Faculty of Engineering

PADDY PLANT (ORYZA SATIVA) DISEASE DETECTION BASED ON IMPROVED CONVOLUTION NEURAL NETWORK

Marceila Suzie Anak Ambrose

Bachelor of Engineering

Electrical and Electronics Engineering with Honours

2023

UNIVERSITI MALAYSIA SARAWAK

Grade: ____________

Please tick (√)

Final Year Project Report ☑

Masters ☐

PhD ☐

DECLARATION OF ORIGINAL WORK

This declaration is made on the ...26...day of...July...2023.

Student's Declaration:

I Marceila Suzie Anak Ambrose

(PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled Paddy Plant (Oryza Sativa) Disease Detection Based On Improved Convolution Neural Network is my original work. I have not copied from any other students' work or from any other sources except where due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.

26/7/2023
Date submitted

Marceila Suzie Anak Ambrose (72317)
Name of the student (Matric No.)

Supervisor's Declaration:

I Dr Annie Anak Joseph (SUPERVISOR'S NAME) hereby certifies that the work entitled Paddy Plant (Oryza Sativa) Disease Detection Based On Improved Convolution Neural Network (TITLE) was prepared by the above named student, and was submitted to the "FACULTY" as a * partial/full fulfillment for the conferment of Bachelor of Engineering in Electrical and Electronics Engineering with Honours (PLEASE INDICATE THE DEGREE), and the aforementioned work, to the best of my knowledge, is the said student's work.

Received for examination by: Dr Annie Anak Joseph
(Name of the supervisor)

Date: 26/7/2023

I declare that Project/Thesis is classified as (Please tick (√)):

☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*
☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*
☑ **OPEN ACCESS**

Validation of Project/Thesis

I therefore duly affirm with free consent and willingly declare that this said Project/Thesis shall be placed officially in the Centre for Academic Information Services with the abiding interest and rights as follows:

- This Project/Thesis is the sole legal property of Universiti Malaysia Sarawak (UNIMAS).
- The Centre for Academic Information Services has the lawful right to make copies for the purpose of academic and research only and not for other purpose.
- The Centre for Academic Information Services has the lawful right to digitalise the content for the Local Content Database.
- The Centre for Academic Information Services has the lawful right to make copies of the Project/Thesis for academic exchange between Higher Learning Institute.
- No dispute or any claim shall arise from the student itself neither third party on this Project/Thesis once it becomes the sole property of UNIMAS.
- This Project/Thesis or any material, data and information related to it shall not be distributed, published or disclosed to any party by the student except with UNIMAS permission.

Student signature marceila suzie
(Date) 26/7/2023

Supervisor signature:
(Date) 28/7/23

Current Address:
No.1, Flat Betik, Ibu Pejabat Polis Daerah Sibu, Jalan Tun Abang Haji Openg, 96000, Sibu, Sarawak

Notes: * If the Project/Thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach together as annexure a letter from the organisation with the period and reasons of confidentiality and restriction.

[The instrument is duly prepared by The Centre for Academic Information Services]

PADDY PLANT (ORYZA SATIVA) DISEASE DETECTION BASED ON IMPROVED CONVOLUTION NEURAL NETWORK

Paddy Plant (Oryza Sativa) Disease Detection Based On Improved Convolution Neural Network

MARCEILA SUZIE ANAK AMBROSE

A dissertation submitted in partial fulfillment
of the requirement for the degree of
Bachelor of Engineering
Electrical and Electronics Engineering with Honours

Faculty of Engineering

Universiti Malaysia Sarawak

2023

ACKNOWLEDGEMENT

I thank God the Almighty for His loyalty and grace in giving me the wisdom, strength, and health to finish this report within the given time. First, I would like to express my special thanks of gratitude to my supervisor, Dr Annie Anak Joseph for her continuous guidance, ideas and advises given during the process of completing this project. Her advises had helped me in every direction to finish this project successfully.

