

Fruit freshness detection by using TensorFlow

BY

MD. Asrafi Rahoman Tonmoy
ID:162-15-7899

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
Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised By

Mr. Gazi Zahirul Islam
Assistant Professor
Department of CSE
Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2020

APPROVAL

This Project titled “**Fruit freshness detection by using TensorFlow**”, submitted by MD. Asrafi Rahoman Tonmoy, ID:162-15-7899 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **JULY 2020**

BOARD OF EXAMINERS



Dr. Syed Akhter Hossain
Professor and Head

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

Chairman



Dr. Sheak Rashed Haider Noori
Associate professor & Associate Head

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

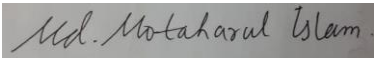
Internal Examiner



Md. Zahid Hasan
Assistant Professor

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

Internal Examiner



Dr. Md. Motaharul Islam
Professor

Department of Computer Science and Engineering
United International University

External Examiner

DECLARATION

I hereby declare that, this project has been done by us under the supervision of **Ms. Nazmun Nessa Moon, Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



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

Co- Supervised by:



Mr. Gazi Zahirul Islam
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



MD. Asrafi Rahoman Tonmoy
ID:162-15-7899
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project successfully.

I really grateful and wish my profound my indebtedness to Ms. **Nazmun Nessa Moon, Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of AI to carry out this project. His 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.

I would like to express my heartiest gratitude to the Almighty Allah and Head, Department of CSE, for his kind help to finish my project and also to other faculty member and the staff of CSE department of Daffodil International University.

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

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

ABSTRACT

We are human being. We need food to live. Because, we get nutrition from food. Most of the nutrition is taken from fruits. Fruit is the source of huge amounts of vitamins and minerals. It also improves our immune system and make us strong. The fruit juice is also delicious and healthy for us. But the rotten fruit and the juice are made from it are unhealthy poisonous to us. Now-a-days people are buying rotten fruits and businessmen are also sold rotten fruits or use them to make juice without knowing. So I think about this and try to do something. Now everything became smart and technology dependent. We can see the use of AI and machine learning everywhere. Following this, I built here a system. A system that is smart, fast and that system can identify fresh and rotten fruits. I use here python CNN model and tensorflow library to classifying the rotten fruit and fresh fruit. This system can identify the percentage of rotten and freshness of fruit by using image data. This system can make our work easy with big farm and juice factory to identify rotted fruit. We hope that it will make our work more accurate in less time.

TABLE OF CONTENTS

| CONTENTS | PAGE NO |
|--|---------|
| APPROVAL | i |
| DECLARATION..... | ii |
| ACKNOWLEDGEMENT..... | iii |
| ABSTRACT | iv |
| LIST OF FIGURES | v |
| LIST OF TABLES..... | vi |
| CHAPTER 1: INTRODUCTION..... | |
| 1.1. Introduction | 1 |
| 1.2. Motivation | 1 |
| 1.3. Rationale Of The Study | 2 |
| 1.4. Research Questions | 2 |
| 1.5. Expected Outcome | 2 |
| 1.6. Report Layout..... | 3 |
| CHAPTER 2: BACKGROUND..... | |
| 2.1. Introduction | 4 |
| 2.2. Related Work..... | 4 |
| 2.3. Comparative Analysis and Summary | 5 |
| 2.4. Scope Of The Problem..... | 6 |
| 2.5. Challenges | 7 |
| CHAPTER 3: RESEARCH METHODOLOGY..... | |
| 3.1. Introduction | 8 |
| 3.2. Research Subject And Instrumentation..... | 8 |
| 3.3. Data Collection Procedure | 8 |
| 3.4. Statistical Analysis | 10 |

| | |
|--|-----------|
| 3.5. Implementation Requirements | 11 |
| CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION | |
| 4.1. Experimental Setup | 13 |
| 4.2. Model Summary | 13 |
| 4.3. Experimental Results and Analysis | 14 |
| 4.4. Discussion | 16 |
| CHAPTER 5: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATIONS FOR FUTURE RESEARCH | |
| 5.1. Summary of the Study | 17 |
| 5.2. Conclusion..... | 17 |
| 5.3. Recommendation..... | 17 |
| 5.4. Implication for future study..... | 18 |
| REFERENCES | 19 |
| APPENDICES | 20 |

