



**Design and Performance Evaluation of Paddy Leaf Disease
Identification Using KNN Algorithm and Keras Model**

by

**Peter Ling Jie Lung
(1632521955)**

A dissertation submitted in fulfillment of the requirements for
the degree of
Master of Science in Microelectronic System Design Engineering

**Faculty of Electronic Engineering Technology
UNIVERSITI MALAYSIA PERLIS**

2021

ACKNOWLEDGEMENT

First of all I would like to express my gratitude to my Master research project supervisor, Assoc. Prof. Ir. Ts. Dr. Mohd Nazrin Md Isa for his guidance and encouragement throughout the duration of this project.

I am also grateful to Associate Professor Ir. Ts. Dr Rozana Aina Maulat Osman and Ts. Dr. Norhawati Binti Ahmad for the assistance on Master program.

Finally, I would like to express my sincere thanks to my parents and my wife for providing me with unfailing support and continuous encouragement and patient that enable me to complete my Master degree successfully.

©This item is protected by original copyright

TABLE OF CONTENTS

	PAGE
DECLARATION OF DISSERTATION	i
ACKNOWLEDGEMENT	ii
TABLE OF CONTENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF ABBREVIATIONS	x
ABSTRAK	xi
ABSTRACT	xii
CHAPTER 1 : INTRODUCTION	1
1.1 Background	1
1.2 Problem Statements	1
1.3 Objective	3
1.4 Scope	4
1.5 Thesis Organization	4
CHAPTER 2 : LITERATURE REVIEW	6
2.1 Type of Paddy Leaf Diseases	6
2.1.1 Bacterial Blight Disease	6
2.1.2 Blast of Leaf	8
2.1.3 Brown Spot of Rice	10
2.2 Case Study on Existing Image Processing and Recognition Systems	11

2.2.1	Classification of Plant Diseases through leaves by machine learning technique	11
2.2.2	Grape Leaf Disease Identification by Machine Learning Techniques	16
2.2.3	Analysis on Prediction of Plant Leaf diseases using Deep Learning	20
2.2.4	Recognition System of Character on Image by using Tensor Flow	21
2.2.5	Cocoa Ripeness identification system by using K-Nearest Neighbor (KNN) Algorithm System	24
2.2.6	Identification of Leaf Disease Using Image Processing Technique	28
2.3	Summary of related work	31
2.4	Conclusion	33
CHAPTER 3 : METHODOLOGY AND IMPLEMENTATION		34
3.1	Introduction	34
3.2	The Proposed System	34
3.3	Python Language and Google Colaboratory Platform	35
3.4	System Development Flow by Stages	35
3.5	Training Image dataset preparation	37
3.6	Paddy Leaf Disease Identification by using KNN Algorithm Method	37
3.6.1	Planning stage	37
3.6.2	Design stage	38
3.6.2.1	Import Module	39
3.6.2.2	Image Acquisition and Feature Extraction	39
3.6.2.3	KNN Model Generation	40
3.6.3	Testing stage	42
3.7	Paddy Leaf Disease Identification by using Keras Model	43
3.7.1	Planning stage	43
3.7.2	Design stage	44
3.7.2.1	Import Module	44

3.7.2.2	Dataset pre-processing	45
3.7.2.3	Keras model training	46
3.7.3	Testing stage	47
	CHAPTER 4 : RESULTS AND DISCUSSION	49
4.1	Introduction	49
4.2	System Performance Comparison	49
4.2.1	System Performance	49
4.2.2	Sample Test Result	50
4.3	Analysis and Considerations	52
4.4	KNN algorithm test on IOS device	54
	CHAPTER 5 : CONCLUSION	60
5.1	Conclusion	60
5.2	Assumptions and further research	61
	REFERENCES	62
	APPENDIX A PYTHON CODING OF PADDY LEAF DISEASE IDENTIFICATION WITH KNN ALGORITHM	64
	APPENDIX B PYTHON CODING OF PADDY LEAF DISEASE IDENTIFICATION WITH KERAS MODEL	65
	APPENDIX C RESULT OF PADDY LEAF DISEASE IDENTIFICATION SYSTEM BY USING KNN ALGORITHM AND KERAS MODEL	67