Besides, I would also like to express my gratitude towards my family members for their infinite love, support, and encouragement in completing my undergraduate studies in UNIMAS. Last but not least, special thanks to my fellow course mates who are willingly helped me out with their abilities. I could not express how much gratitude towards those who helped me but sincerely wish all of them best of everything now and then.

ABSTRACT

Paddy plant *(Oryza sativa L.)* plays a pivotal role in ensuring food security in Malaysia and enhancing the agricultural ecological environment. The early detection of paddy plant diseases is very important in the agriculture sector. It helps to pave the way for effective decision-making in overcoming the issues caused by the paddy diseases. Some existing methods, for example, GoogleNet, InceptionNet and VGGNet are used to detect paddy diseases. However, the applied algorithms of machine learning have some drawbacks in terms of time consuming, complex architecture, high computational cost, and the ability of the model to perform with more various types of paddy disease. Therefore, in this project, a hybrid model based on an improved Convolution Neural Network is developed to detect paddy diseases. The combination of the model is between DenseNet-201 and CAE network, and this hybrid model known as CAE-DenseNet. The model is designed, trained, and tested by using MATLAB 2023a. One of the purposes of the combination is to improve the feature extraction technique. Then, by using the large data size with the application of data augmentation, the performance of the model improved. The performance of proposed model CAE-DenseNet is compared with other existing models such as original CAE network, AlexNet, GoogleNet, ResNet-50, VGG-16 and original DenseNet-201. The results analysis shows that the performance of existing models is average because of their limitations to train the data and detected the paddy diseases. The performance of CAE-DenseNet model also evaluated with cross validation and confusion matrix. Overall, the proposed model CAE-DenseNet has a high detection accuracy rate of 88.15%. The accuracy result based on 10-K fold cross validation is 91.11% and the value for precision, recall and F1 score from confusion matrix are 92.03%, 84.70% and 88.21% respectively. Therefore, it proved that the hybrid model CAE-DenseNet able to detect the paddy diseases efficiently.

ABSTRAK

Tanaman padi *(Oryza sativa L.)* memainkan peranan penting dalam memastikan jaminan makanan di Malaysia dan meningkatkan persekitaran ekologi pertanian. Pengesanan awal penyakit tanaman padi amat penting dalam sektor pertanian. Ia membantu untuk membuka jalan dalam membuat keputusan yang berkesan dalam mengatasi isu-isu yang disebabkan oleh penyakit padi. Beberapa kaedah sedia ada, contohnya, GoogleNet, InceptionNet dan VGGNet digunakan untuk mengesan penyakit padi. Walau bagaimanapun, algoritma pembelajaran mesin yang digunakan mempunyai beberapa kelemahan dari segi memakan banyak masa, seni bina yang kompleks, kos pengiraan yang tinggi dan keupayaan model untuk mengesan lebih banyak pelbagai jenis penyakit padi. Dalam projek ini, sebuah model hibrid berdasarkan Rangkaian Neural Konvolusi yang ditambah baik telah dibina untuk mengesan penyakit padi. Gabungan model adalah antara rangkaian DenseNet-201 dan CAE, dan model hibrid ini dikenali sebagai CAE-DenseNet. Model ini direka bentuk, dilatih dan diuji dengan menggunakan MATLAB 2023a. Salah satu tujuan penggabungan model adalah untuk meningkatkan teknik pengekstrakan ciri. Kemudian, dengan menggunakan saiz data yang besar dengan aplikasi penambahan data, prestasi model bertambah baik. Prestasi model CAE-DenseNet yang dicadangkan telah dibandingkan dengan model lain yang sedia ada seperti rangkaian asal CAE, AlexNet, GoogleNet, ResNet-50, VGG-16 dan DenseNet-201 yang asal. Analisis keputusan menunjukkan bahawa prestasi model sedia ada adalah sederhana kerana keterbatasan mereka untuk melatih data dan mengesan penyakit padi. Prestasi model CAE-DenseNet juga dinilai dengan pengesahan silang dan matriks kekeliruan. Secara keseluruhannya, model CAE-DenseNet yang dicadangkan mempunyai kadar ketepatan pengesanan yang tinggi iaitu 88.15%. Keputusan ketepatan berdasarkan pengesahan silang 10 kali ganda adalah sebanyak 91.11% dan nilai untuk ketepatan, imbas semula dan skor F1 daripada matriks kekeliruan adalah masing-masing sebanyak 92.03%, 84.70% dan 88.21%. Oleh itu, model hibrid CAE-DenseNet terbukti dapat mengesan penyakit padi dengan cekap.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

