PLANT DISEASE IDENTIFICATION USING AUTOENCODER

JANICE ONG AUN NEE

A project report submitted in partial fulfilment of the
requirements for the award of the degree of
Master of Engineering (Computer and Microelectronic Systems)

School of Electrical Engineering
Faculty of Engineering
Universiti Teknologi Malaysia

FEBRUARY 2021

DEDICATION

This project report is specially dedicated to my father and my mother, Ooi Sock Hong and Ong Beng Hoe, who have been my source of strength and inspiration when I thought of giving up. Not to forget, to my family members, for all their overflowing supports and prayers. I also give thanks to God for all His blessings, for the supportive and kind people surrounds me.

ACKNOWLEDGEMENT

I would like to express my deepest appreciation to all those who provided me the possibility to complete this final year project successfully. A special gratitude I give to my supervisor, Dr Musa Bin Mohd Mokji for encouragement and guidance. His constructive comments, thoughtful ideas and short notes had guided me a lot of positive outcomes to this project.

Furthermore, I would like to thank to my beloved family who had always supported and encouraged me through this tough journey in UTM. Their endless love and motivation have brought me the strength to complete this master program. My sincere appreciation also extends to all my dearest fellow friends and Intel's colleagues, a big thank you for their endless support and understanding during my work and study full time, especially during this difficult time in pandemic. Lastly, I wish to express my sincere thanks to all those who had contributed in completing my final year project. Thanks a lot.

ABSTRACT

Plant diseases limit the crop production and have received more attention from experts and farmers. Plant disease identification is carried out by experienced people or needs microscopic identification. However, trained people or professionals are not always available, and the manual approach may lead to bias or errors, costly and time-consuming, especially when some of plant disease symptoms are similar. It has also not easily been understood and identified that attacking crop could be due to parasitic organisms like fungus or bacteria besides the insect. To reduce the damage on the crops, plant disease early detection should be carried out in an automated way for early detection, prevention and control. Many methods have been proposed to do automated detection, but it is not easy to target which feature is the best for the classification. Thus, the objective of this project is to develop an automatic feature extraction method in identifying the severity of two types of plant diseases, namely early blight and late blight, which are caused by microorganism attacks. The main classifier module will be governed by autoencoders as an automatic feature extraction to identify the plant diseases. The MATLAB software was used to develop the autoencoder module. With the data set ready from Plant Village leaf images, this project identified two plant diseases into three severity levels, low, mild and severe at 72.7% accuracy.

ABSTRAK

Penyakit tumbuhan membatasi pengeluaran makanan dan mendapat perhatian lebih dari pakar dan petani. Pengenalpastian penyakit tumbuhan dilakukan oleh orang yang berpengalaman atau memerlukan pengenalan mikroskopik. Walau bagaimanapun, orang terlatih atau profesional tidak selalu didapati, dan pendekatan manual menyebabkan bias atau kesalahan, mahal dan memakan masa, terutama apabila beberapa gejala penyakit tumbuhan adalah serupa. Ia juga tidak mudah memahami dan mengenalpasti tumbuhan yang menyerang sama ada disebabkan oleh organisma parasit seperti kulat atau bakteria selain serangga. Untuk mengurangkan kerosakan pada tanaman, pengesanan awal penyakit tanaman harus dilakukan secara automatik untuk pengesanan, pencegahan dan pengendalian. Banyak kaedah telah dicadangkan untuk melakukan pengesanan automatik, tetapi tidak mudah untuk mendapatkan ciri daun yang terbaik untuk klasifikasi. Oleh itu, objektif projek ini adalah untuk mengembangkan kaedah pengekstrakan ciri daun automatik dalam mengenal pasti keparahan dua jenis penyakit tumbuhan, iaitu penyakit “early blight” dan “late blight”, yang disebabkan oleh serangan mikroorganisma. Modul pengkelasan utama akan diatur oleh *autoencoder* sebagai pengekstrakan ciri automatik untuk mengenal pasti penyakit tanaman. Perisian MATLAB akan digunakan untuk mengembangkan modul *autoencoder*. Dengan set data dari gambar daun Plant Village, projek ini mengenal pasti dua penyakit tumbuhan kepada tahap penyakit rendah,sederhana dan teruk dengan kejituan sebanyak 72.7%.

