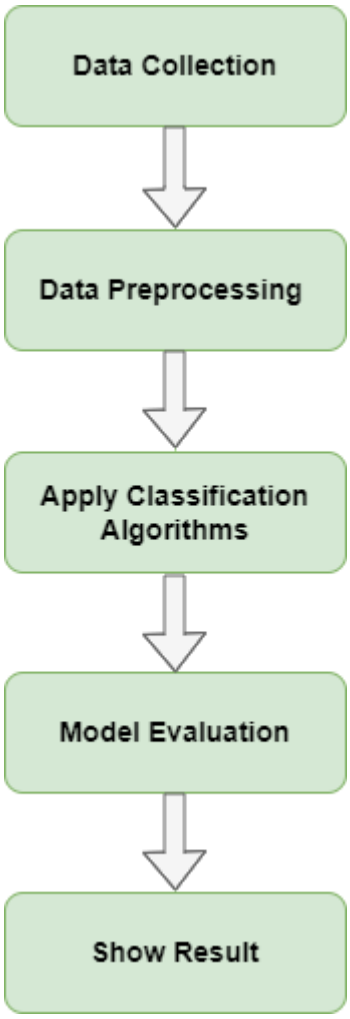


Figure 3.1: Methodology Diagram

We'll follow these steps to do our study. Firstly, we have to collect data then the data have to be pre-processed following some steps. Then our classification algorithms have to be applied which are CNN, KNN, SVM. Finally, we can evaluate the model and get our predicted result. The steps are discussed in details below as follows:

3.3 Data Collection

We worked with some Bangladeshi vegetables in our study that are almost available. Our data is divided into two categories: Fresh vegetables and Rotten vegetables. Six types of vegetables were used in our research. There were mostly works on Fresh fruits and Rotten fruits detection and a very few on vegetables but there is no work with Bangladeshi vegetables data. We collected our data manually from home, Kawran Bazar and some from Google. The vegetables are as follows:

1. Tomato
2. Chili
3. Beans
4. Onion
5. Brinjal
6. Cauliflower

Some fresh vegetables samples:

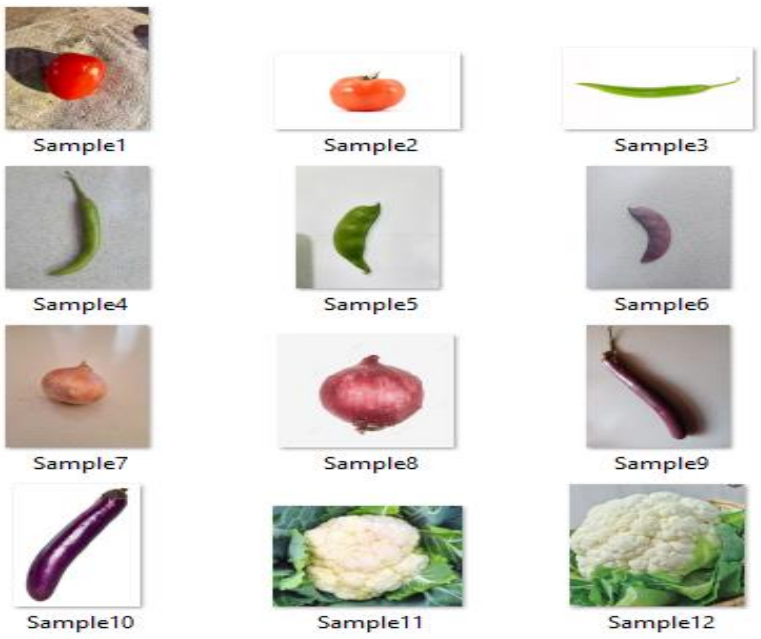


Figure 3.2: Fresh Vegetables samples

Some rotten vegetables samples:

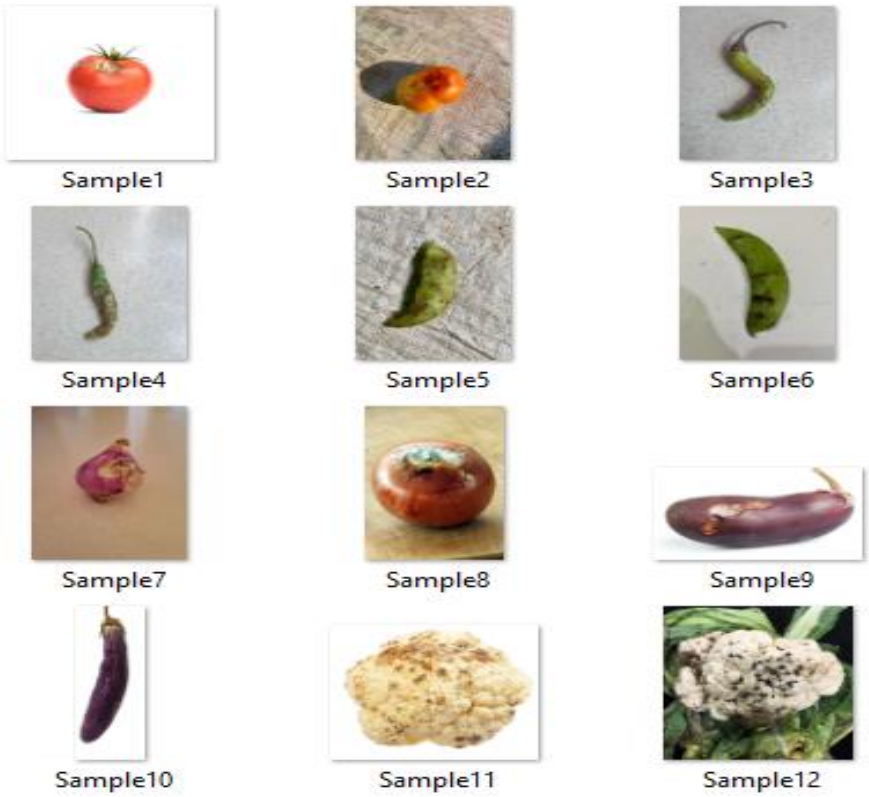


Figure 3.3: Rotten Vegetables samples

The amount of data we have collected for the fresh and rotten ones are listed below in a table.

Vegetables	Amount
Fresh Tomato	200
Fresh Chili	200
Fresh Beans	200
Fresh Onion	200
Fresh Brinjal	160
Fresh Cauliflower	180

Rotten Tomato	180
Rotten Chili	180
Rotten Beans	160
Rotten Onion	160
Rotten Brinjal	140
Rotten Cauliflower	140

Table 3.1: Data collection

3.5 Data Pre-processing

After collecting our data of fresh & rotten vegetables, we have to preprocess the dataset. Data pre-processing is a technique where raw data is prepared to make useful format for machine learning models. It is crucial step to make a machine learning model. As when we start to make a model, we don't have clear or formatted data. Hence for this, firstly we have to apply preprocesses steps to the dataset. It's necessary to get our expected result and also it increases the accuracy.

We have collected total 2100 images for our research. For classification, our data is divided into two categories – Fresh and Rotten. From the above Table 3.1 we see the amount of our all types of fresh and rotten vegetables dataset. After collecting the vegetables images, I have split them into train datasets and test datasets. In both of the sets, there are total 6 types of vegetables and these 6 types are divided into fresh & rotten categories. The below figure explains the labelling easily.

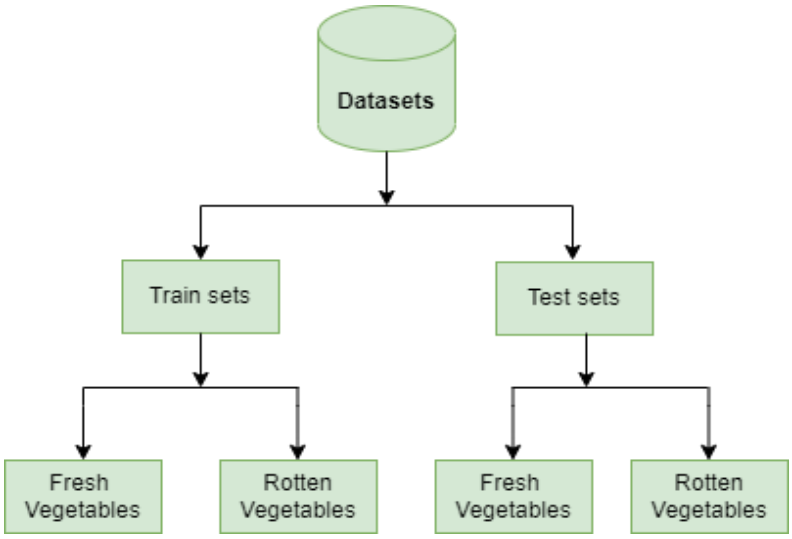


Figure 3.4: Dataset labelling

We have 80% data in the training sets and 20% data in the test sets. We have listed these in the below Table 3.2 for showing it in an easy way.

Vegetables	Amount of train sets	Amount of test sets
Fresh Tomato	160	40
Fresh Chili	160	40
Fresh Beans	160	40
Fresh Onion	160	40
Fresh Brinjal	128	32
Fresh Cauliflower	144	36
Rotten Tomato	144	36
Rotten Chili	144	36
Rotten Beans	128	32
Rotten Onion	128	32
Rotten Brinjal	112	28
Rotten Cauliflower	112	28

Table 3.2: Train and Test sets amount

Firstly, for the processing we have to examine and check all the datasets to ensure that there is no irrelevant data or information. This will be effective to do the processing. As we have fresh and rotten – 2 types of data so, it can be mixed together. Hence, we need to ensure that there is no such data that belongs to another category. And the unnecessary data will be removed in this way.