LIST OF ABBREVIATIONS

AE	Autoencoder
BN	Batch Normalization
CAE	Convolutional Autoencoder
CNN	Convolution Neural Network
DCNN	Deep Convolution Neural Network
DenseNet	Densely Connected Network
FC	Fully Connected
FPA	Flower Pollination Algorithm
GPU	Graphics Processing Units
InceptionNet	Inception Network
mAP	Mean Average Precision
MATLAB	Matrix Laboratory
MF	Median Filtering
MLP	Multi-layer Perceptron
OWELM	Optimal Weighted Extreme Learning Machine
RCNN	Recurrent Neural Network
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
RGB	Red,Green,Blue
SSD	Single Shot Multi-Box Detector
SVM	Support Vector Machine
VGGNet	Visual Geometry Group Network

CHAPTER 1

INTRODUCTION

1.1 Research Background

Paddy plant (Oryza sativa L.) is the globe's second most significant food crop after wheat, with the highest paddy production and consumption occurring in Asia [1][2]. Paddy plants play a pivotal role in ensuring food security in Malaysia and enhancing the agricultural ecological environment [3]. As stated by [4], the paddy plant is very important in Malaysia under the food sub-sector for various reasons. First, it is known as the main food for most people [4]. Second, the cultivation of the paddy crop is not only the backbone of Malaysia's economic sector but also for small-scale farmers and landless agricultural workers. However, there are several factors that contribute to the reduction in paddy cultivation in Malaysia and one of them is paddy plant diseases which cause major paddy yield losses. There is also a negative impact on the environment when the number of pesticides or fertilizers used by the farmers is inappropriate due to the lack of knowledge in identifying the accurate types of paddy disease.

In Malaysia, paddy plants are infected by a few devastating diseases such as Hispa rice, brown spot, bacterial leaf blight and rice blast which are caused by a wide range of phytopathogens that include bacteria, viruses, and fungi, resulting in reduction of crop yield in agriculture such as poor productivity and quality of rice production [5]. Thus, the diagnosis of paddy plant diseases at an early stage may guide the way to make decision effectively in handling paddy production. Identifying paddy diseases requires skill and knowledge. In addition, manual observation to determine the type of paddy illness is subjective and consume much time. Occasionally, the disease recognized by farmers or specialists may be deceiving. Therefore, automatic paddy plant disease detection is required as it is one of the implementations of artificial intelligence based on machine learning [6].

According to [7], researchers have used several different approaches of computer vision technology for the diagnosis and detection of crop diseases in agriculture over the past decades. The computer can automatically identify and assess the status of crops in real-time, which can increase production efficiency in comparison when using human labor. Since the leaf is the primary component that displays paddy diseases most strongly, using computer vision to analyze photos of the crop can help discover diseases earlier than manual observation. The conventional algorithm to identify paddy diseases are based on the extraction, color selection, texture, form, and other parameters including the segmentation and edge recognition. However, the accuracy for classification of images task has significantly upgraded because of the rapid development of deep learning. The technique of deep learning is frequently implemented for the detection and classification of paddy diseases as it can achieve the criteria to classify the huge dataset.