TABLE OF CONTENTS

TITLE		PAGE
DECLARATION		**iii**
DEDICATION		**iv**
ACKNOWLEDGEMENT		**v**
ABSTRACT		**vi**
ABSTRAK		**vii**
TABLE OF CONTENTS		**viii**
LIST OF TABLES		**x**
LIST OF FIGURES		**xi**
LIST OF ABBREVIATIONS		**xii**
LIST OF APPENDICES		**xiii**
CHAPTER 1	**INTRODUCTION**	**1**
1.1	Research Background	1
1.2	Problem statement	1
1.3	Objectives	2
1.4	Scope of Work	2
1.5	Report Outline	3
CHAPTER 2	**LITERATURE REVIEW**	**5**
2.1	Overview	5
2.2	Introduction of autoencoder	5
2.3	Related works	7
CHAPTER 3	**RESEARCH METHODOLOGY**	**15**
3.1	Project flow	15
3.2	Plant disease severity level	16
3.3	Proposed System Methodology	17
	3.3.1 Training phase for autoencoder	19

3.3.2 Testing phase 21

3.4 Summary 23

CHAPTER 4 RESULTS AND DISCUSSION 25

4.1 Performance Comparison with different Max Epoch 25

4.2 Performance Comparison with different Hidden Size 26

4.3 Performance Comparison with different L2 Regularization 26

4.4 Performance Comparison with single layer and double layer of autoencoder classification 28

4.5 Analysis on the best accuracy trained model (Experiment 4b) 31

4.6 Summary 33

CHAPTER 5 CONCLUSION AND RECOMMENDATIONS 34

5.1 Project Achievement 34

5.2 Recommendation for Future Works 35

REFERENCES 36

Appendices A - E 39 - 44

LIST OF TABLES

TABLE NO.	TITLE	PAGE
Table 2.1	Summary of the related works	13
Table 3.1	Severity of Early blight and Late blight	17
Table 4.1	Performance Comparison with different Max Epoch	25
Table 4.2	Performance Comparison with different Hidden Size	26
Table 4.3	Performance Comparison with different L2 Regularization	26
Table 4.4	Performance comparison with single layer and double layer of autoencoder classification with other parameters	29
Table 4.5	Analysis on the best accuracy trained model (Experiment 4b)	31
Table 4.6:	Performance the trained model with double autoencoder layers	33

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
Figure 2.1	Structure of autoencoder	5
Figure 2.2	Main steps of hyperscale imaging study	8
Figure 2.3	Rice disease images pre-processing procedure	10
Figure 2.4	The structure of stacked autoencoder	11
Figure 3.1	Overall project proposed flow chart	18
Figure 3.2	Single layer autoencoder stacked network	20
Figure 3.3	Two layer autoencoder stacked network	20
Figure 3.4	The classification of test set	22
Figure 4.1	Weight of the first layer autoencoder from Experiments A-I	27
Figure 4.2	Weight of the autoencoders a) first layer autoencoder	31
Figure 4.3	Mean Squared Error with L2 and Sparsity Regularizers	31
Figure 4.4:	Accuracy of the autoencoder a) single layer b) double layers	32

LIST OF ABBREVIATIONS

AE1	-	Autoencoder 1
AE2	-	Autoencoder 2
CNN	-	Convolution Neural Network
CPU	-	Central Processing Unit
DL	-	Deep Learning
Exp	-	Experiment
FN	-	False Negative
FP	-	False Positive
GA	-	Genetic Algorithm
HIS	-	Hue, Saturation, Intensity
MaxEpoch	-	Maximum epoch
RGB	-	Red Green Blue
SGDM	-	Spatial Gray-Level Dependence
SVT	-	Support Vector Machine
SL	-	Supervised Learning
TN	-	True Negative
TP	-	True Positive

LIST OF APPENDICES

Appendix A Renaming code	39
Appendix B Pre-processing, reading images and convert to grey images	40
Appendix C Training Autoencoders	42
Appendix D Testing and result tabulating	43
Appendix E More plant leave images with different severity	44