Next, we removed the background of all the images. ‘rembg’ which is a third-party library was used to do this work. By using this tool, we can remove background of all our images at once. It’s really difficult to do this with the help of photoshop or manually one by one and it also waste our time. But this library done our work easily and fast.

Next step is spot mark on all the images. Here, we additionally make use of the "contours" third-party library which is used for shaping and image detection. Many different stains and color alterations appear on fruits and vegetables when they begin to decay. The color that has altered from its natural color or the spots is basically how the machine decides if the vegetable is fresh or not. Therefore, in this situation, marking is required. The system will determine the precise percentage based on the size of the spot. It’s a very important step as depending on this, the system can predict if the vegetable is rotten or fresh. From the below Figure 3.4 we can understand this easily.

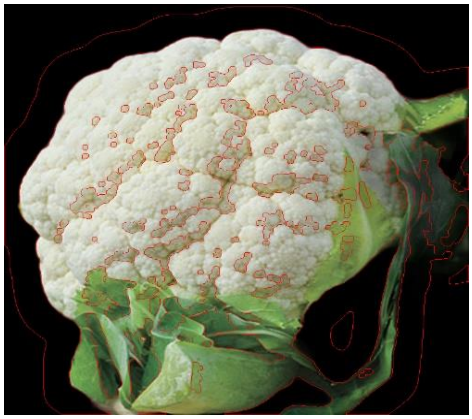


Figure 3.5: Mark Spot Sample

Next, we have to resize all the images of our dataset because our images have different sizes. It's import to have same size in all the datasets. NumPy will be used for this. For simpler analysis, images are turned into NumPy arrays. The neural network's input size is 500*500 pixels. Since the machine can only handle integer & floating types of values, it is necessary to convert the scaled image to NumPy. Each pixel is converted to a float variable using NumPy. All data are transferred into one dimension. That's how data is saved for the train sets.

3.5 Classification Algorithms

For classification of our data, we have used three algorithms: CNN, KNN and SVM. The basic idea about these algorithms is given below:

3.5.1 Convolutional Neural Network

For acquiring our output, we will use CNN as our proposed model which is a subset of machine learning and used for image classification or recognition. It can detect important features automatically without analyzing the whole image. And also, without any human direction. CNN model helps the machine to remember the features of images then it can predict if a new image is added.

Convolutional layer comes first after taking input which is used for extracting features. Here, we used Conv2d layer. It starts with filter size of 16 and 3x3 karnel then it activates ReLu function.

Then comes pooling layer. Here 2x2 max-pooling layer reduced dimensionality of feature map. In this way it reduced number of parameters. Flatten function was used for single dimension. Then comes fully connected layer which does classification was used with ReLu activation. Dropout was used to remove the overfitting. Lastly a dense layer was used with sigmoid activation.