Deep learning typically uses Convolution Neural Networks (CNN) for graphical image recognition since it has the advantages of extracting features directly and employed as a classifier. It also uses some operations such as convolution and pooling. Each of the images is treated by the computer with a three-dimensional array along with the length, width, and depth. The operation of CNN allows the computer to work with the images in the algorithm since there is a limitation where the computer is unable to take input data because it can only operate on binary signals. The use of CNN to train image data makes it easier to handle and manipulate the structures of array and easier to carry out the operation of both pooling and convolution. The structure of CNN has been gradually enhanced by the researchers to get better results. Therefore, in recent years, they started to propose a hybrid deep learning model based on an improved CNN to detect crop diseases in a more efficient way.

Some of the previously proposed models are based on the architecture InceptionNet and DenseNet [8]. The basic hybrid block namely IDNet is combining the direct bypass connection of the DenseNet model and the branched architecture of the InceptionNet model. In the basic IDNet block, hybrid activation and pooling operations are also used in the hybrid block. From visualization experiments, this hybrid activation mode shows the hybrid activation mode responds to object semantic regions more flexibly. The results indicate that when using a lower number of parameters, this hybrid model can achieve equal or better detection accuracy. In the paper proposed by [9], a deep transfer learning approach is used in which DenseNet has been trained first on the ImageNet and Inception module are combined to create a new network and this hybrid model is called DENS-INCEP. They make modification for the original DenseNet model by substituting a global pooling layer in its full connection layer. An Inception model and a fully connected ReLU layer with 512 neutrons are then added into the system, followed by a Softmax layer that was fully connected. On the other hand, [10] proposed a method with deep Convolutional Autoencoder (CAE) approach as feature extractor and a Support Vector Machine (SVM) integrating as a classifier. In their research, first step is Autoencoder (AE) reconstructing the original images and then extracting the features of these images by using only the encoder part. The output from extracted features become input for SVM to perform image classification. Another similar work was done by [11]. The authors presented a novel hybrid model of CAE and Convolution Neural Networks (CNN). First, the dimensionality of the input leaf images has been reduced by training the CAE network. Then, CNN classified images into two categories, infected or healthy paddy. All the results achieved by these hybrid models able to proceed with more than 95% accuracy but there are still some limitations that need to be overcome such as the complexity of the model and the training time.

Identification of paddy diseases based on a hybrid model is advantageous as it accelerates the performance of the model to give high accuracy in detecting diseases. However, a more capable and effective hybrid deep learning model-based strategy to identify and categorized the paddy diseases is still required. Based on recent literature studies done, it has indicated that the existing methods currently used to identify plant leaf diseases have some drawbacks in terms of complexity of the model, the learning time and the ability of the model to perform with more various types of paddy disease. Therefore, a new proposed system for identifying plant leaf diseases is suggested in this study, and a comparison of the results obtained with the earlier approaches is described.

1.2 Problem Statement

In the developing countries such as Malaysia, detecting the paddy plant diseases still employs many conventional means. Farmers depend on traditional methods such as virtual inspection of the leaf color patterns and its structures to monitor the paddy plant's health condition. Additionally, manual examination to determine the types of paddy infection is subjective and requires a lot of time. The disease recognized by farmers or specialists may also be deceiving. Hence, this leads to the use of excessive or inappropriate amounts of pesticides or weeds.

Recent research in the field reveals that paddy plant diseases are only moderately well identified by the earlier detection algorithms. It can be demonstrated by the precision and training data for identification and classification generated by the earlier methods. Some of previous methods, for example, AlexNet and Support Vector Machine (SVM) have faster training time but achieved low accuracy while InceptionNet and VGGNet are able to achieve high accuracy but have longer time training. Next, most of the traditional machine learning algorithms are developed in laboratories, and their robustness is insufficient to meet the demands of real-world agricultural applications. The other major disadvantage is a lot of previous research works also used a huge amount number of training parameters. Higher learning time or a powerful computer are needed to train a model with a high number of training parameters.