CHAPTER 1

INTRODUCTION

1.1 Research Background

There are more than 800 million people in this world who do not have sufficient food consumption. 1.3 billion of them live on less than $1 daily and more than 10% of food production lost is due to plant disease globally [1]. Monitoring of plant disease and health plays an important role in crops cultivations and agriculture production. In the early days, plant disease was detected via manual monitoring and analysis. However, this is ineffective and late plant disease discovery could lead to a significant reduction in the crop's quality and quantity. Tomato and potato are some of the most widely cultivated food crops. These two plants are also commonly attacked by plant diseases, namely early blight and late blight. These two plant diseases are caused by microorganism and fungi. The early disease can cause the lower leaves to be completely destructed, expand rapidly, causes the stem to collapse quickly and even lead to the death of the infected plant and affect the yield. Therefore, early plant disease detection is needed to reduce the damage to the crops.

1.2 Problem statement

Traditionally, plant disease detection is done visually by professional or experienced people. However, these trained people are not always available and sometimes the process even needs microscopic identification. Plant disease detection in image-based an essential solution and a crucial part of precision agriculture to detect plant diseases and classify the severity of variability in crops. Besides, some plant disease is extremely difficult to control especially when the plant disease can be spread quickly and cause the crops to be destructed in a very short time. As disease outbreaks are becoming more concerned across the globe, early disease detection can be used for

disease diagnosis control and analysis. If early detection cannot be done and the plant disease cannot be controlled from being spread out, this can lead to catastrophe and famine [2]. From the literature review, several methods have been proposed to do automated detection. Most of the feature extraction process involved handcrafted, complex algorithm or complicated mathematical formula. Thus, it is not easy to target which extraction feature is the best for the classification. Besides, it is difficult to detect the plant disease for the symptoms that are very similar. For example, early blight and late blight has similar characteristic or symptoms. These features need to be extracted for a better classification process.

1.3 Objectives

The aim for this project is to develop the feature extraction and classify the plant disease severity. The objectives of this project are listed as below:

(a) To develop an automatic feature extraction method using autoencoder in identifying two plant diseases, namely early blight and late blight.

(b) To classify early blight and late blight into three severity levels, low mild and severe.

1.4 Scope of Work

The scope of this project is to focus on the feature extraction of the images of early blight and late blight disease and classify them into three severity level. The main classifier module will be governed by autoencoders as an automatic feature extraction to identify the plant diseases. The engineering software, MATLAB will be used to develop the autoencoder module.

1.5 Report Outline

This project report is divided into 5 chapters. Chapter 1 is the project background, objectives, and the scope of work. Chapter 2 is the literature review for this project. Simple autoencoder and previous related works were discussed. Chapter 3 covers the research methodology. In this chapter, the project flow and proposed methodology was discussed. Chapter 4 shows the results obtained from the proposed method. Lastly, Chapter 5 summarized the conclusions, discussion and recommendation for future work.

REFERENCES

[1] R. N. Strange and P. R. Scott, “Plant Disease : A Threat to Global Food Security,” no. February 2005, 2016, doi: 10.1146/annurev.phyto.43.113004.133839.

[2] P. S. Ojiambo, J. Yuen, F. Van Den Bosch, and L. V. Madden, “Epidemiology: Past, present, and future impacts on understanding disease dynamics and improving plant disease management — A summary of focus issue articles,” *Phytopathology*, vol. 107, no. 10, pp. 1092–1094, 2017, doi: 10.1094/PHYTO-07-17-0248-FI.

[3] Q. V. Le, “Building high-level features using large scale unsupervised learning,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, pp. 8595–8598, 2013, doi: 10.1109/ICASSP.2013.6639343.

[4] A. Sankaran, M. Vatsa, R. Singh, and A. Majumdar, “Group sparse autoencoder,” *Image Vis. Comput.*, vol. 60, pp. 64–74, 2017, doi: 10.1016/j.imavis.2017.01.005.

