PLANT DISEASE CLASSIFICATION SYSTEM USING DEEP LEARNING

MOHD AZIZUL BIN ASMAD LATIP

FACULTY OF COMPUTING AND INFORMATICS UNIVERSITI MALAYSIA SABAH 2022



PLANT DISEASE CLASSIFICATION SYSTEM USING DEEP LEARNING

MOHD AZIZUL BIN ASMAD LATIP

THESIS SUBMITTED IN PARTIAL FULFILMENT FOR THE BACHELOR'S DEGREE IN COMPUTER SCIENCE WITH HONOURS

(NETWORK ENGINEERING)

FACULTY OF COMPUTING AND INFORMATICS UNIVERSITI MALAYSIA SABAH 2022



DECLARATION

I acknowledge that this Bachelor's Degree Thesis is the result of my own efforts and work, except for quotations, reference and summaries, each of which I have explained the source.

14 January 2022

Azizul

MOHD AZIZUL BIN ASMAD LATIP BI18110214



CONFIRMATION OF REPORT

NAME	: MOHD AZIZUL BIN ASMAD LATIP
MATRIC NUMBER	: BI18110214
TITLE	: PLANT DISEASE CLASSIFICATION SYSTEM USING DEEP LEARNING
DEGREE	: BACHELOR OF COMPUTER SCIENCE WITH HONOURS (NETWORK ENGINEERING)
VIVA'S DATE	: 28/12/2022

CERTIFIED BY;

1. SUPERVISOR

Signature

DR. MUZAFFAR BIN HAMZAH



ACKNOWLEDGEMENT

First and foremost, I am grateful to ALLAH S.W.T for allowing me to complete my Final Year Project 2 and for providing me with the strength I require to perform my responsibilities as a Computer Science student.

I'd like to thank my supervisor, Dr. Muzafar bin Hamzah, for his support, direction, recommendations, criticism, and friendship during the completion of this research. Thank you also to all of the helpful academics in the Faculty of Computing and Informatics for their time, advice, and suggestions as the project progressed at university.

Above all, I'd want to express my gratitude to my family for their outward support when I informed them I wanted to continue my studies, as well as my friends for supporting me to complete this project. The concepts would not have been accomplished without their help.



ABSTRACT

In agriculture production, the unlimited and no disease plant product from farming become important things due this product is necessities in terms foods and mainly source for highly nutrient to community regardless country in the world. However, an increase in the human population requires an increase in agricultural production. Generally, the most important thing in agriculture that affects the quantity and quality of crops is plant diseases and make crop disease become a major threat to food security. In general, a farmer knows that his plant is attacked by a disease through direct vision. But, this process is sometimes inaccurate. Plant diseases are not only a threat to food security at the global scale, but can also have disastrous consequences for smallholder farmers whose livelihoods depend on healthy crops. With the development of machine learning technology, plant disease classification can be done automatically using deep learning. In deep learning, a convolutional neural network (CNN) is a class of deep neural networks, most commonly applied to analysing visual imaginary. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on the shared-weight architecture of the convolution kernels that scan the hidden layers and translation invariance characteristics.



ABSTRAK

Dalam pengeluaran pertanian, hasil tanaman tanpa had dan tiada penyakit daripada pertanian menjadi perkara penting kerana produk ini adalah keperluan dari segi makanan dan terutamanya sumber khasiat tinggi kepada masyarakat tanpa mengira negara di dunia. Namun begitu, pertambahan populasi manusia memerlukan pertambahan pengeluaran pertanian. Secara amnya, perkara terpenting dalam pertanian yang mempengaruhi kuantiti dan kualiti tanaman adalah penyakit tumbuhan dan menjadikan penyakit tanaman menjadi ancaman utama kepada keselamatan makanan. Secara umumnya, seorang petani mengetahui bahawa tumbuhannya diserang penyakit melalui penglihatan secara langsung. Tetapi, proses ini kadangkala tidak tepat. Penyakit tumbuhan bukan sahaja mengancam keselamatan makanan pada skala global, tetapi juga boleh membawa kesan buruk kepada petani pekebun kecil yang mata pencariannya bergantung kepada tanaman yang sihat. Dengan perkembangan teknologi pembelajaran mesin, klasifikasi penyakit tumbuhan boleh dilakukan secara automatik menggunakan pembelajaran mendalam. Dalam pembelajaran mendalam, rangkaian saraf konvolusi (CNN) ialah kelas rangkaian saraf dalam, yang paling biasa digunakan untuk menganalisis khayalan visual. Ia juga dikenali sebagai invarian anjakan atau rangkaian neural tiruan invarian ruang (SIANN), berdasarkan seni bina berat kongsi inti konvolusi yang mengimbas lapisan tersembunyi dan ciri invarian terjemahan.