Therefore, in this research, a hybrid model is proposed to decrease the number of training parameters that eventually speed up the learning time and accelerate the performance of the model in detecting paddy diseases. Other than that, the purpose of the hybrid model is to reduce image noise and the complexity of the model. The aim is also to achieve high accuracy for the detection of paddy diseases. Then, the performance proposed model is evaluated and compared with state to the art methods.

1.3 Research Questions

i. What are the methods used for the detection of paddy plant diseases?

ii. How to detect paddy plant diseases more precisely?

iii. How is the performance of the proposed network in detecting paddy plant diseases?

1.4 Research Objectives

The major purpose for this study is to detect paddy plant diseases based on improved Convolution Neural Network. Specifically, the study discoursed the following objectives:

i. To investigate the previous techniques used to detect the paddy plant diseases.

ii. To develop the hybrid model for the detection of paddy plant diseases based on improved Convolution Neural Network.

iii. To evaluate the performances of the proposed model in detecting the diseases of the paddy plant with the previous methods.

1.5 Hypothesis

The hypothesis of this project is that the proposed network based on an improved Convolution Neural Network can outperform the existing network for the detection of paddy plant leaf diseases. To make this possible, the shortcomings of the prior detection techniques for paddy plant leaf diseases are being discovered. Many research papers are deeply analyzed, enable for the for the application of knowledge regarding the reasons behind the poor and mediocre performance of the earlier techniques. The previous study also helps to create a suitable registering framework for the proposed method. This has implications for model complexity for paddy disease detection, training duration, and accuracy of identification and classification of whether the paddy plants are healthy or not. Therefore, a crucial conclusion from this study is that the performance of the suggested approach may be better than that of the earlier methods. In this project, the hypothesis is done by comparing the performance of the proposed approach and earlier methods.

1.6 Expected Outcomes

In the outcome of this study, through the comparison of the performance between proposed and previous method, it is expected that the proposed network has higher accuracy result compared to the existing methods. Other than that, in term of learning time, it is expected that the proposed network has faster learning time to train the model compared to the previous model. Next, for the proposed model, the error rate in detecting and classifying the paddy diseases correctly into the specific categories also expected to be lower than the error rate existing one. Therefore, the proposed method able to achieve optimum results and perform effectively.

1.7 Project Motivation

The process to detect and identify the type of paddy plant leaf diseases can be completed by using the previous automated disease diagnosis models to categorized paddy plant leaf diseases. The use of an automated model for plant leaf disease detection is advantageous because it requires less time and effort than more time-consuming conventional techniques. This paper surveys many earlier classification techniques for paddy diseases.

In this study, a suitable registering framework has been developed to identify the drawbacks of the earlier detection techniques for paddy plant leaf diseases based on the survey that was conducted for this study. Therefore, the understanding of the shortcomings of the earlier approaches serves as the motivation for this research. The suggested methods attempt to outperform the existing ones in terms of accuracy, time complexity, and fewer training parameters.

Additionally, there are no other research that provide a more thorough comparison of this kind in the recent literature. As a result of this performance comparison of several prior classification methods, the information regarding the involved directions for future research is discovered. Furthermore, the source for the dataset which consists of healthy and infected paddy images used for training the model are acquired from PlantVillage website. Therefore, as the dataset utilized has a synthetic nature comparable to the real-time photos, the findings of this study will be significant to help farmers to control the yield of their paddy crops by detecting devastating diseases.

1.8 Project Scope

In the project scope, the algorithm for detection of paddy plant diseases is developed through the MATLAB software. The dataset of paddy images is acquired from PlantVillage website, and the selected images consist of healthy and infected paddy leaf. There are several main processes involved in detecting and classifying paddy diseases. The first step is image acquisition, then image processing or pre-processing followed by feature extraction and classification. The flow chart for each process is also presented in this report.