[5] B. A. Olshausen and D. J. Fieldt, “Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1 ? Coding V1 Gabor-wavelet Natural images,” *Vis. Res*, vol. 37, no. 23, pp. 3311–3325, 1997.

[6] Y. Dandawate and R. Kokare, “An automated approach for classification of plant diseases towards development of futuristic Decision Support System in Indian perspective,” *2015 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2015*, pp. 794–799, 2015, doi: 10.1109/ICACCI.2015.7275707.

[7] C. Xie, Y. Shao, X. Li, and Y. He, “Detection of early blight and late blight diseases on tomato leaves using hyperspectral imaging,” *Sci. Rep.*, vol. 5, pp. 1–11, 2015, doi: 10.1038/srep16564.

[8] M. S. Hossain, R. M. Mou, M. M. Hasan, S. Chakraborty, and M. Abdur Razzak, “Recognition and detection of tea leaf’s diseases using support vector machine,” *Proc. - 2018 IEEE 14th Int. Colloq. Signal Process. its Appl. CSPA 2018*, no. March, pp. 150–154, 2018, doi: 10.1109/CSPA.2018.8368703.

[9] Y. Tian, C. Zhao, S. Lu, and X. Guo, “SVM-based Multiple Classifier System for recognition of wheat leaf diseases,” *World Autom. Congr. Proc.*, pp. 2–6, 2012.

[10] H. Al-Hiary, S. Bani-Ahmad, M. B. M. Reyalat, and Z. ALRahamneh, “Fast and Accurate Detection and Classification of Plant Diseases,” *Int. J. Comput. Appl.*, vol. 7491 LNCS, no. PART 1, pp. 72–81, 2012, doi: 10.1007/978-3-642-32937-1_8.

[11] Y. Lu, S. Yi, N. Zeng, Y. Liu, and Y. Zhang, “Identification of rice diseases using deep convolutional neural networks,” *Neurocomputing*, vol. 267, pp. 378–384, 2017, doi: 10.1016/j.neucom.2017.06.023.

[12] E. Fujita, Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, “Basic investigation on a robust and practical plant diagnostic system,” *Proc. - 2016 15th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2016*, pp. 989–992, 2017, doi: 10.1109/ICMLA.2016.56.

[13] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, “Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification,” *Comput. Intell. Neurosci.*, vol. 2016, 2016, doi: 10.1155/2016/3289801.

[14] A. M. Abdu, M. M. Mokji, and U. U. Sheikh, “Machine learning for plant disease detection: An investigative comparison between support vector machine and deep learning,” *IAES Int. J. Artif. Intell.*, vol. 9, no. 4, pp. 670–683, 2020, doi: 10.11591/ijai.v9.i4.pp670-683.

[15] H. F. Pardede, E. Suryawati, R. Sustika, and V. Zilvan, “Unsupervised Convolutional Autoencoder-Based Feature Learning for Automatic Detection of Plant Diseases,” *2018 Int. Conf. Comput. Control. Informatics its Appl. Recent Challenges Mach. Learn. Comput. Appl. IC3INA 2018 - Proceeding*, pp. 158–162, 2019, doi: 10.1109/IC3INA.2018.8629518.

[16] V. Zilvan, A. Ramdan, E. Suryawati, R. B. S. Kusumo, D. Krisnandi, and H. F. Pardede, “Denoising Convolutional Variational Autoencoders-Based Feature Learning for Automatic Detection of Plant Diseases,” pp. 1–6, 2020, doi: 10.1109/icicos48119.2019.8982494.

[17] M. M. Yang, A. Nayeem, and L. L. Shen, “Plant classification based on stacked autoencoder,” *Proc. 2017 IEEE 2nd Inf. Technol. Networking, Electron. Autom. Control Conf. ITNEC 2017*, vol. 2018-Janua, pp. 1082–1086, 2018, doi: 10.1109/ITNEC.2017.8284906.

[18] W. Jia, M. Yang, and S. H. Wang, “Three-Category Classification of Magnetic Resonance Hearing Loss Images Based on Deep Autoencoder,” *J. Med. Syst.*, vol. 41, no. 10, 2017, doi: 10.1007/s10916-017-0814-4.