TABLE OF CONTENT

	Page
TITLE	Ι
DECLARATION	П
COMFIRMATION OF REPORT	III
ACKNOWLEDGEMENT	IV
ABSTRACT	V
TABLE OF CONTENT	VII-VIII
LIST OF TABLE	IX
LIST OF FIGURE	Х
LIST OF ABBREVIATIONS	XI
CHAPTER 1: INTRODUCTION	
1.1 Problem Background/Motivation	1
1.2 Problem Statements	2
1.3 Research Objective	3
1.4 Project Scope	3

CHAPTER 2: SYSTEMATIC LITERATURE REVIEW

4-5
6-7
8-10
11-27
28-32
33





CHAPTER 3: METHODOLOGY

34
35
35
35
35
36
36
37
37

CHAPTER 4: SYSTEM ANALYSIS AND DESIGN

4.1 Image Acquisition	
4.2 Image Pre-processing	39
4.3 Feature Extraction and Classification	39
4.3.1 Feature Extraction	40
4.3.2 Classification	40
4.4 Dropout	40
4.5 Activation Function	40
4.6 CNN Architecture	41
4.6.1 Proposed CNN model	41
4.7 Proposed Data Flow Diagram	42
4.7.1 DFD Level 0	42
4.7.2 DFD Level 1	42
4.2.3 DFD Level 2	43
4.8 Entity Relationship Diagram	44
4.9 Graphical User Interface	45-46



CHAPTER 5: IMPLEMENTATION

5.1 Image Acquisition	47
5.2 Image Pre-processing	48
5.3 Proposed CNN Model	49
5.3.1 Model Summary	50
5.3.2 Compiling and Fitting	50
5.4 Pre-trained AlexNet and ResNet50	
5.4.1 Pre-trained AlexNet	52
5.4.2 Pre-trained ResNet50	54
5.5 System Implementation	

CHAPTER 6: TESTING AND RESULT

6.0 Proposed CNN result	57
6.1 Testing System	58-59
6.2 AlexNet and ResNet50	60
6.0.1 AlexNet	60
6.0.2 ResNet50	61
6.0.3 Conclusion	62

CHAPTER 7: CONCLUSION

7.0 Advantages and Disadvantages	
7.0.1 Advantages	63
7.0.2 Disadvantages	63
7.1 Achieved Objectives	
7.2 Future Works	

REFERENCE



LIST OF TABLE

		Page
Table 2.4.1:	RQ1 Extracted Result	12
Table 2.4.2:	RQ2 Extracted Result	18
Table 2.4.3:	RQ3 Extracted Result	25
Table 3.3.1:	Software Requirement	35
Table 3.3.2:	Hardware Requirement	35
Table 6.2:	Testing Result	60



LIST OF FIGURE

Figure 2.2:	Baseline CNN architecture	6
Figure 2.3(a):	PRISMA SLR	9
Figure 2.3(b):	Review Process	10
Figure 2.4.1:	Problem Background	13
Figure 2.4.2(a):	Methodology	19
Figure 2.4.2(b):	Model	19
Figure 3.1:	Iterative Process	34
Figure 4.1:	Sample test dataset	38
Figure 4.2:	Sample train dataset	39
Figure 4.7.1:	Level 0	43
Figure 4.7.2:	Level 1	44
Figure 4.7.3:	Level 2	44
Figure 4.8:	Proposed ERD	45
Figure 4.9.1:	Proposed GUI	46
Figure 4.9.2:	Uploaded image	46
Figure 4.9.3:	Treatment	47
Figure 5.1:	Organize datasets	48
Figure 5.2.1:	Pre-processing datasets	49