LIST OF TABLES

	PAGE	
Table 2.1	Manu Disease Images by types in Dataset	14
Table 2.2	Review of existing Image Recognition and Classification System	32
Table 3.1	Quantity of image samples by each paddy leaf disease type	37
Table 4.1	Summary of System Computational Time by Model	50
Table 4.2	Training Dataset KNN Algorithm Prediction Result and Accuracy	52
Table 4.3	Training Dataset Keras Model Prediction Result and Accuracy	52

LIST OF FIGURES

		PAGE
Figure 1.1	Paddy Crop	2
Figure 2.1	Paddy Rice infected with Bacterial Blight Disease	7
Figure 2.2	Paddy rice infected with leaf blast disease.	8
Figure 2.3	Paddy Leaf infected with brown spot disease.	10
Figure 2.4	Flow diagram of the automated plant leaf disease identification system.	12
Figure 2.5	(a) Image Input before Contrast Enhancement (b) Image after Contrast Enhancement	13
Figure 2.6	Image input and Alternaria Alternata as classifier output	15
Figure 2.7	Image input and Anthracnose as classifier output	15
Figure 2.8	Flow chart of the plant disease identification system.	16
Figure 2.9	(a) Grape Leaf Black Rot Disease (b) Grape Leaf Esca Disease (c) Grape Leaf Leaf Blight Disease	16
Figure 2.10	(a) Original Grape Leaf sample Image (b) Grape Leaf Sample Image after perform Grabcut segmentation	17
Figure 2.11	(a) Image during Acquisition (b) Thresholded Image. (c) Extension. (d) Corrasion.	18
Figure 2.12	(a) HSV Image. (b) h-channel Image. (c) Thresholded Image. (d) Binary XOR operation for Corrasion Contour Image and thresholded HSV image. (e) Segmented Disease Area	18

Figure 2.13	Classification of infected leaf part by Semisupervised learning. (a) Image Acquisition (b) BGR Image Conversion (c) HSV Image Conversion (d) Blue pixels mask for filtration.	19
Figure 2.14	Illustration of sequential neural network	21
Figure 2.15	Flow diagram of the character recognition system by Tensor Flow System	22
Figure 2.16	Tensor Flow Execution Engine	23
Figure 2.17	Result of the system after analyze the image.	24
Figure 2.18	Flowchart of system	25
Figure 2.19	3 level of cocoa ripeness	25
Figure 2.20	Testing result for Raw Cocoa	27
Figure 2.21	Testing result for Medium Cocoa	28
Figure 2.22	Testing result for Mature Cocoa	28
Figure 2.23	Three types of Potato Leaf Disease	29
Figure 2.24	Flowchart of the system	29
Figure 2.25	KNN classification between 2 classes	30
Figure 3.1	General Flow Diagram	35
Figure 3.2	Flow chart of paddy leaf disease identification system with KNN method	38
Figure 3.3	Python function library that required by system	39
Figure 3.4	Code for Read Image, Resize and Feature Extraction	40
Figure 3.5	Command code for generate KNN model	40

Figure 3.6	Command code to perform gridsearch	41
Figure 3.7	Optimized n_neighbor Value is given	42
Figure 3.8	Command code for KNN model accuracy assessment	42
Figure 3.9	Command code for implementation stage	43
Figure 3.10	Flow chart of Paddy Leaf Disease identification system with Tensorflow	43
Figure 3.11	Command code for system initial stage	45
Figure 3.12	Command code for image dataset creation	45
Figure 3.13	Command code for dataset cache and prefetch	46
Figure 3.14	Command code for data augmentation	46
Figure 3.15	Command code for Keras model generation	47
Figure 3.16	Command code for Keras model accuracy assessment	47
Figure 3.17	Command code for implementation stage	48
Figure 4.1	Paddy leaf samples with disease type identified by manual, predicted by KNN algorithm system and Keras model system	51
Figure 4.2	Flow diagram of the python script	55
Figure 4.3	Command code of the python script	55
Figure 4.4	Prompt window when system start run	56
Figure 4.5	Load Bacterial blight disease sample image into system	57
Figure 4.6	Load Blast of leaf disease sample image into system	58
Figure 4.7	Load Brown spot disease sample image into system	59

