Rice leaves disease detection by using TensorFlow

BY

Tasnim Hasan Chowdhury
ID:162-15-8094

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

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

7th October 2020

APPROVAL

This Project titled "Rice leaves disease detection by using TensorFlow''

, submitted by Tasnim Hasan Chowdhury, ID:162-15-8094 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 7th October 2020.

BOARD OF EXAMINERS

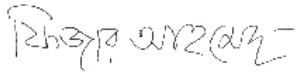
Dr. Syed Akhter Hossain **Chairman**

Professor and Head

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University



Dr. Fizar Ahmed **Internal Examiner**

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Abdus Sattar **Internal Examiner**

Assistant Professor

Department of Computer Science and Engineering

Faculty of Science & Information Technology

Daffodil International University

Dr. Mohammad Shorif Uddin **External Examiner**

Professor

Department of Computer Science and Engineering

Jahangirnagar University

DECLARATION

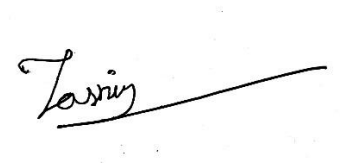
I hereby declare that, this project has been done by us under the supervision of **Ms. Nazmun Nessa Moon, Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree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

Supervised by:

_____

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

Submitted by:



_____

Tasnim Hasan Chowdhury
ID:162-15-8094
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

This research work focused on automatic detection method for image analysis on rice leaves under wide range of natural condition for further analysis. In image processing the syndrome is an essential part for feature eradication and regulation. However, some of the threats are still flawed to predict the inflammation. To meet those threats, the expected algorithm focuses on an exact problem to predict the inflammation from early warning. Bacterial Leaf Blight and Brown Spot are a major bacterial and fungal inflammation respectively in rice (Oryza sativa) crops, it causes harvest loss and reduce the grains quality. Various hybrid techniques for image analysis and regulation algorithms were analyzed and an automatic detection method has been proposed for identifying the exact inflammation in rice leaves under different natural condition. This system can classify the percentage of infected and healthy of rice leaves by using image data. This system can make our work easy with big agro farm to identify infected plants. I hope that it will make our work more classified in less time.

TABLE OF CONTENTS

CONTENTS **PAGE NO**

LIST OF FIGURES

FIGURES **PAGE NO.**

LIST OF TABLES

TABLES **PAGE NO.**

CHAPTER 1

INTRODUCTION

1.1 Introduction

Picture classifier is a critical consider for investigating fake insights and profound learning consider. It is one of the foremost utilized innovations and inquire about subject to solve our issues approximately lifestyle issues and actualities. Each day in our life we are confronting diverse sorts of issues which can be effectively handle and fathomed by using artificial insights and classifiers. Such a problem like knowing which natural product is nice or spoiled which nourishment is nice for our wellbeing or terrible indeed ready to find which malady the natural product or the vegetable have been influenced and ought to we eat it or not. To unravel this issue, I took an issue which is knowing which natural product is nice and natural product which is spoiled. I will utilize counterfeit insights and algorithms search a CNN and relapse calculations to discover whose Apple any kind of natural product is new and which one is spoiled. Natural products in super shop are regularly not new but since of the lighting condition ready to get it this. By utilizing my framework anybody can know that by fair filtering natural product and it'll be a charm for them for sparing the money and buying the new natural product from the shopping center shops.

1.2 Motivation

These days' parcels of individuals are getting terrible natural products and vegetables from shopping shopping centers and super shops without knowing it is nice or it is terrible. Since it is time expending reality testing nourishment one by one they never check natural products some time recently buying. As a result of this truth numerous individuals are getting terrible fruit every day and they need to eat that with the chance of harming their body and wellbeing. So my main motivation was to indicate this issue and settle it. In the event that they have a framework which can able to

filter nourishment in few moment they will continuously test and check natural products some time recently buying. So that the chance of getting awful natural products will be diminished and their wellbeing and cash both will be protected. So this investigate is centered to play down this issue and making a system with picture preparing which is able make client able to filter quick and get the result.

1.3 Rationale of the Study

It is no question there are parts works at the Convolutional neural organize (CNN) in picture classifier. I can see numerous works with pictures, a few of them can discover cancer cell or a few of them can identify malady in leaf. But working with the Google Tensorflow library is more productive and solid. It can appear exact yield with less preparing information. This made me to be interested to work with this.

At the display time, I can see the utilize of AI all over and able to too see the emergency with the nourishment new. So I choose to do something new to identify new natural product and spoiled natural product with the assistance 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 the diseases 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 various leaves 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 rice leaves I will be able to get good accuracy range in order to find the best comparison possible among their healthy to infected 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 plant 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 the disease of the plant and proved that CNN is the best algorithm model for image classifier. In future I will try to use my system in big farms, fields or agro institute to ensure healthy plants.