49

Page

Figure 5.3:	Implementation model	50
Figure 5.4:	Model Summary	51
Figure 5.5:	Compiling and fitting model	51
Figure 5.6.1:	Load pre-trained AlexNet	52
Figure 5.6.1.2:	Summary pre-trained AlexNet	53
Figure 5.6.1.3:	Updating Output layer	53
Figure 5.6.1.4:	Compiling pretrained AlexNet	54
Figure 5.6.2:	Load pre-trained ResNet50	54
Figure 5.6.3:	Fine-tuning Dense layer	55
Figure 5.6.4:	Compile and fit ResNet50	55
Figure 5.7.1:	Import library	56
Figure 5.7.2:	Main	56
Figure 5.7.3:	Classification	57
Figure 6.1.1:	Epochs-Accuracy	58
Figure 6.1.2:	Epochs-Loss	59
Figure 6.3.1.0:	Epochs-Accuracy and Epochs-Loss	60
Figure 6.3.1.1:	Epochs-Accuracy and Epochs-Loss	62
Figure 6.3.2.0:	Epochs-Accuracy and Epochs-Loss	63
Figure 6.3.2.1:	Epochs-Accuracy and Epochs-Loss	64



LIST OF ABBREVIATIONS

CNN	- Convolutional Neural Network
ML	- Machine Learning
GUI	- Graphical User Interface
ERD	- Entity Relationship Diagram
DFD	- Data Flow Diagram
RQ	- Research Question
RO	- Research Objectives





CHAPTER 1

INTRODUCTION

This chapter discusses an overview of the study conducted. The project title is Plant Disease Classification System Using Deep Learning. It consists of problem statements and motivation, research question, research objectives, and the scope of the study. Problem statements describe the problems that arise and make the selected projects to be undertaken. To tackle this problem statement, research question has been devised. Research objectives are the goals list for the projects to be achieved. At last, the scope of the study discusses about the how the user is related with our propose solution and the way the user using our propose solution.

1.1 PROBLEM BACKGROUND / MOTIVATION

In crops, the leaves become one of main area indicating to get information about the quantity and quality of the agriculture product. The existence of disease's plant can be affect by few factor, such as climate change, presence of weed, and soil infertility. Apart from that, plant disease affecting regardless position either individually, community, and also global. By individually such as the farmers, will make their product will decrease due the good product is separated with bad product. In by community, will make the price of agriculture product such as vegetables and fruit in market will increasing due the decreasing quantity of agriculture product? By global, will threat to the growth of several agricultural products and a source of economic losses. The failure to detect disease in plants will lead to the using pesticide or fungicide with higher consumption that will costly. Therefore, the visual



inspections by expert and biological review must be carried out through plant diagnosis when required. However, this method is usually time-consuming and cost ineffective. To deal this issue, the intelligent technology such as Machine Learning will necessary to help to detect plant disease. The motivation of this project is using Deep Learning by specifically Convolutional Neural Network algorithms for doing classification task. This is because, this algorithms strikingly effective for real-life object classification, recognition, and classification purposes. Before this algorithms, the conventional machine learning (ML) been used, but this technology not improvise as deep learning to classification purposes. To address the task of object classification, is performed in a single platform by using deep learning architectures. The Convolution Neural Network (CNN) was implement as a techniques to detect disease of plant.

1.2 PROBLEM STATEMENT

In this study, how different pre-trained model that used by each researcher to implement their CNN model influenced their result in terms accuracy of their model using the same datasets. This is because, their approach resulting different output due factor such as using different resources or parameter of model. This can be refer to (Sembiring et al., 2021), (Mohanty et al., 2016), and (Mahmudul Hassan et al., 2021) that their accuracy of their pre-trained model is influenced by their parameter of model.

Machine learning previous approach such as KNN and SVM cannot ensure high accuracy due the hand-crafted feature and tedious classification task. This can be seen when (Gharghory, 2020) and Amit Wagle et al mention that typical machine learning before CNN is usually resulting low accuracy due this approach is hand-crafted and tedious task. Thus, for his propose solution is both using CNN AlexNet model as a classifier. This is because, CNN algorithms strikingly effective for real-life object classification. Thus, implemented of CNN model is more favourable than previous approach such as KNN and SVM.