LIST OF ABBREVIATIONS

API	Application Programming Interface
BGR	Blue Green Red
GLCM	Gray Level Co-occurrence Matrix
GMM	Gaussian Mixture Model
HSL	Hue Saturation Lightness
HSV	Hue Saturation Value
IDE	Integrated Development Environment
KNN	K Nearest Neighbors
MNIST	Modified National Institute of Standards and Technology
PLRV	Potato Leaf Roll Virus
RGB	Red Green Blue
SVM	Support Vector Machine

©This item is protected by original copyright

Penciptaan Dan Penilaian Prestasi Sistem Pengenalpastian Penyakit Daun Padi Dengan Menggunakan Algoritma KNN Dan Model Keras

ABSTRAK

Projek ini bertujuan untuk membina sistem pengenalpastian penyakit daun padi dengan menggunakan bahasa pengaturcaraan Python. Padi sawah ialah tanaman makanan yang amat penting di dunia. Walaupun padi dapat dituai dua kali setahun, produktiviti dapat dipengaruhi sekiranya dijangkit penyakit. Penyakit daun padi ialah salah satu penyakit yang boleh menyebabkan kerugian besar pada tanaman padi dan mengurangkan pendapatan petani. Cara tradisional untuk mengenalpasti penyakit daun padi dengan mata kasar sangat membazir masa dan tenaga kerja. Oleh itu, sistem pengenalpastian penyakit daun padi amat diperlukan untuk menggantikan cara tradisional. Kaedah klasifikasi penyakit daun padi ini menekankan pada penggunaan teknik memproses gambar dan pembelajaran mesin. Projek ini memberi tumpuan kepada pengenalpastian penyakit karah daun padi, penyakit hawar seludang dan penyakit bintik daun. Berdasarkan tinjauan sistem yang sedia ada, terdapat banyak konsep dan teknik yang dapat diterapkan ke dalam sistem baru. Dalam penyelidikan ini, sistem pengenalpastian penyakit daun padi dibina oleh dua modul untuk perbandingan. Salah satunya menggunakan modul *Scikit-Learn* dengan model algoritma *K-Nearest Neighbor (KNN)*; manakala yang lain adalah dengan menggunakan modul *Tensor Flow* dengan model Keras. Algoritma KNN menyimpan segala kes latihan dan menjalankan pengenalpastian kes baru dengan merujukkan persamaan dari kes sedia ada. Keras ialah *Application Programming Interface (API)* peringkat tinggi yang digunakan dalam *Tensor Flow* dengan fokus dalam pembelajaran mendalam. Keras menyusun, membina dan memperbaiki neural model semasa menjalankan latihan. Aliran sistem dimulakan dengan pemerolehan gambar set data latihan, pengekstrakan ciri gambar dan pelatihan model neural. Setelah menjalani latihan model, kedua-dua sistem diuji dengan contoh gambar untuk menjalankan pengenalpastian penyakit daun padi. *Google Colab Notebook* digunakan untuk penulisan program dan simulasi. Skrip Python dapat dilancarkan dengan peranti pintar dan menjalankan pengenalpastian penyakit daun padi di sawah padi. Sumber gambar baru yang mengandungi daun padi boleh dipilih dari kamera atau album foto di peranti pintar. Skrip Python dapat menjalankan pengenalpastian penyakit pada peranti pintar dengan hasil output yang betul. Prestasi sistem pengenalpastian penyakit daun padi oleh algoritma KNN dan model Keras dinilai dan dibandingkan dari aspek masa pengiraan komputasi dan ujian sampel untuk kedua-dua model tersebut. Tiga jenis penyakit daun padi berjaya dikenali. Kadar ketepatan dapat ditingkatkan lagi dengan pertambahan lebih banyak input gambar berlabel. Dari segi masa pengiraan komputasi, algoritma KNN dapat menyelesaikan semua tugas dalam 4mins 13sec, sementara model Keras memerlukan jumlah masa 18mins 30sec. Dari hasil ujian sampel, ketepatan sistem pengenalpastian penyakit daun padi oleh algoritma KNN dapat mencapai 81.33%. Ketepatan sistem pengenalpastian penyakit daun padi oleh model Keras adalah 74.67%. Algoritma KNN menawarkan prestasi yang lebih baik dalam sistem pengenalpastian penyakit daun padi.

