DISEASE DETECTION OF SOLANACEOUS CROPS USING DEEP LEARNING

NURUL HIDAYAH BINTI A. HALIM

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DISEASE DETECTION OF SOLANACEOUS CROPS USING DEEP LEARNING METHOD

NURUL HIDAYAH BINTI A. HALIM

This report is submitted in partial fulfilment of the requirements for the degree of Bachelor of Electronic Engineering with Honours

Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka

2022

UNIVERSITI TEKNIKAL MALAYSIA MELAKA
FAKULTI KEJUTERAAN ELEKTRONIK DAN KEJURUTERAAN KOMPUTER

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : Disease Detection of Solanaceous Crops using Deep Learning

Sesi Pengajian : 2021/2022

Saya NURUL HIDAYAH BINTI A. HALIM mengaku membenarkan laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (✓):

☐	**SULIT***	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
☐	**TERHAD***	(Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan.
✓	**TIDAK TERHAD**	

Disahkan oleh:

(NURUL HIDAYAH BINTI A. HALIM)

(PM. DR. SYAFEEZA BINTI AHMAD RADZI)

Alamat Tetap: No 21 Lorong Masjid, Kampung Jimah Baru, 71950, Seremban, Negeri Sembilan

PROF. MADYA DR. SYAFEEZA BINTI AHMAD RADZI
Profesor Madya
Fakulti Kejuruteraan Elektronik Dan Kejuruteraan Komputer
(FKEKK)
Universiti Teknikal Malaysia Melaka (UTeM)
Hang Tuah Jaya
76100 Durian Tunggal, Melaka, Malaysia
06-270 2372 / 011-3234 5124
syafeeza@utem.edu.my

Tarikh : 21 Jun 2022

Tarikh : 21 Jun 2022

*CATATAN: Jika laporan ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali tempoh laporan ini perlu dikelaskan sebagai SULIT atau TERHAD.

DECLARATION

I declare that this report entitled "Disease Detection of Solanaceous Crops using Deep Learning" is the result of my own work except for quotes as cited in the references.

Signature : ……………………………………

Author : NURUL HIDAYAH BINTI A. HALIM

Date : 21 JUNE 2022

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Bachelor of Electronic Engineering with Honours.

Signature : ....

Supervisor Name : PM. DR. SYAFEEZA BINTI AHMAD RADZI

Date : 21 JUNE 2022

DEDICATION

I dedicate all this hard work to my creator Almighty God Allah SWT for always eases everything in my journey.

ABSTRACT

Nowadays, Artificial Intelligence has become a synonym in our lives. The technology of AI has helped so many people to solve their problems. In agriculture, crop diseases are more likely to occur because of many factors such as shifting weather, lack of nutrition, and pest attacks. This study looks into the solanaceous crops disease for leaf and its fruit. Four types of solanaceous crops were used to be investigated in this project, namely tomato, pepper, eggplant, and potato. The deep learning method of the CNN architecture-based YOLOv5 model was applied in this project to detect the disease of solanaceous crops. The work involved four types of crops, including 23 classes of healthy and disease crops that infected the leaf and fruits. The total dataset of all 23 classes is 16580 images and has been divided into three parts: training set, validation set, and testing set. The dataset that was used for training is 88% of the total dataset (15000 images), 8% of the dataset performed a validation process (1400 images), and the rest of the 4% dataset was used for the test process (699 images). The performance of YOLOv5 has been compared with YOLOv4 to prove that this model is more robust and better in terms of accuracy. The training, validation, and testing dataset simulation was performed on a Google Colab notebook.

ABSTRAK

Pada masa kini, Kepintaran Buatan telah menjadi sinonim dalam kehidupan kita. Teknologi AI telah membantu ramai orang untuk menyelesaikan masalah mereka. Dalam industri pertanian, penyakit tanaman lebih kerap berlaku kerana banyak faktor seperti perubahan cuaca, kekurangan nutrisi dan serangan serangga perosak. Kajian ini mengkaji penyakit tanaman solanaceous untuk daun dan buahnya. Empat jenis tanaman solanaceous digunakan untuk projek ini iaitu tomato, lada, terung dan kentang. Kaedah pembelajaran mendalam model YOLOv5 berasaskan seni bina CNN telah digunakan dalam projek ini untuk mengesan penyakit tanaman solanaceous. Projek ini melibatkan empat jenis tanaman, termasuk 23 kelas tanaman sihat dan berpenyakit yang menjangkiti daun dan buah. Jumlah set data kesemua 23 kelas ialah 16580 imej dan telah dibahagikan kepada tiga bahagian: set latihan, set pengesahan dan set ujian. Set data yang digunakan untuk latihan ialah 88% daripada jumlah set data (15000 imej), 8% set data melakukan proses pengesahan (1400 imej), dan selebihnya set data 4% digunakan untuk langkah ujian (699 imej).). Prestasi YOLOv5 telah dibandingkan dengan YOLOv4 untuk membuktikan model ini lebih teguh dan lebih baik dari segi ketepatan. Latihan, pengesahan dan simulasi set data ujian telah dilakukan pada buku nota Google Colab.