Moreover, low accessibility by farmers to detect disease of plant due the lacking knowledge about disease of plant and also lacking available automatic system to detect. This problem can be seen when (Swathika et al., 2021), (Rubini & Kavitha, 2021), (Jadon, 2020), (Moyazzoma et al., 2021), (Deepalakshmi et al., 2021), and (Li et al., 2020) face the same problem. In their research, farmers unable to identify crops diseases, this is due the farmer's lack of knowledge about the crops specific detail whether in condition or time. Thus, the

2



UNIVERSITI MALAYSIA SABAH

accessibility farmers to detect disease through their vision is low. So, the system that enable to detect disease by classification is needed.

1.3 RESEARCH OBJECTIVES

The purpose of this final year project is to conduct project-based research. Simultaneously, this initiative ensures that it is active in product development. This project's goals are as follows:

I. RO1

a. To examine and assess 2 pre-trained model which is AlexNet and ResNet of the plant disease classification system via literature review.

II. RO2

b. To design and implement appropriate CNN model of plant disease classification system.

III. RO3

c. To test the implementation plant disease classification system as a propose solution.

1.4 PROJECT SCOPE

In this project, the datasets of images will be divided to 2 part which is train datasets and test datasets. Train datasets will defined and test data to train the model. The model that implement is Sequential model that will defined and trained by using train datasets. In the Sequential model, which is hidden layer. The few convolutional layer with parameter such as 'relu' in each layer and 'softmax' layer in final layer was added to increase the accuracy and speed. The image of test data will used to test data and check the accuracy. The test data can be test with insert the directory of where image of test data is placed. The new image test datasets can also downloaded from internet or take picture and place the image in favorable directory. Finally, a web application will be used to see the result. The farmers can use this web application by take the picture of their leaves and stored the image in folder.





CHAPTER 2

SYSTEMATIC LITERATURE REVIEW

2.1 Introduction

Plant diseases are the primary threat to agricultural progress across the world, causing significant losses each year. The treatment of plant diseases has gotten a lot of press. The goal of this study was to diagnose plant illness by looking at the leaves. In general, trees may be turned into beautiful furniture, and their leaves have decorative value. However, because existing plant diseases have resulted in significant economic losses, thus this study focused on the necessity to identify plant leaf disease quickly and accurately.

In recent years, a lot of research have successfully used deep learning models to achieve varying degrees of accuracy on laboratory/field pictures. For example, Sembiring et al, Swathika et al, Rubini & Kavitha, Moyazzoma et al, and Deepalakshmi et al. When evaluated on data that is identical to that used during training, these accurate classifiers achieve high accuracy, but when tested on different data, they fail horribly. For example, Mohanty et al, whereas training models achieved accuracy and confidence of more than 99 percent, confidence plummeted to less than 40 percent when evaluated against pictures from reliable internet sites. Given the relatively small sizes of datasets currently available (tens of thousands of images versus millions of images required for training models from scratch), transfer learning, or tweaking the final output of a trained neural network to a new dataset,



has become a popular tool of choice for researchers. Examples of using this technique is Sembiring et al, Rubini & Kavitha, Chen, Chen, et al, and etc.

For objective of this study is to examining or investigate the state of the art for deep learning approach in plant disease classification. In this study, technique that use by researcher to implement their model is examine on how their type approach influenced their result in terms accuracy and performance of their model. Also, study on how researcher utilized hyper parameter and pooling combination to get a better result. Thus, by gaining knowledge from studying research paper, we will apply this technique to our model as we can get a good result.

In this study, we have conducted our research question to 3 question as our method review which is for RQ1 is **What is the problem background that urge implementation of plant disease classification system?** For RQ2 is **What method and process is used to implement plant disease classification system?** For RQ3 is **What is the implementation result of the plant disease classification system?** This RQ will help us to achieve our objectives that is to investigate the state of art for deep learning approach in plant disease classification.

The contribution from this study is related to our 3 research question. For our RQ1, we have divide the main problem to 10 part as reference to vary problem that urge implementation. For RQ2, the method and process is use have investigate and classified them and also compare their result in table form. Next for RQ3, the finding each research paper have identified and discussing them in this literature review

The rest of this paper is organized as follows: Section 2 presents an overview of the process of baseline Convolutional Neural Network and their related work. Section 3 introduces the methodology for our research question and our search strategy. Section 4 shows the result. Section 5 holds the discussion and, finally, Section 6 reports the final conclusions.