I will make my system open source and freely available to all. So that anyone like magistrate or agriculture inspector of Bangladesh can use this system to identify healthy or infected plants. It will make their work easier. For my system normal people also can identify the infected plants or can send infected plants images to government 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 leaves image data locally it become too 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 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 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 this section, 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 disease vs. disease image dataset. These data manly selected because those are available label data and for my project I need labeled data. I used their 2684 images as training data and 671 images as validation 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 than 6000 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 validation and train, I also use here validation folder for check train data validation. Then I divided those validation and train folder's data in more folder like Brown spot, Hipsa, leaf Blast etc.

3) Labeling Data

In this phase I renamed all images as their name and also numbered them sequentially. I use here (Fig 3.3.1) a software to labeling and the name of the software is 'Bulk Rename Utility'.

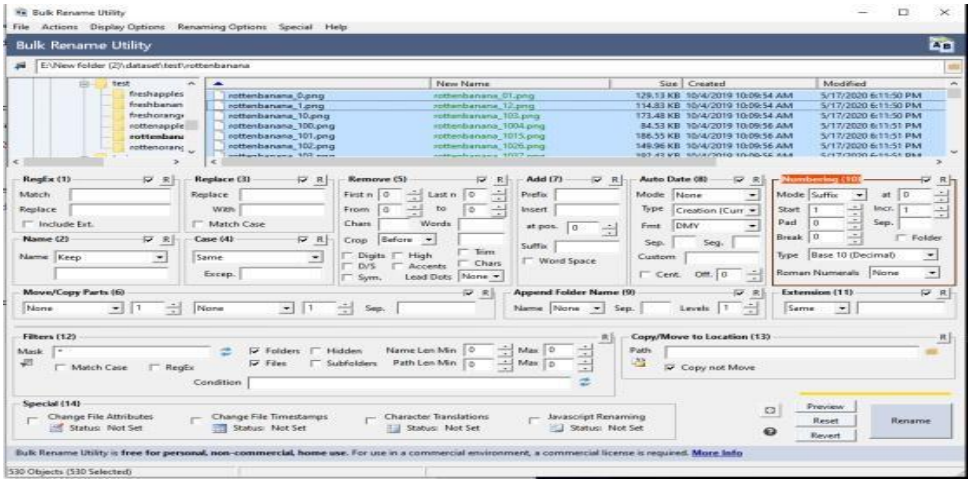


Fig 3.3.1: Data labeling by Bulk Rename Utility

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 which is shown in fig 3.3.2.

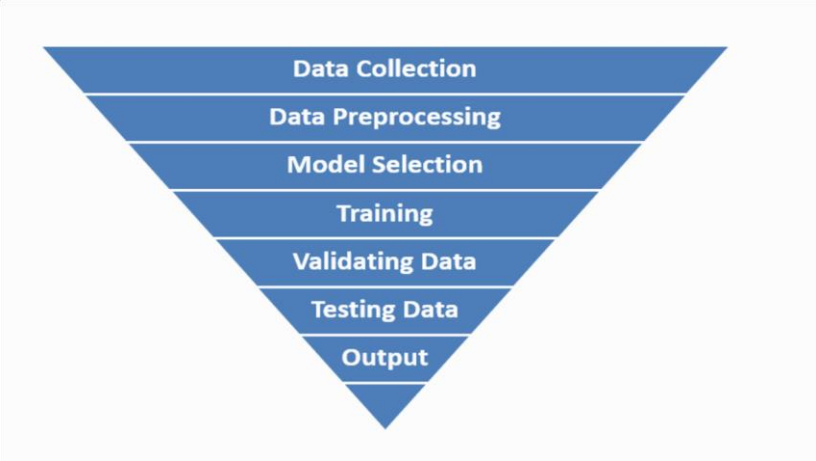


Fig 3.3.2: Methodology at a Glance

3.4 Statistical Analysis

The amount of my total image data is more than 7000 that I collect but after preprocessing I get total 6709 data. The accurate data amount are given below, we see train data's in table 3.4.1 & validation data's in table 3.4.2.