Design And Performance Evaluation Of Paddy Leaf Disease Identification Using KNN Algorithm And Keras Model

ABSTRACT

This research aims to establish paddy leaf disease identification system by using Python programming language. Paddy rice is one of the important food crops in the world. Although paddy rice can be harvested twice per annum, the productivity could be affected if the paddy plantation is infested by diseases. Paddy leaf disease is one of the diseases that can cause huge loss to paddy crops and bring down the farmers' income. Traditional way of identifying paddy plant diseases by human eyes is very time consuming and labour intensive. Hence it is necessary to set up paddy rice disease identification system to replace this traditional way. This disease classification system emphasizes on the usage of image processing technique and machine learning. This research work focuses on identifying three main types of paddy leaf diseases; Blast of Leaf, Bacterial Blight and Brown Spot. To date, there are a lot of concepts and techniques were applied to the related system. In this work, the paddy leaf disease identification system uses two different modules as for performance comparison. One is using *Scikit-Learn* module with *K-Nearest Neighbour* (KNN) algorithm; another is by using *Tensor Flow* with *Keras model*. KNN algorithm implements supervised algorithm to save all the training cases and assigns classifier to new cases based on measurement of similarity. Keras is a high level *Application Programming Interface* (API) running in *Tensor Flow* that focus in deep learning. Keras compiles, constructs and improves neural network model during training epochs. Paddy leaf disease identification system comprises of training images acquisition unit, image feature extraction unit, and neural model training unit. Upon completion of training on databases of related images, both systems are tested with test samples image to perform classification of the aforementioned three types of paddy leaf diseases. The *Google Colab Notebook* is used for the two algorithms realization and simulation. Python script can be ran on smart devices to load the KNN data file or Keras model files and perform on-site paddy leaf disease identification. When running on these smart devices, the input image can be chosen from either camera or photo album. The Python script is able to execute the disease identification on the smart device with correct output results. The performance of paddy leaf disease identification system by KNN algorithm and Keras model is evaluated and compared from the aspects of computational time and samples test for both models. Simulation results showed that the three major paddy leaf diseases were successfully recognized by using both models. In term of accuracy of disease detection, it can be further enhanced if more inputs of labelled images are feed as training data. For the computational time, the proposed system with the KNN algorithm able to complete all tasks within 4mins 13sec, while the Keras model required total processing time of 44mins 35sec. In terms of computational accuracy, the KNN algorithm achieved up to 81.33%, whereas the accuracy of the paddy leaf disease identification system by the Keras model is 74.67%. Thus it is clearly showed that the KNN algorithm performed better for the case of paddy leaf diseases identification system.

CHAPTER 1 : INTRODUCTION

1.1 Background

This chapter summarizes the conducted research study. The research aims to develop a python-based image processing-based system to detect the presence of paddy disease based on its leaf images. Chapter 1 consists of background, problem statement, research objectives, research scope, and thesis organization. The background glanced through the identification of projects and related issues. Problem statement further elaborated the issues that encountered and emphasize the necessary to establish the system. The objectives are the main goals and targets that set throughout the projects. Scope of study will discuss about the purpose of the project and expectations from this project. Lastly, thesis organization gives a summary of the sequence for each chapter in this thesis.