2.2 Background

A convolutional neural network (CNN) is a form of artificial neural network that is specifically designed to process pixel data and is used in image recognition and processing. CNNs are image processing, artificial intelligence (AI) systems that use deep learning to perform both generative and descriptive tasks, with image and video recognition as a common application (What Is Convolutional Neural Network? Definition from WhatIs.Com et al).

UNIVERSITI MALAYSIA SABAH

5

The CNN configurations are generally made up of a number of particular components that represent the differences in different architectures. The overall architecture of a CNN is depicted visually on figure 2.2(a), with its key parts such as the input layer, convolutional layer, pooling layer, and a flattening process, in which data is inputted into a collection of dense layers, which reflect the result achieved in the output layer. With the preceding layer, convolutional layers carry the outputs of convolution filters or kernels. The weights and biases, which may be learnt in each iteration using an optimization function, are the key parameters of these filters or kernels. The optimization function's goal is to generate kernels that are good data representations with no errors. Down sampling is done with **pooling layers** to minimise the size of the neuron and alleviate the performance issue of over-fitting. The most common form of pooling operation in pooling layers is max pooling, which captures the maximum value of the pooling window. Non-linearity is added to the network via activation function layers. There are several activation functions in the literature, including sigmoid, tanh, and ReLU, which is the most commonly utilised. Dropout layers are used to solve the problem of over-fitting by shutting down the neurons in the network at random. To calculate the scores or probability of classes, fully connected layers are used. The inputs to the classifier are the outputs of the fully connected layers; the softmax classifier is the most well-known.



Figure 2.2: Baseline CNN architecture

CNNs have been used in numerous research to classify and detect diseases. When it comes to Amit Wagle et al, and this author proposed network-based on AlexNet and comparing with the traditional Support Vector Machine (SVM). They using Transfer Learning for the pre-trained AlexNet network for a different amount of data for training, and the result are validated with a SVM and Deep learning classifier. As a result, AlexNet performed well with an accuracy of 91.15% as compared to SVM giving 88.96% and 89.69%. From





the accuracy result, we can confirm that using Deep learning is better than traditional approach which is SVM on doing classification task in terms accuracy result.

Luaibi et al developed a plant disease classification model, by doing experiment of CNN model such as AlexNet and ResNet with and without data augmentation. Data augmentation involves the process of creating new data points by manipulating the original data. Thus, the dataset in this experiment will increase and increase the accuracy. As a result, the trained models with data augmentation give the best results with 95.83% and 97.92% for ResNet and AlexNet respectively.

Based from the existing review that we have study, we have gaining knowledge about on how architecture of CNN work essentially and how the researcher manipulate part of architecture to increase the result in terms accuracy and performance. Thus, this research is conducted to examine and assess deep learning approach that apply in agriculture.

2.3 Methodology

In order to be able to analyse the main reason that urge the researcher implement plant disease classification system. For **RQ1**, we should first recognize the problem background as main objective of study. Thus, the first objective of this study is doing research on diverse problem background from diverse research article to analyse the main reason researcher implement Deep Learning based solution to overcome problem background. We achieve that by study 60 research article main problem and extract in table then analyse it.

For **RQ2**, by grouping the literature per publication in range year 2016-2021, we are able to identify possible approach in the research paper as their propose solution. The propose solution is highlight in each research paper as main focus such as specific model and technique that researcher use to doing experimentation while research is progress. The model each research paper is difference between each other and also revolution model appear by mixing with other solution to get a better result. The technique that researcher use will divided to two strategy which is Transfer Learning and from scratch to implement model. After having knowledge about the model specification that researcher used as tool to detect disease of crops. Next, we will discuss about how the process implementation of their approach by study their specific model and experimentation. Thus, analysing the methods by grouping them in table might give an idea on how the difference model and technique will influence their result.



7

For **RQ3**, we will discuss about the finding from article extracted result. For systematic work, we have done grouping our finding in table form to make extracting task more efficient. The table will show each article finding from their experimentation result. From that, will apply the useful and related knowledge with our project progress.

The result from future work will apply to our implementation system as a reference as our main objective is to implement system that sustainability in terms accuracy and user interface. Thus, we will list our future work in each article and apply the knowledge. All the process has been concluded in diagram 2.3(a) below.