LIST OF FIGURES

| FIGURE NAMES | PAGE NO. |
|---|----------|
| Fig 3.1: Data labeling by Bulk Rename Utility..... | 09 |
| Fig 3.2: Methodology at a Glance..... | 11 |
| Fig 4.1: The CNN model used in my project..... | 14 |
| Fig 4.2: Training accuracy vs validation accuracy And Training loss vs validation loss..... | 15 |

LIST OF TABLES

| TABLES NAMES | PAGE NO. |
|---|-----------------|
| Table: 3.1 Train image data amount..... | 11 |
| Table: 3.2 Test image data amount..... | 11 |
| Table: 4 Fresh vs rotten accuracy..... | 14 |

CHAPTER 1

INTRODUCTION

1.1 Introduction

Image classifier is an important study for exploring artificial intelligence and deep learning study. It is one of the most used technologies and research topic to solve our problems about daily life issues and facts. Every day in our life we are facing different kinds of problems which can be easily tackle and solved by using artificial intelligence and classifiers. Such a problem like knowing which fruit is good or rotten which food is good for our health or bad even we can find which disease the fruit or the vegetable have been affected and should we eat it or not. To solve this problem, I took a problem which is knowing which fruit is good and fruit which is rotten. I will use artificial intelligence and algorithms search a CNN and regression algorithms to find whose Apple any kind of fruit is fresh and which one is rotten. Fruits in super shop are often not fresh but because of the lighting condition we can understand this. By using my system anyone can know that by just scanning fruit and it will be a charm for them for saving the money and buying the fresh fruit from the mall shops.

1.2 Motivation

Nowadays lots of people are getting bad fruits and vegetables from shopping malls and super shops without knowing it is good or it is bad. Because it is time consuming fact testing food one by one they never check fruits before buying. As a result of this fact many people are getting bad fruit every day and they have to eat that with the risk of injuring their body and health. So my main motivation was to specify this problem and fix it. If they have a system which will able to

scan food in few second they will always test and scan fruits before buying. So that the chance of getting bad fruits will be decreased and their health and money both will be preserved. So this research is focused to minimize this problem and creating a system with image processing which will make user able to scan fast and get the result.

1.3 Rationale of the Study

It is no doubt there are lots works at the Convolutional neural network (CNN) in image classifier. I can see many works with images, some of them can find cancer cell or some of them can detect disease in leaf. But working with the Google Tensorflow library is more efficient and reliable. It can show accurate output with less training data. This made me to be interested to work with this.

At the present time, I can see the use of AI everywhere and we can also see the crisis with the food fresh. So I decide to do something new to detect fresh fruit and rotten fruit with the help of AI.

1.4 Research Questions

- Can I detect the freshness or rotten percentage?
- Can I use it on our mobile phone?
- Which type of people will get the benefit?
- How much accuracy can we get?

1.5 Expected Output

Expected outcome of this research based project is to build a system to help the people who need the fruit freshness detector.

1.6 Report Layout

In **chapter 1**, all about this project is written here. The reason of choosing this project, how will this project be completed, project motivation, expected outcome and so on is discussed briefly.

In a word, chapter 1 is elaboration of introduction of this project.

In **chapter 2**, related works on this area which were studied are showed. Their findings and limitations are summarized and hence the scope and challenges of the research are also mentioned.

In **chapter 3**: research methodology will discusses Research Subject and Instrumentation, Data Collection Procedure, Statistical Analysis and Implementation Requirements

In **chapter 4**: experimental results and discussion Experimental Results and Descriptive Analysis

In **chapter 5**: presents a short conclusion. And list of reference

CHAPTER 2

Background

2.1 Introduction

In this chapter, I will discuss related research or project about image classifier or related to detect something with image data. In the first section I will discuss about previous related work, then in the second section I will show the outcome or a summary of my study of the related work and then I will discuss about the benefits and challenges that I face to do this project.