1.2 Problem Statements

Technology nowadays is getting more advanced. Handy equipments such as smartphone and tablets have significant increase of usage in life. With these, the techs in phones have been improvised over a time. Thus, the usage of smart phone has been discovered for almost all fields and applications. Some research and studies implies that on-board sensors and computational power in smartphone can be used in complex algorithm process, such as image processing for the agricultural products. One of these areas is the early detection of agricultural products' diseases. With early detection and

solving of these plants' diseases, the impact can be minimized and subsequently improve product's quality assurance. (Gesellschaft, 2011)

Among all agricultural plants from worldwide, paddy rice plays so important role that over half of our world's population deeply relies on it. Paddy is considered as one of the main food supplies. (Nayar, 2014) It is widely grown in Asia, especially Malaysia and Thailand.



Figure 1.1 Paddy Crop
Source: (Jose, 2009)

Paddy is typically harvested twice per annum. However, some problems faced by paddy farmers because the paddy fields exposed to be attacked by diseases, pests, bacteria, and fungi. One of these main loss factors is paddy leaf diseases. If the paddy crop had been infested by the diseases, the rest might also get exposed to the risk of impact. As outcome, the farmer's income will be decreased.

Disease is known as the abnormal condition that causes the plant malfunction in growth and fruiting. It is crucial to get the symptoms detection, treatment, and

prevention of plant diseases, especially in the early stages, in order to guarantee the paddy crops yield. (D. Y. Kim et al., 2018)

Manual handling of the examination on plant leaf specimens is tedious, as it consumes a lot of time and workload. Automated plant disease identification shows the importance of image processing application, since it works with simply based on the symptoms on plant leaves. (A. K. Singh et al., 2018)

Paddy leaf disease identification system through image processing will be very helpful once established. In this research, a system is built from python code that functions to detect and classify the paddy leaf disease by using image processing technique and machine learning.

1.3 Objective

- i) To design paddy leaves identification system by using Python programming language.
- ii) To perform pattern recognition with implementation of machine learning technique through K-Nearest-Neighbours (KNN) Method and Keras model.
- iii) To analyse performance of the developed system in terms of it detection accuracy and computational time.

1.4 Scope

The scope of this research is to establish paddy disease identification system using Python programming language. Current traditional way of identifying paddy plant diseases by human eyes takes a lot of time and labor intensive. The research work begins with the study on current existing plant disease detection system and procedures to train and classify the disease type from acquired input images.

Next, study and implementation of techniques on image acquisition and image processing. Feed the paddy leaf images with diseases and classifiers into the training system to generate an appended image database file and model file. In this project, KNN algorithm method and Keras model with Tensor Flow is used to train the system. Then, predict test sample images by identification system based on training model. The performance of both systems is reviewed and compared.

Finally, the prototype system is loaded in smart devices to check the function. This is to perform paddy leaf disease identification on the paddy field site.

1.5 Thesis Organization

There are 5 chapters in this thesis. Each chapter is organized with several subheadings.

Chapter 1 introduces briefly on the background, problem statement, research objective and research scopes.

Chapter 2 will focus on introduction of the paddy plant, the literature review of current existing methods and technique.

Chapter 3 will explain on the methodology and how the experiments are carried out.

Chapter 4 will discuss about the results obtained from the experiments and research. All the results are arranged accordingly to research process flow for better understanding.

Finally, Chapter 5 will conclude the whole project.

©This item is protected by original copyright

CHAPTER 2 : LITERATURE REVIEW

2.1 Type of Paddy Leaf Diseases

Paddy is the primary food supply for over three billion people of the world. Paddy rice is the seed of the grass species *Oryza Sativa*. As one of the most utilized food plant, this type of cereal grass is so important among the worldwide. (Zeigler, 2017)

Many factors has been identified that can affect the yield of paddy rice crops. One of the critical factors is the paddy leaf disease. The following subsections summarize the type of paddy leaf diseases, symptom of the diseases and the solution for each of them. This research will be focused on the three most dangerous bacterial blight disease, blast of leaf disease and brown spot disease.