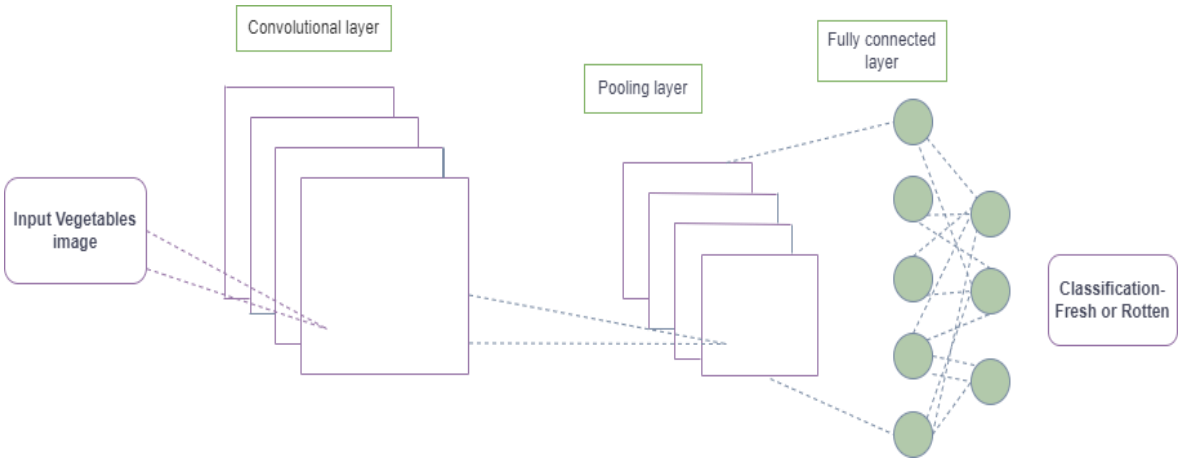


Figure 3.6: Convolutional Neural Network Structure

3.5.2 K-Nearest Neighbor

For comparison with our model, we're using KNN classification here. It will classify the fresh and rotten vegetables. We know, KNN is based on supervised learning method. It is used mostly for classification problems. In our work, we'll use KNN classification to categorize each of the classes in an image. KNN method assumes similarity between existing data and new data and then places the new data into that category which is most like to existing data. When new data enters, it can be easily classified as all the available data are stored in KNN. As we are working with fresh and rotten 2 types of vegetables so to classify the vegetables into fresh and rotten, we'll apply this KNN method. It stores the data at train phase and when new data enters it'll classify them into these categories. KNN mainly works by finding the K value which represents the number of nearest neighbor. To find K value, we'll use Euclidean Distance between all data point. Then the new data point will be assigned to the category that has maximum number of neighbor.

Euclidean Distance formula between two data point is:

$$d = \sqrt{[(x_2 - x_1)^2 + (y_2 - y_1)^2]} \quad (3.1)$$

3.5.3 Support Vector Machine

SVM is based on supervised learning which is used to find solution for 2 group classification problem and also in regression. But it mostly solves classification problems like – image classification, object detection etc. SVM create decision boundary called hyperplane which divide n-dimensional space in classes. Thus, it quickly can classify new data points when needed. SVM selects extreme vectors that creates hyperplane and those extreme classes known as support vector.

There are two types of SVM. Linear SVM is used when it is simple to segregate data using a hyperplane and a straight line. Non-linear SVM can't be used just by drawing a straight line. For linear input we've two dimensions, x & y and for non-linear we've another dimension, z.

The equation formula is: $z = x^2 + y^2$ (3.2)

3.6 Output

We got our result after applying these above algorithms. We have worked with six types of vegetables. And our proposed model can classify the fresh and rotten ones successfully. We have got better result of the fresh tomato than the other vegetables. We got lowest for the fresh brinjal vegetable. Then in the rotten ones we got better result for tomato and less for cauliflower.

Here are some fresh vegetables result samples:

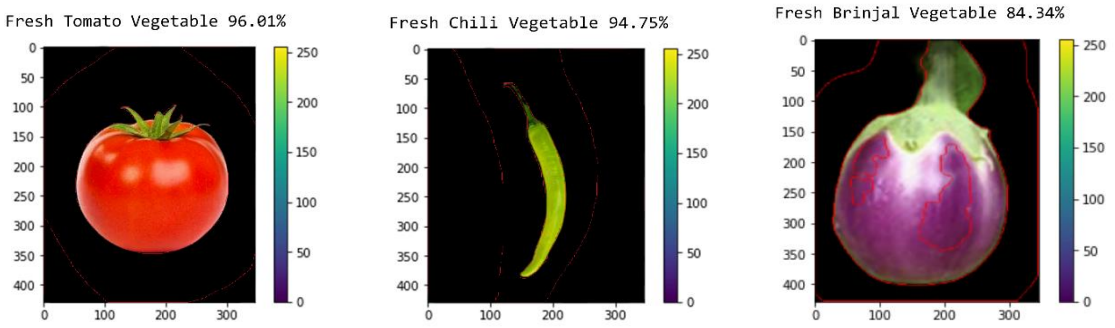


Figure 3.7: Fresh Vegetables Output

Here are some rotten vegetables result samples:

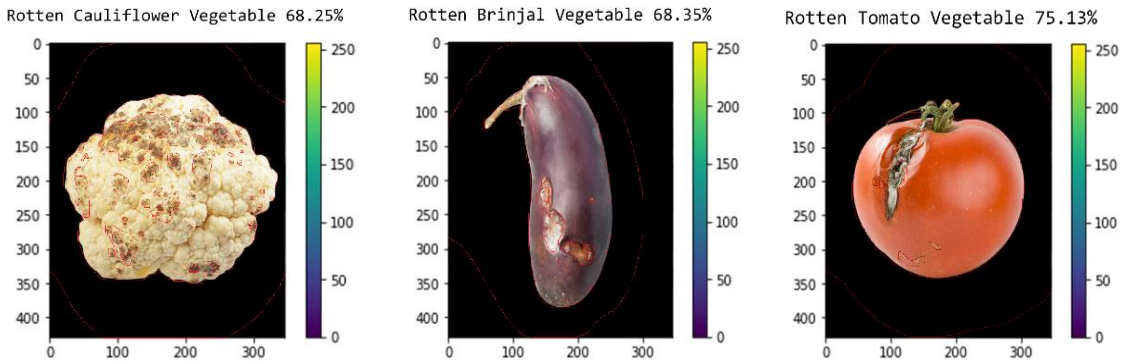


Figure 3.8: Rotten Vegetables Output

We can view from the above samples that we got accuracy of 96.01% for fresh tomato data, like this for chili data 94.75%. Similarly, for rotten tomato we obtained 75.13%. That's how we got our outcomes for all the data we used in this work.