2.2 Related Works

In a research done by Deepika Jaswal, Sowmya.V, K.P.Soman [1] in Image Classification Using Convolutional Neural Networks classified different kinds of image, scene and environments. They used CNN layers to make the classifier to do the classification. They proved that CNN can work with both face and non-face classification, building and nature classification, dense and rational classification. They used 250 epoch in their classification. They got about 77% accuracy in forest and agricultural classification and about 91% of accuracy in residential vs agricultural classification. They got about 97% of accuracy by working with green regions. They also proved that CNN gives the best accuracy in the green scene rather than the building scene. So it works not only with material or building images but also with nature and fruits. So from this research we can ensure that CNN is a great algorithm to work with nature and nature products like fruit and vegetables classification.

A research done by Bingquan Huo and Fengling Yin [2] on the topic Research on Novel Image Classification Algorithm based on Multi-Feature Extraction and Modified SVM Classifier. In their research they worked with image object classifiers and SVM, Kernel and optimization. They used a common vector machine(SVM), K-neighbours neural networks and random forest to make the model in order to classify images. They found that SVM has the best classification result and

reduced amount of training time etc. They successfully overcome the most common problem in computer vision classification and object detection by their proposed model. But their accuracy and result accuracy was not very high in their proposed model. So I can decide from the research that CNN will work better than SVM in scene and nature image classification.

2.3 Comparative Analysis and Summary

After researching some research papers and projects I decided to go with CNN(Convolutional Neural Network) because,

- It works best with nature scene detection, green object detection and classification, environment and forest classification. Even it works best with the fruit which we all need.
- CNN has the best accuracy among others image classification algorithm with the accuracy of 90% or above with proper training and labelling,
- It's easy to use and lots of resources to work with in further development
- It also works best in comparing images like fresh fruit and rotten fruit. Face and non face etc.
- I can obtain good results and accuracy by using CNN layers properly and with proper training.

As I decided to use CNN (Convolutional Neural Network) as my main classification model I will work with CNN layers and deep learning algorithms. I used tensorflow and keras in backend to implement model. My main goal was using an in house database and anaconda. But because of the wide use colab and easy, fast implementation way I decided to go with google colab. I used colab and google own GPU in runtime to make the best out of it. As a result I had to use mount drive for database use using google drive.

My main goal is making a classifier of fresh and rotten fruits and further comparing them in order to get the output. So I used the Pooling layer, Con2d and activation layers for creating CNN layers. I worked with apple, banana and other easy to use objects as in first. I can not use hard to detect object classifiers which may lead us to bad accuracy and output. By using apple, banana and mango I will be able to get good accuracy range in order to find the best comparison possible among their good to rotten range values. As if I have a good range between good and bad then I will be able to say how much bad or good that fruit is and is it edible or not.

2.4 Scope of the Problem

The main focus of this research work is making e system that can identify freshness of the fruit and proved that CNN is the best algorithm model for image classifier. In future I will try to use my system in big factories, supper shop or fast-food shop to ensure healthy foods.

I will make my system open source and freely available to all. So that anyone like magistrate or street food inspector of Bangladesh can use this system to identify rotten fruit or food. It will make their work easier. For my system normal people also can identify the rotten food or can send rotten food images to govt server and magistrate can understand that which area need to observe.

2.5 Challenges

1) Data Collection

Since there are available image data in online, data collection had not that much difficult for this research project. But when I start collecting mangoes image data locally it become to much difficult because most of the seller did not like that we touch theirs products without any reason like buying. For more accuracy I need more train data. But I was not able to to buy huge amount of fruits. Then I decide to collect data from **Kaggle**.

2) Model selection

Model selection is a quiet tough task for any researcher. Because research project success depends on Data set and Model selection. The right choice will lead you at your goal quickly and the wrong one will ruin it. I test different types of model with the test data and try to find the best one. I also try to work with Matlab. After trying everything I choose a CNN algorithm because Google already did something great to make our work easy. When I see that Google built a library like Tensorflow_for this type of images classifier work and I can work with Google CoLab when working with images need an expensive type of GPU, but Google provides free virtual GPU for my work I decided to work on this model.

3) Data labeling

Data labeling was the most important part of my work. Because it make code more faster to run and less time consuming for data training. I use **Bulk Rename Utility** to labeling my data.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In in this session, I will describe about my research methodology and procedures. Moreover tools for the research project, data collection, research topic, pre-processing, processing, statistical analysis and its implementation will be discusses in this session.