2.1.1 Bacterial Blight Disease

Bacterial blight is a severe paddy rice disease. When infested at small seedlings level, the infested leaves turn from yellow color to straw-colored and sooner it will wilt and decease. When infested at mature level, it typically occurs during tillering to panicle formation. Tillering is the growth of additional side shoots and produce multiple stems which starting from the initial single seedling. Panicle is the end terminal component of rice tiller. The symptom of bacteria blight disease is as shown in Figure 2.1. (PEAT GmbH, 2015)



Figure 2.1 Paddy Rice infected with Bacterial Blight Disease
Source: (PEAT GmbH, 2015)

Initially it exhibits as soft green or grayish green streaks first on leaves. As it gets worse, these streaks will merge and become long yellowish streak line. Leaves tend to yellow and eventually die. Finally, the leaf will drop off milky bacterial ooze which later dried and form white crust.

Bacterial blight disease is caused by bacteria named *Xanthomonas oryzae pv. Oryzae*. This disease spread easily under condition of bad weather, high humidity and warm temperatures. Higher level of nitrogen in fertilizer or closer plantation tends to help the spread of Bacterial Blight disease. Majority of grain will break if infest during paddy panicle growth period. As preventive before planting, it is recommended to treat the paddy seeds with antibiotic and copper oxychloride. Below listed additional measures are recommended to reduce the bacterial blight disease once discover and confirm:-

- i) Check drainage and Nursery system.
- ii) Check nitrogen level in fertilization if excessive.
- iii) Apply an extra dose of potash along with the last dose of nitrogen.
- iv) Clear the infested weeds from channels and surroundings to minimize the spread risk.
- v) Dry the paddy fields between seasons in order to extinguish the bacteria in surrounding.

2.1.2 Blast of Leaf

A Paddy Leaf blast symptom is white spots in round or oval shapes, with dark color margins on the leaves, as shown in Figure 2.2. It will infest on every parts of the paddy plant.



Figure 2.2 Paddy rice infected with leaf blast disease.
Source: (PEAT GmbH, 2015)

Once infested, leaves will exhibit yellow to light green chlorotic, eye-shaped lesions with pointed ends. The lesions' size relies on plant age and infested time. As the symptom enlarge by time, the leaves eventually dry out and die. Paddy leaf blast is caused by a type of fungus named *Magnaporthe Grisea*. The fungus is able to survive on the straw after harvest and bring forward to next season. There are several factors that encourage leaf blast disease, such as low temperature, rain season, high nitrogen level in soil or fertilization, low silicon levels and low moisture in soil.

On chemical preventive control for blast leaf, the seed can be treated with thiram. Farmers can use fungicides that contain *Azoxystrobin* or *Triazoles* or *Strobilurins* to spray at nursery, tillering, heading and panicle to control paddy leaf blast. As recommendations, below measures to control the blast leaf disease:-

- i) Check nitrogen level in fertilization if excessive.
- ii) Monitor flood level of the field if sufficient water supply.
- iii) Prevent drainage of paddy field.
- iv) Maintain control of weeds and alternative hosts.
- v) If soil is silicon deficient, use silicon fertilization.
- vi) Clear the infested weeds from channels and surroundings to minimize the spread risk.

2.1.3 Brown Spot of Rice

This Brown spot disease symptom is as shown in figure 2.3. Brown circle necrotic spots are obvious on the young leaves. On mature plants, the reddish margins can be observed.



Figure 2.3 Paddy Leaf infected with brown spot disease.
Source: (PEAT GmbH, 2015)

This disease is caused by *Cochliobolus Miyabeanus*, which is a type of fungus. It is capable to survive in seed level up to four years and spread to other plants through airborne spores. Typically brown spot disease is due to mishandling of soil fertility, in from the aspects of micronutrients. The brown spot disease also can be caused by other factor, such as high humidity (86%-100%), long period of moisture on leaves and high temperature (16-36°C).