CHAPTER 4
RESULTS AND DISCUSSION

4.1 Introduction

Classification process of the fresh and rotten vegetables are described in previous part. And we saw that the model can successfully detect the expected outcome. In this chapter, we'll see the experimental results of the models like- model loss, accuracy in details.

4.2 Experiment Result

We obtained our model's accuracy and loss graph after applying the proposed model to our dataset. To obtain a graph of the model's accuracy and loss, train accuracy & validation accuracy are compared.

Accuracy of a model decides how correctly a model can predict or detect the results. Based on training, data or the input, accuracy is used to indicate which model is better at finding patterns among datasets. There is formula for finding accuracy which is given below:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (4.1)$$

Here,

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

A result where the model properly predicted positive class is referred to as true positive. Similarly, true negative is a result for which the model accurately predicts a negative class. When a model predicts the positive class inaccurately, the result is false positive. False

negative occurs when the model predicts a negative class inaccurately. We got our model accuracy implementing this equation 4.1 The model accuracy graph is given below:

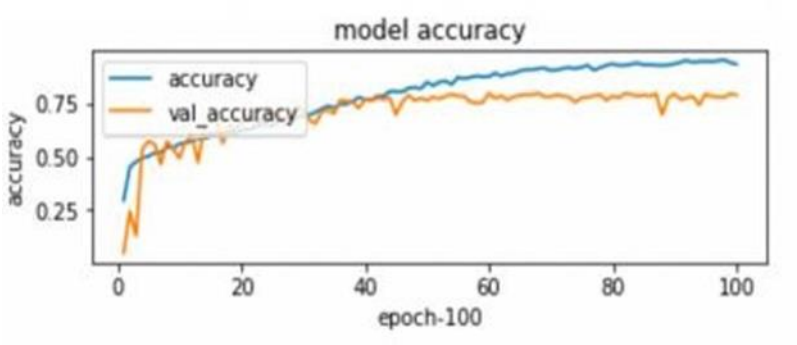


Figure 4.1: Training Accuracy vs. Validation Accuracy

In the above Figure 4.1, training accuracy vs validation accuracy is showed. The blue line represents training accuracy and the orange line represents validation accuracy. We used 100 epochs in this model. The accuracy for training and validation increases gradually. We got like 93% for training accuracy.

Model loss is a function which compares predicted & target values. It indicates how bad was the model to predict a sample. The below Figure 4.2 shows the training loss vs validation loss where validation loss is decreasing.

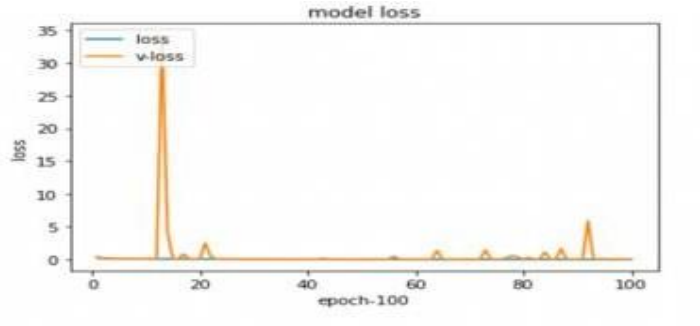


Figure 4.2: Training loss vs. Validation loss

We have also used KNN and SVM to compare with our model. SVM got 65% training accuracy and KNN got 63%. All the classification result is given below in the Table 4.1

Algorithm Techniques	Training Accuracy	Validation Accuracy
CNN	93%	76%
KNN	63%	57%
SVM	65%	55%

Table 4.1: Algorithms accuracy result

4.3 Discussion

After analyzing all the data that we gathered and developing a model utilizing the provided data, we were able to obtain the expected outcomes. We saw our model is doing good in terms of accuracy after training & testing the datasets. It can detect the fresh & also the rotten vegetables with an accuracy of 93%. We also saw that our model gives better outcome than the two models which we used to compare. So, CNN does a good job in fresh & vegetable's classification.