3.2 Research Subject and Instrumentation

I notice that data is the main part of the research. It is a very critical part for a researcher to find out perfect data and perfect algorithm or model for his research work. He/She also need to study about related research papers. Then he/She should need to make several decisions:

- Which data should collect?
- How to ensure that collected data are ok?
- How should each data be organized?
- How should each data be labeled?

3.3 Data Collection Procedure

In experiment, I used cats vs dogs image dataset. These data manly selected because those are available label data and for my project I need labeled data. I used their 5000 cats and dogs images as training data and 2500 images as testing data

1) Data Pre-processing

Data pre-processing refers to the pre phase of processing datasets. Generally raw data sets are not able to perform operations and generate expected outcome. As a result, data pre-processing is required. And it is considered to be one of the most important part of research. In this phase I collect more then 20000 image data from different sources and try to unnecessary or noisy data.

2)Data Organizing

In this phase I divided data and store them in two data folder test and train, I also use here validation folder for check train data validation. Then I divided those test and train folder's data in more folder like freshapple, rottenapple, freshbanana or rottenbanana etc.

3)Labeling Data

In this phase I renamed all images as their name and also numbered them sequentially. I use here a software to labeling and the name of the software is 'Bulk Rename Utility'.

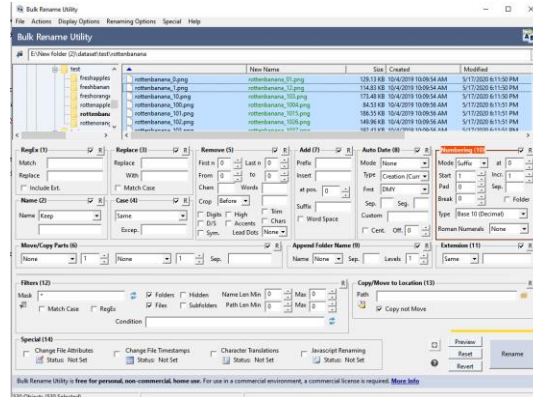


Fig 3.1: Data labeling by Bulk Rename Utility

4)Data Storing

In this part I store all the data in google drive because it make our work easier in google drive.

I can use those online stored data in my project by following some simple step or code.

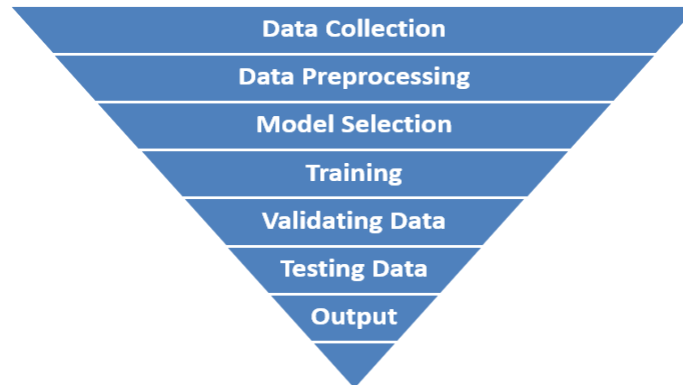


Fig 3.2: Methodology at a Glance

3.4 Statistical Analysis

The amount of my total image data is more then 16,000 that I collect but after preprocessing I get total 13,599 data. The accurate data amount are given below,

Table 3.1: Train image data amount

| Fruits Name | Amount |
|--------------------|---------------|
| Fresh Apple | 1,693 |
| Fresh Banana | 1,581 |
| Fresh Oranges | 1,466 |
| Rotten Apple | 2,342 |
| Rotten Banana | 2,224 |
| Rotten Orange | 1,595 |

Table 3.2: Test image data amount

| Fruits Name | Amount |
|--------------------|---------------|
| Fresh Apple | 395 |
| Fresh Banana | 381 |
| Fresh Oranges | 388 |
| Rotten Apple | 601 |
| Rotten Banana | 530 |
| Rotten Orange | 403 |

3.5 Implementation Requirements

- **Python 3.8**

Python 3.8 is Python version. It is a high level programming language. Most of the research use it to do their research. It is highly recommended programming language for AI based work and it is very popular among new generation's programmer because it is very easy to learn and understand.

- **Google CoLab**

Google CoLab is free to use open source distributor of Python programming language. We can work here online in our browser as like as Jupiter notebook , but the main benefit of this Google CoLab is , it provide us free online virtual GPU access.