Below recommendation on preventives measures:-

- i) Apply Fungicides such as *Iprodione*, *Propiconazole*, *Azoxystrobin*, *Trifloxystrobin* for seeds treatment.
- ii) Apply mixture of cattle manure with chemical fertilization.
- iii) Use Silicon Fertilization.
- iv) Ensure balanced nutrient supply and soil nutrient level regularly.
- v) Clear weeds in the surrounding.
- vi) Clear the infested weeds from channels and surroundings to minimize the spread risk.

2.2 Case Study on Existing Image Processing and Recognition Systems

2.2.1 Classification of Plant Diseases Through Leaves by Machine Learning Technique

A study was conducted by Deepa, Rashimi N and Chinmai Shetty aimed to develop an automatic plant disease detection system through machine learning techniques. Figure 2.4 shows the flow diagram of the system. The training dataset consist of total 58 images, as shown in Table 2.1. After feed in the training dataset, the system will be able to perform plant disease identification. (Deepa et al., 2021)

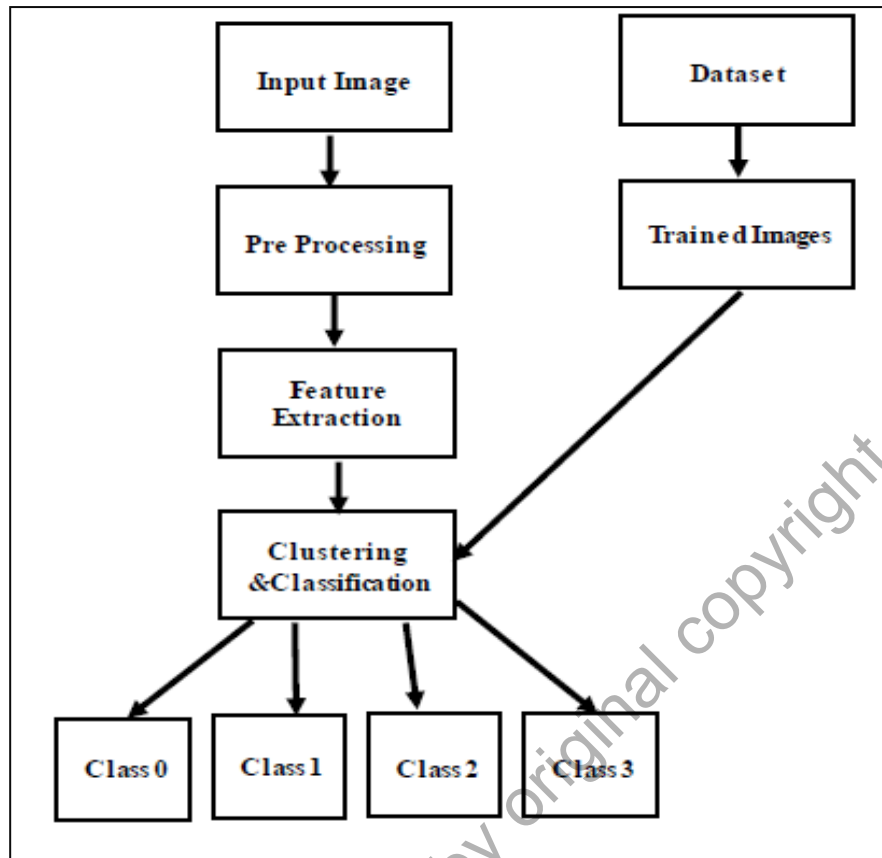


Figure 2.4 Flow diagram of the automated plant leaf disease identification system.
Source: (Deepa et al., 2021)

Initially, colored photo of plant leaf taken by camera and input to the system. After the image input, image processing technique is performed to clear all noise signal in image data and enhance the image features, such as contrast. The contrast enhancement is done by protraction the input intensity to new value. Figure 2.5 shows the image pre and post enhancement.