CHAPTER 5

CONCLUSION AND LIMITATIONS

5.1 Findings & Contributions

The main goal of our study was finding difference between rotten vegetables and the fresh vegetables from given input. In the farming sector, it's really important to be able to determine which are the good ones or rotten ones or there can be huge loss. It will be a help to the farmers to avoid losses. They can identify the rotten ones easily. And with automation system, time will also be saved because manually doing this kind of tasks require lots of time. There are many works on fruits fresh & rotten classification and some were implemented using Kaggle data. The other countries already built systems like this. But there is no particular work with Bangladeshi vegetables. And with this system, our farmers will get benefit. Here we have worked with some known Bangladeshi vegetables. We successfully detected our expected results. In our study a CNN model was used that concentrates on finding and classifying the rotten/unhealthy and healthy/fresh vegetables. For our inquiry, we first gathered data manually in many ways. After preprocessing, we split up the train & test data into our task. We have used total 2100 data where we used 1680 images for train sets and 420 images for test sets. Later we got our outcome.

We used Anaconda application to implement the code and testing. The CNN model's convolutional layers constructed model's accuracy and also loss graphs. The outcome demonstrated that our proposed CNN model can distinguish between the rotten & fresh data of vegetables with a better accuracy. We saw, for various vegetables data it gives different results. Then we also calculated training accuracy and validation accuracy. The proposed model obtained 93% accuracy. We have also used two other machine learning models – KNN & SVM for comparison with our model with same data. We saw from the outcomes that CNN does better result in classifying the datasets into fresh & rotten categories. So, our model can be used to classifying this type of work with less time.

5.2 Limitations

We have some limitations in our study. Our main limitation was datasets. In the deep and machine learning models, accurate data gathering is very essential. For image classification, we need clear, good quality images to do processing or else the model can't give a good result. As we're working with vegetables classification into fresh and rotten, hence we needed very clear and accurate data to have a better accuracy. We had a good amount of fresh vegetables data but we had less rotten datasets. We couldn't find a large amounts of rotten ones. That's why, our result for rotten vegetables isn't that much good. We couldn't obtain better accuracy for rotten datasets and also for some fresh datasets. We couldn't obtain SVM & KNN model accuracy well as they're very time-consuming models. My device was a little bit slow too and it takes much time to run models. That's why, we couldn't try enough to train the models. Hence, KNN & SVM accuracy isn't good.

5.3 Recommendation for Future Works

As CNN requires a huge set of data, we will use more datasets in the future. Because of datasets our accuracy couldn't do better in some vegetables. We couldn't do well for the rotten ones so we'll collect more images for the rotten ones to have better accuracy. And there are different kinds of vegetables so we will work with more types of vegetables in the future. And SVM and KNN models couldn't do well. We will try to find better accuracy for these two models. As the world is fast in developing and technology is doing really well in this modern world so it will really be great to have an application for this. The application will be helpful for saving more time. So, I will try to implement an application in the future.

REFERENCES

Ashraf, S., Kadery, I., Chowdhury, M. A., Mahbub, T. Z., & Rahman, R. M. (2019). Fruit Image Classification Using Convolutional Neural Networks. *International Journal of Software Innovation (IJSI), 7*(4), 51-70.

A. Kausar, M. Sharif, J. Park and D. R. Shin, "Pure-CNN: A Framework for Fruit Images Classification," 2018 International Conference on Computational Science and Computational Intelligence (CSCI), 2018, pp. 404-408, doi: 10.1109/CSCI46756.2018.00082.

Bhargava, A. and Bansal, A., 2021. Fruits and vegetables quality evaluation using computer vision: A review. Journal of King Saud University-Computer and Information Sciences, 33(3), pp.243-257.

Bhargava, A., & Bansal, A. (2019, December 20). Automatic Detection and Grading of Multiple Fruits by Machine Learning. *Food Analytical Methods*, *13*(3), 751–761. https://doi.org/10.1007/s12161-019-01690-6

C. C. Foong, G. K. Meng and L. L. Tze, "Convolutional Neural Network based Rotten Fruit Detection using ResNet50," *2021 IEEE 12th Control and System Graduate Research Colloquium (ICSGRC)*, 2021, pp. 75-80, doi: 10.1109/ICSGRC53186.2021.9515280.

D. Karakaya, O. Ulucan and M. Turkan, "A Comparative Analysis on Fruit Freshness Classification," *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2019, pp. 1-4, doi: 10.1109/ASYU48272.2019.8946385.

Gilani, Labiba & Tahir, Syed Fahad & Rasheed, Usama & Saqib, Hafsa & Hassan, Mehdi & Alquhayz, Hani. (2022). Fruits and Vegetables Freshness Categorization Using Deep Learning. Computers, Materials & Continua. 71. 5083-5098. 10.32604/cmc.2022.023357.

Jana, S., Parekh, R., Sarkar, B. (2021). Detection of Rotten Fruits and Vegetables Using Deep Learning. In: Uddin, M.S., Bansal, J.C. (eds) Computer Vision and Machine Learning in Agriculture. Algorithms for Intelligent Systems. Springer, Singapore.