- **Bulk Rename Utility**

It is a very useful windows software to labeling any kind of data as you want. It is free and easy to use.

- **Hardware/Software Requirements**

- Operating System (Windows 7 or above)
- Web Browser(preferably chrome)
- Hard Disk (minimum 4 GB)
- Ram(more than 4 GB)

CHAPTER 4

Experimental Results and Discussion

4.1 Experimental Setup

In my model implementation and code implementation I have collected the data first. Procedure given below,

- As I worked with freshness detection (rotten / good) I had to collect both good and bad fruit images from different areas
- For a larger part of my project I have collected images from the shop, bazar and home by taking pictures.
- I also scrap the internet to find out more images about this topic and I also collected data from online databases like kaggle.
- Then I have finalized and normalized the data in order to perform my training.
- After labelling data data was usable and good for further processes.

After that I had to preprocess my image in below steps,

- First I cropped all images with same dimensions
- Then I had to resize and reshape all image in order to normalize dataset
- Then used a pooling layer to remove all negative pixels to make final usable images.

4.2. Model Summary

In my model I used a sequential model with layers of conv2d, maxPooling2 then finally a flatten layer. There are three CNN layers. Finally I wrapped these layers with a flatter layer and a dense layer. As a result we will get accuracy with my model. From other research I found that three layers is best for creating a CNN model. By adding a dense and a flatten layer we can optimise the model and get best output from the model with best accuracy.

Model: "sequential_7"

| Layer (type) | Output Shape | Param # |
|-------------------------------|----------------------|----------|
| conv2d_21 (Conv2D) | (None, 150, 150, 16) | 448 |
| max_pooling2d_21 (MaxPooling) | (None, 75, 75, 16) | 0 |
| conv2d_22 (Conv2D) | (None, 75, 75, 32) | 4640 |
| max_pooling2d_22 (MaxPooling) | (None, 37, 37, 32) | 0 |
| conv2d_23 (Conv2D) | (None, 37, 37, 64) | 18496 |
| max_pooling2d_23 (MaxPooling) | (None, 18, 18, 64) | 0 |
| flatten_7 (Flatten) | (None, 20736) | 0 |
| dense_14 (Dense) | (None, 512) | 10617344 |
| dense_15 (Dense) | (None, 1) | 513 |

Total params: 10,641,441
Trainable params: 10,641,441
Non-trainable params: 0

Fig 4.1:The CNN model used in our project

4.3 Experimental Results & Analysis

After I ran my dataset and created the model from the given data, I found the output what I wanted. I compared fresh fruit with fresh fruit which gave about 90% accuracy and I also compared fresh vs. rotten fruits which gave about 49%-68% of matching accuracy. The final output chart has given below in table 1

Table 4.1: Fresh vs rotten accuracy

| Fresh vs. Fresh | Fresh vs. Rotten | Rotten vs. Rotten |
|-----------------|------------------|-------------------|
| 89%-100% | 49%-68% | 89% |

So from the table it is clear that when we get an output comparison value about 49% it's the worst stage and 68% it's still edible. When it's above 69 it's good and when it is above 80% we can say it is fresh and healthy.

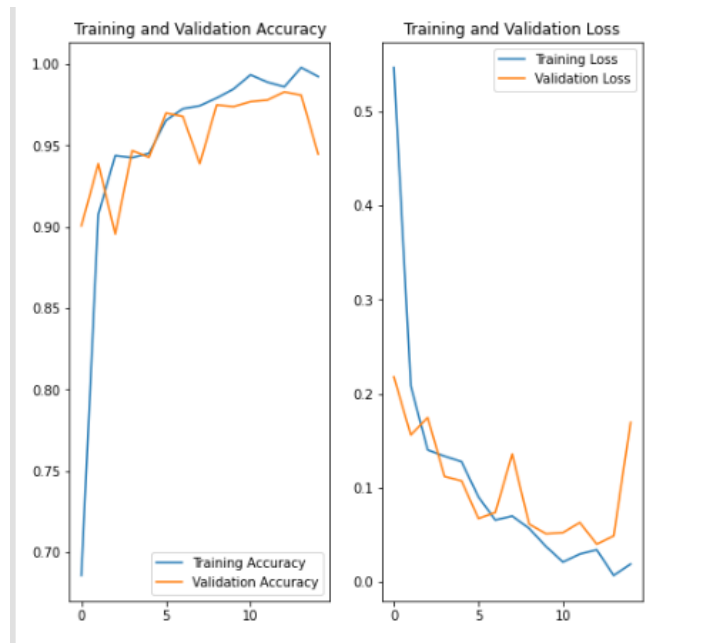


Fig 4.2: Training accuracy vs validation accuracy
And Training loss vs validation loss