Figure 2.3(a): PRISMA systematic review process

The strategy for collecting the relevant literature is twofold, first is a keyword search in a list of scientific libraries and second is the collection of the papers from the deep learning approach in agriculture. The scientific libraries included in the search is only one which is Scopus, this is because, Scopus easier to cite via Mendeley which is reference manager. The literature extraction comprises of two independent keyword searches in the libraries above using the search keywords "Deep Learning" and "Plant Disease classification." The search query has been made as short as possible in order to retrieve as many publications as possible that include the phrases.

The selected literature body collected from both strategy should commit to a set of inclusion criteria. For first criteria, the literature should address Deep Learning as an area of research, either main or secondary. Therefore, the keywords "Deep learning" should exist as a whole and continuously in at least one of the fields: title, keywords or abstract. Additionally, possible composites of the keywords should be examined, e.g., Classification System using Machine Learning. Secondly, be research papers, i.e., articles that have been peer-reviewed.

8





and published in a scientific journal. Next, be written in English. Lastly, have a document body that is more than one page long. As a result, there are no books, extended abstracts, presentations, presentation notes, keynotes, or articles published in a language other than English in the literature.

The review process is done based on figure below which is the step has been taken. Firstly, collecting all the literature. The library search, at this point includes a search of the keywords in the whole text body in order to include the maximum amount of papers. Secondly, applying inclusion/exclusion criteria. The literature collection resulting from the previous step are searched for the keywords in the field's title, abstract, keywords. Next, verifying rejected papers. The rejected literature from the previous step is searched for only the terms "Plant Disease classification" and "Deep Learning" in the field's title, abstract, keywords and evaluated if they are related literature. This would avoid rejecting papers with different combinations of the keywords. And lastly, verifying included papers. The included literature that resulted from the two previous steps is verified manually by reading the abstract and conclusion. In this step, we make sure that the papers included in the review provide results that are directly or indirectly related to the field of Deep Learning.



Figure 2.3(b): Review process



2.4 Result

2.4.1 RQ1: What is the problem background that urge implementation of plant disease classification system?

Problem	Low	Large	Negative	Increasing	Low efficient	Large
Backgroun	Accessibi	scale of	impact of	of world	result by	resources
d	lity.	crops.	decreasing	population.	expert.	of CNN
			crops			model.
			product.			
Source	1. (Swathi	1. (P.	1. (Chen,	1.(Sachdev	1. (Jaiswal et	1.(Sembirin
Info	ka et al.,	Bansal et	Chen, et al.,	a et al.,	al., 2021)	g et al.,
	2021)	al., 2021)	2020)	2021)	2. (Shrivastava	2021)
	2. (Rubini	2. (Genaev	2. (Sachan &		et al., 2019)	2.(Mohanty
	&	et al.,	Rajpal,		3. (Amit	et al.,
	Kavitha,	2021)	2020)		Wagle, n.d.)	2016)
	2021)	3. (Kathires	3. (Mohapatr		4. (Icasmi,	3. (Mahmud
	3. (Jadon,	an, 1911)	a et al.,		n.d.)	ul Hassan
	2020)	4. (Ahmad	2021)		5. (Khattak et	et al.,
	1. (Moyaz	et al., n.d.)	4. (Gomaa &		al., n.d.)	2021)
	zoma et		Abd El-Latif,		6. (Goyal et	4. (S.
	al., 2021)		n.d.)		al., 2021)	Bansal &
	2.(Deepal				7. (Zekiwos &	Kumar,
	akshmi				Bruck, 2021)	2021)
	et al.,				8. (Xiao et al.,	5. (Sharma
	2021)				2021)	et al.,
	3. (Li et				9. (Fountas et	2020)
	al., 2020)				al., 2020)	6. (Mensah
					10. (Saleem,	Kwabena
					Khanchi, et	et al.,
					al., n.d.)	2020)
					11. (Liu et al.,	7. (Fountso
					n.d.)	p et al.,
					12. (Yan et al.,	n.d.)
					n.d.)	
					13.(Maeda-	
					Gutiérrez et	
					al., n.d.)	
					14. (Khatoon	
					et al., 2021)	

Table 2.4.1: RQ1 Extracted Result



UNIVERSITI MALAYSIA SABAH

10