Table 3.4.1: Train image data amount

Disease Name	Amount
Healthy Leaves	1,191
Brown Spot	418
Hipsa	452
Leaf Blast	623

Table 3.4.2: Validation image data amount

Disease Name	Amount
Healthy Leaves	297
Brown Spot	105
Hipsa	113
Leaf Blast	156

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 programing 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 & 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 disease detection (bacterial leaf blast, Hipsa, brown spot) I had to collect both good and bad leaf images from different areas
- For a larger part of my project I have collected images from the farm, field and botanical research center 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 optimize the model and get best output from the model with best accuracy Which we can see in fig 4.2.1.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
keras_layer_1 (KerasLayer)   (None, 2048)              21802784
_________________________________________________________________
flatten (Flatten)            (None, 2048)              0
_________________________________________________________________
dense (Dense)                (None, 512)               1049088
_________________________________________________________________
dropout (Dropout)            (None, 512)               0
_________________________________________________________________
dense_1 (Dense)              (None, 3)                 1539
=================================================================
Total params: 22,853,411
Trainable params: 1,050,627
Non-trainable params: 21,802,784
_________________________________
```

Fig 4.2.1: 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 specific disease with specific disease which gave about 90% accuracy and I also

compared one disease vs. another disease leaves which gave about 60%-79% of matching accuracy. The final output chart has given below in table 4.3.1

Table 4.3.1: Disease vs. Disease accuracy

One Disease Vs. Another Disease	Specific Disease vs specific Disease
60%-79%	80%-100%

So from the table it is clear that when we get an output comparison value about 60% it’s the worst stage and 79% it’s still edible. When it's above 80 it's good and when it is above 80% we can say it is fresh and healthy. The accuracy & loss graph is shown in the fig 4.3.1 is given below.

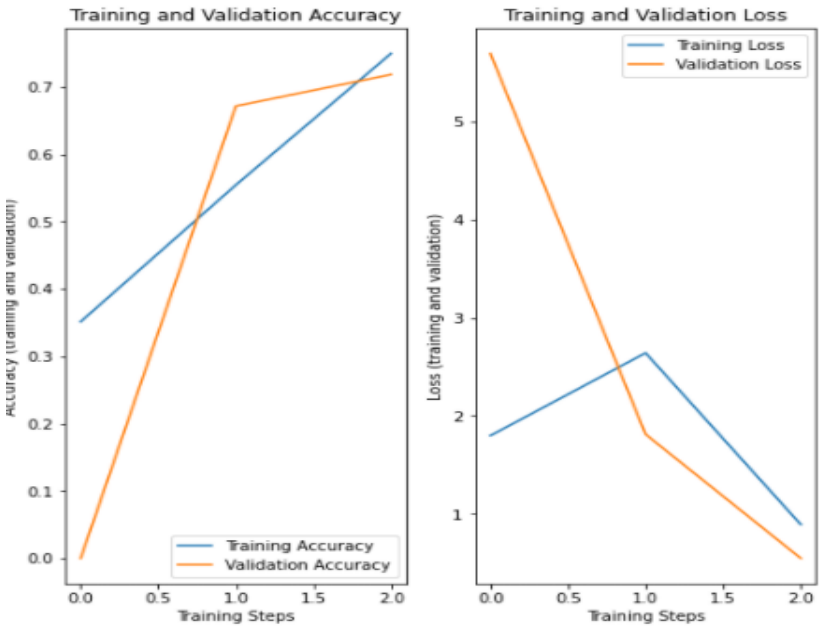


Fig 4.3.1: Training accuracy vs. validation accuracy And Training loss vs. validation loss

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 specific disease and specific disease (brown spot). It is also able to differentiate between one disease and another disease in a range of 60% to 79%. So, we will be able to predict the range of infected and how much it is from starting very bad at 60% and to mild rotten at 79%.

CHAPTER 5

SUMMARY, CONCLUSION, RECOMMENDATION & IMPLECATION FOR FUTURE RESEARCH

5.1 Summary of the Study

The research project is developed for finding the diseases of rice leaves 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 healthy & unhealthy rice or plants one by one shop for government agricultural inspector is very difficult and time consuming. It is also difficult and waste of time for farmers & owner of the farms to find out inappropriate plants. 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

[1] Deepika Jaswal, Sowmya.V, K.P.Soman – ”Image Classification Using Convolutional Neural Networks”,-2014

[2] Bingquan Huo and Fengling Yin-”Research on Novel Image Classification Algorithm based on Multi-Feature Extraction and Modified SVM Classifier”,-2015

[3] Samer Hijazi, Rishi Kumar, and Chris Rowen, IP Group, Cadence,-” Using Convolutional Neural Networks for Image Recognition”.

[4] An introduction to convolutional neural networks [Online]available at:http://white.stanford.edu/teach/index.php/An_Introduction_to _Convolutional_Neural_Networks

[5] Tutorial on deep learning [Online] available at : http://deeplearning.net/tutorial/lenet.html

[6] 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 ware major among them one is selecting best algorithm and another is data collecting. Before working with CNN algorithm I tried many way 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.

Rice leaf

ORIGINALITY REPORT

12%	12%	5%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	5%
2	www.ijser.org Internet Source	1%
3	Saket Kumar Ojha, Kasiviswanath Challa, Madhu Kiran Vemuri, Naga Sri Vardhan Yarlagadda, B L N Phaneendra Kumar. "Land Use Prediction on Satillite images using Deep Neural Nets", 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019 Publication	1%
4	www.sersc.org Internet Source	1%
5	www.ijcaonline.org Internet Source	1%
6	tensorflow.classcat.com Internet Source	1%
7	A Vinay, Aditya Lokesh, Vinayaka R Kamath, K	