4.4 Discussion

After revising my model and dataset I can say that this research has concluded that this classifier can be used in any kind of comparing dataset to compare and predict the accuracy among them. I successfully defined the accuracy of 89% of predicting rotten and fresh fruits (apple). It is also able to differentiate between fresh apple and rotten apple in a range of 49 to 68%. So, we will be able to predict the range of rotten and how much it is from starting very bad at 49% and to mild rotten at 68%.

CHAPTER 5

Summary, Conclusion, Recommendation and Implication for Future Research

5.1 Summary of the Study

The research project is developed for finding differences between fresh fruit and rotten fruit and prove that CNN is the best image classifier algorithm. For this research I needed image data. So data sets are collected from different online and offline sources. After that data pre-processing rules are maintained for all the data to make them compatible with the systems environment. Data sets are trained for data handling purpose. Here some of data sets are trained data sets and some are test data sets.

5.2 Conclusions

In this era of technology data bought us new opportunity as well as new complexity. Handling new data require new method, sometimes new technology. Finding bad and rotten food or fruit one by one shop for govt food inspector is very difficult and time consuming. It is also difficult and waste of time for juice factory and supper shop owner to identify bad products. So this simple system can solve all of those problems. This system can be a simple but efficient.

5.3 Recommendations

Several adaptive algorithms may develop to identify objects from image data. Because our future is AI based and image data in machine learning is very important for AI sector. It make our machine or technology more productive. So everyone should need to work with image classifier. It may change the entire concept what you have been understood earlier.

5.4 Implication for Further Study

- To make this more efficient I shall make more data
- I can use more type or class of data .
- In future I will build a complete open source working platform with huge amounts of image data.

REFERENCES

- I. Deepika Jaswal, Sowmya.V, K.P.Soman – "Image Classification Using Convolutional Neural Networks",-2014
- II. Bingquan Huo and Fengling Yin-"Research on Novel Image Classification Algorithm based on Multi-Feature Extraction and Modified SVM Classifier",-2015
- III. Samer Hijazi, Rishi Kumar, and Chris Rowen, IP Group, Cadence,-" Using Convolutional Neural Networks for Image Recognition".
- IV. An introduction to convolutional neural networks [Online]available at:http://white.stanford.edu/teach/index.php/An_Introduction_to_Convolutional_Neural_Networks
- V. Tutorial on deep learning [Online] available at : <http://deeplearning.net/tutorial/lenet.html>
- VI. Yi Yang and Shawn Newsam, "Bag-Of-Visual-Words and Spatial Extensions for Land-Use Classification",ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), 2010
- VII. <https://www.kaggle.com/>

APENDIX RESEARCH REFLECTION

During project activities, I faced several problems. But two problems were major among them one is selecting best algorithm and another is data collecting. Before working with CNN algorithm I tried many ways to solve my problem and I failed to get best output. To collecting data I faced too much difficulty. Because locally data collection is time consuming and hard for student. Country like Bangladesh, people didn't take it positively. So, I have to collect all the required data from online sources. And after a long time and a lot of attempts and hard work I got succeed.

Plagiarism Report

ORIGINALITY REPORT

| | | | |
|------------------|------------------|--------------|----------------|
| 7 % | 6 % | 2 % | % |
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|----------|--|----------------|
| 1 | dspace.daffodilvarsity.edu.bd:8080 Internet Source | 3 % |
| 2 | www.tensorflow.org Internet Source | 2 % |
| 3 | www.sersc.org Internet Source | 1 % |
| 4 | www.ee.ucla.edu Internet Source | <1 % |
| 5 | www.micansinfotech.com Internet Source | <1 % |
| 6 | R. Punitha Lakshmi, Melingi Sunil Babu, V. Vijayalakshmi. "Voxel based lesion segmentation through SVM classifier for effective brain stroke detection", 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017 Publication | <1 % |