JOURNALS, G. R. D. (2020). Fruits Classification using Convolutional Neural Network. GRD Journals .

K. Roy, S. S. Chaudhuri, S. Bhattacharjee, S. Manna and T. Chakraborty, "Segmentation Techniques for Rotten Fruit detection," *2019 International Conference on Opto-Electronics and Applied Optics (Optronix)*, 2019, pp. 1-4, doi: 10.1109/OPTRONIX.2019.8862367.

Liu, X., Zhao, D., Jia, W., Ji, W., & Sun, Y. (2019). A Detection Method for Apple Fruits Based on Color and Shape Features. *IEEE Access*, *7*, 67923–67933. https://doi.org/10.1109/access.2019.2918313

Mukhiddinov, M., Muminov, A., & Cho, J. (2022, October 26). Improved Classification Approach for Fruits and Vegetables Freshness Based on Deep Learning. Sensors, 22(21), 8192. https://doi.org/10.3390/s22218192

Nosseir, A., & Ahmed, S. E. A. (2019, February 14). Automatic Classification for Fruits' Types and Identification of Rotten Ones Using k-NN and SVM. *International Journal of Online and Biomedical Engineering (IJOE)*, *15*(03), 47. https://doi.org/10.3991/ijoe.v15i03.9832

Palakodati, S.S.S., Chirra, V.R., Dasari, Y., Bulla, S. (2020). Fresh and rotten fruits classification using CNN and transfer learning. Revue d'Intelligence Artificielle, Vol. 34, No. 5, pp. 617-622.

Pathak, R. and Makwana, H., 2021. CLASSIFICATION OF FRUITS USING CONVOLUTIONAL NEURAL NETWORK AND TRANSFER LEARNING MODELS. Journal of Management Information and Decision Sciences, 24, pp.1-12.

Raikar, M. M., S M, M., Kuchanur, C., Girraddi, S., & Benagi, P. (2020). Classification and Grading of Okra-ladies finger using Deep Learning. *Procedia Computer Science*, *171*, 2380–2389. https://doi.org/10.1016/j.procs.2020.04.258

Tbk, Bharath & Prashar, Deepak & Vaidya, Gayatri & Kumar, Vipin & Kumar, S. & Sammy, F. (2022). A Novel Model to Detect and Classify Fresh and Damaged Fruits to Reduce Food Waste Using a Deep Learning Technique. Journal of Food Quality. 2022. 1-8. 10.1155/2022/4661108.

Vishnubhatla, A. (2021). Classification of Fresh Vegetables Through Deep Learning and Neural Networks. In: Mallick, P.K., Bhoi, A.K., Marques, G., Hugo C. de Albuquerque, V. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, vol 1317. Springer, Singapore. https://doi.org/10.1007/978-981-16-1056-1_29