1.9 Report Outline

The report consists of five chapters which are the introduction, literature review, and methodology. These chapters in the report are outlined as below:

Chapter 1: Introduction

In this chapter, the research background gives the elaboration of the project introduction and some of the related works. The problem statement describes the issues outlined in existing related works through literature study, moving on to creating objectives based on research questions to be achieved through this project. Then, the motivation of the project is provided.

Chapter 2: Literature Review

In this second chapter, the history of work link to paddy plant diseases are highlighted and described. Based on articles, journals, proceedings, and conference papers, a literature study was conducted on existing methodologies and processes. The review is not only limited to methodology, but a critical evaluation and comparison between materials are provided. Then a literature review concludes all the relatable item provided for evidence.

Chapter 3: Methodology

In this chapter, methodology and techniques used to detect and classify paddy plant diseases are explained. The overall project is being implemented by inspiring project management techniques and skills. The flowchart for the proposed implementation is presented for algorithm development. Then, the Gantt chart was provided to record and report the project's progress on a weekly basis for the whole semester of execution.

Chapter 4: Results and Discussion

In this chapter, the outcomes of the project are stated and further discussed. The comparison of the results between proposed method and the previous method are done to analyze the performance of the model in detecting the paddy diseases. The comparison of the results including the accuracy and learning time of the proposed and existing models. The performance evaluation for both proposed and previous algorithm also done and being compared to achieve the third objectives of this project.

Chapter 5: Conclusion

This chapter discussed and emphasized the conclusions, the limitations, future works, and future recommendations of the project. The chapter summarized up the overall approach and established the framework future research.

Chapter 2

LITERATURE REVIEW

2.1 Overview

This section presents an overview of related and comparable previous research on the detection and classification of paddy diseases. The following is a summary of this chapter where types of plant leaf diseases are discussed in Section 2.2. The effects of plant leaf diseases are discussed in Section 2.3, and the prior approach for identifying plant leaf diseases is discussed in Section 2.4. Section 2.5 is the introduction of some deep learning models. Then, Section 2.6 discuss the previous methods of paddy disease detection while Section 2.7 is the summary of existing methods and fundamental of research gap. Lastly, the chapter is summarized in Section 2.8.

2.2 Paddy Plant Diseases

Pathogenic organisms such as fungi, bacteria, viruses, protozoa, as well as insects and parasitic plants, are the primary causes of paddy diseases. Every species of plant, whether it is in the wild or in a garden, is susceptible to diseases. It is very crucial to identify the infected areas of the plant at the early stage to stop the disease part from spreading to other areas. Typically, the infected part occurred on the leaves, roots, or stems of the plants with leaves is one likely most to be infected by diseases.

Paddy plant leaf diseases come in a variety of forms. For instance, bacterial leaf blight, rice blast, brown spot and Hispa rice. In this project, the dataset of paddy images which consists of infected and healthy paddy are obtained from PlantVillage Image website, one of the largest datasets that collected the images of various types of plant species. Figure 2.1 to Figure 2.4 shows four common diseases that infect the paddy plant.

Figure 2.1: Bacterial Leaf Blight [12]

Bacterial leaf blight is caused by the Gram-negative bacterium *Xanthomonas oryzae pv. oryzae (Xoo)* [12]. This disease usually invades the paddy plant through wounds or hydathode water pores. Then, xylem vessels of paddy leaves are moved and colonized, resulting in tannish grey to white lesions along the veins. Bacterial leaf blight disease can lead to yield reduction of up to 20–50% yield loss and even possibly up to 100%.

According to [12], the symptoms of this disease's early stage can be seen on the leaf blades as soon as tillering reaches its highest position. Typically, the disease starts in the lower portions of plants and subsequently moves up to the upper sections of the plants. Due to the severity of the disease, the above part of the leaf or the whole area of the leaf blade changed to a pale yellow before drying up. In a strong disease assault, yellow to white stripes formed just inside the leaf blade margins. These stripes later turned pale yellow and necrotic.