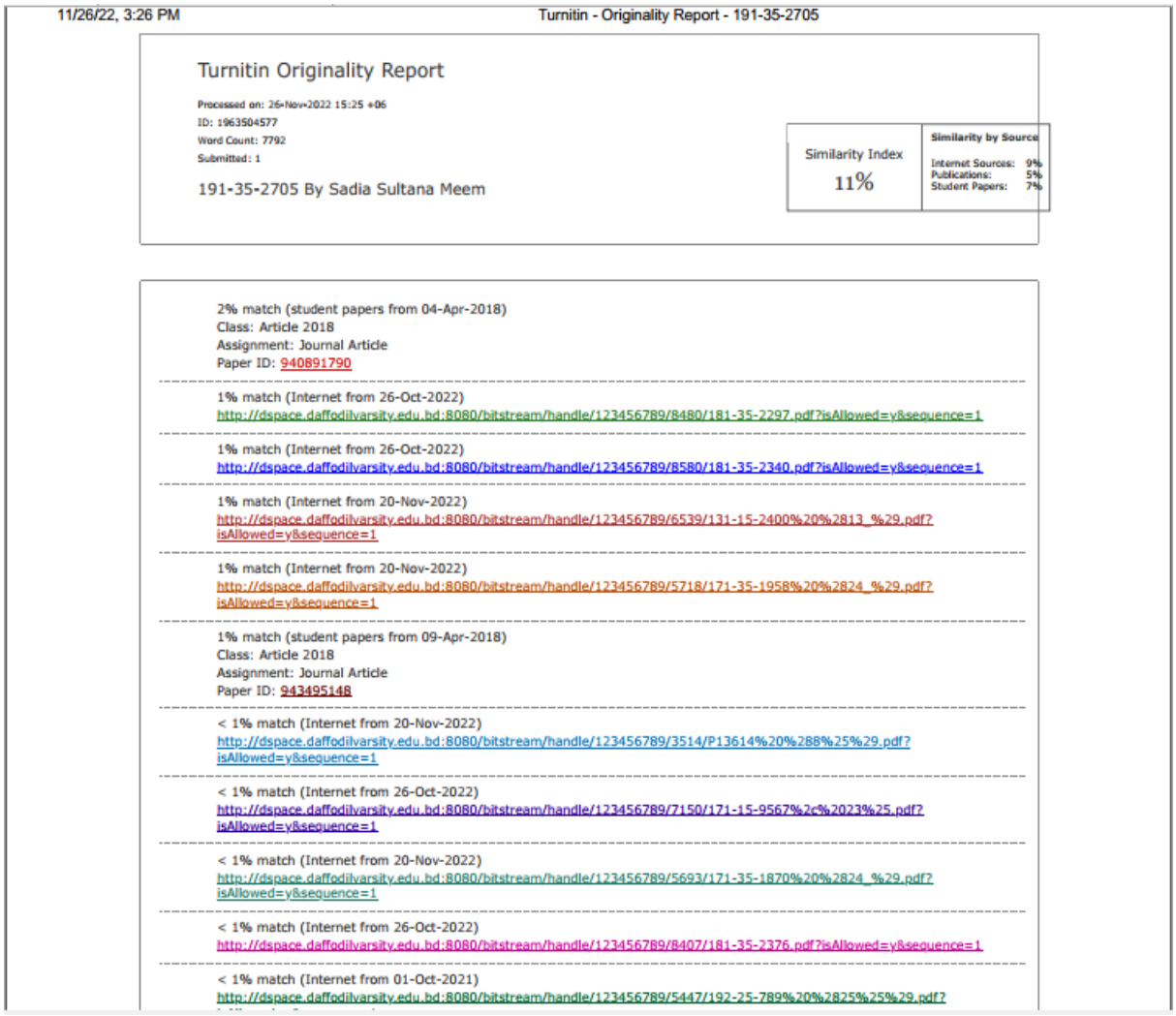
X. Liu, D. Zhao, W. Jia, W. Ji and Y. Sun, "A Detection Method for Apple Fruits Based on Color and Shape Features," in *IEEE Access*, vol. 7, pp. 67923-67933, 2019, doi: 10.1109/ACCESS.2019.2918313.

Xue, J., Huang, L., Mu, B., Wang, K., Li, Z., Sun, H. & Zhao, H. (2022). Detection of Rotten Fresh-Cut Cauliflowers based on Machine Vision Technology and Watershed Segmentation Method. *American Journal of Biochemistry and Biotechnology*, *18*(2), 155-167.

Xu, Y., Zhai, Y., Chen, Q., Kong, S., & Zhou, Y. (2022). Improved Residual Network for Automatic Classification Grading of Lettuce Freshness. *IEEE Access*, *10*, 44315–44325. https://doi.org/10.1109/access.2022.3169159

Zuo, Haichun. (2022). Analysis and Detection of Tomatoes Quality using Machine Learning Algorithm and Image Processing. 10.21203/rs.3.rs-2016895/v1.

PLAGIARISM REPORT



11/26/22, 3:26 PM Turnitin - Originality Report - 191-35-2705

Turnitin Originality Report

Processed on: 26-Nov-2022 15:25 +06
ID: 1963504577
Word Count: 7792
Submitted: 1

191-35-2705 By Sadia Sultana Meem

Similarity Index	Similarity by Source
11%	Internet Sources: 9% Publications: 5% Student Papers: 7%

2% match (student papers from 04-Apr-2018)
Class: Article 2018
Assignment: Journal Article
Paper ID: 940891790

1% match (Internet from 26-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/8480/181-35-2297.pdf?isAllowed=y&sequence=1

1% match (Internet from 26-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/8580/181-35-2340.pdf?isAllowed=y&sequence=1

1% match (Internet from 20-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/6539/131-15-2400%20%2813_%29.pdf?isAllowed=y&sequence=1

1% match (Internet from 20-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5718/171-35-1958%20%2824_%29.pdf?isAllowed=y&sequence=1

1% match (student papers from 09-Apr-2018)
Class: Article 2018
Assignment: Journal Article
Paper ID: 943495148

< 1% match (Internet from 20-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/3514/P13614%20%288%25%29.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 26-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/7150/171-15-9567%2c%2023%25.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 20-Nov-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5693/171-35-1870%20%2824_%29.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 26-Oct-2022)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/8407/181-35-2376.pdf?isAllowed=y&sequence=1

< 1% match (Internet from 01-Oct-2021)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5447/192-25-789%20%2825%25%29.pdf?

LIBRARY CLEARANCE