ACKNOWLEDGEMENTS

Foremost, I want to thank my lovely parents, A. Halim Bin Sodimin and Zarina Binti Khamis, for always supporting me in all conditions and giving spirit to never give up during my degree journey, especially during my final year semester.

My sincere thanks to Muhamad Haziq Bin Shaari for his willingness to borrow his laptop each time my laptop faced a problem. To my little sister, Nurul Umairah Binti A. Halim, thank you for always aiding me in cleaning the house when I am busy doing my final year project and thesis.

Finally, to my supervisor, Associate Professor Dr. Syafeeza Binti Ahmad Radzi, thank you so much for constantly guiding me and sharing your knowledge with me during this project. I would also like to express my gratitude to my panels Dr. Khairuddin bin Osman and Dr. Siti Fatimah Binti Sulaiman, for their encouragement and insightful comments that helped me improve my work on this project. Last but not least, I hope your kindness will be repaid.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AI	:	Artificial Intelligence
CNN	:	Convolutional Neural Network
YOLO	:	You Only Look Once
EXIF	:	Exchangeable Image File Format
CPU	:	Central Processing Unit
YAML	:	Yet Another Markup Language
mAP	:	Mean Average Precision
IoU	:	Intersection Over Union
SDG	:	Sustainable Development Goals

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

This chapter will begin with an overview of this project's background, objective, problem statement, the scope of work, and thesis outline for understanding the project's primary purpose.

1.1 Project Background

Agriculture has long been a vital economic and social sector. In general, crop disease detection has usually been carried out manually using visual inspection or microscope techniques, which are time-intensive and prone to inaccuracy leading to different human vision and error information [1]. The primary purpose of this project is to detect various diseases on solanaceous crops that infected at leaf and their fruit. In agriculture, detecting diseases as soon as possible is essential to prevent them from infecting other healthy crops. To overcome this problem, a deep learning model based on YOLOv5, which is the latest object detection algorithm model in the YOLO family, has been used for this project [2]. The performance of YOLOv5 will be evaluated to prove the accuracy of disease detection of solanaceous crops using this model.

1.2 Objective

These are several objectives to achieve at the end of this project:

1. To implement a deep learning model through an object detection algorithm (YOLOv5) to detect the diseases on the leaf and fruit of solanaceous crops.
2. To analyze the performance of the YOLOv5 model in terms of detection accuracy.

1.3 Problem Statement

It is difficult for manpower to detect crop diseases early in the crop growing process accurately. Small mistakes or missteps are usually unavoidable when using the workforce, especially when identifying the plant's type of disease because human eyes are prone to errors and are time-consuming. However, disease and pest control challenges still haunt some local farmers. This project aims to help the workforce use AI towards the Industrial Revolution 4.0. As a result, disease detection requires regular crop monitoring throughout the growing period. Today, deep learning-enabled developments in computer vision have led to a situation where disease diagnosis is dependent on automatic recognition. Therefore, this project will use an object detection technique that implements the YOLOv5 model to detect the type of disease of solanaceous crops. The output of this project is expected to aid farmers in improving their agricultural abilities to detect disease in the solanaceous crops by using this system model.

1.4 Scope of Work

This project aims to develop a system to detect and localize images of affected diseases of solanaceous crops based on the deep learning method. This project needs thousands of samples to produce an excellent accuracy for implementing the system. The dataset of disease and healthy leaf and fruit of solanaceous crops were collected throughout the Plant village dataset on GitHub and Kaggle and captured by own mobile camera. The dataset has been divided into several classes: their own healthy and disease leaf and fruit names. After dividing the dataset of their class, the annotation process is performed, drawing the bounding box around the area of the healthy and disease images of the solanaceous crops. Data augmentation was applied to increase the dataset for performance accuracy to solve this problem. This process is done on the Roboflow.ai website, which provides free space for annotation, generates auto data augmentation, and exports files to specific formats that can be used for this system. The simulation process to train this system is conducted in Google Colab notebook as it provides free GPU to run the simulation. The proposed object detection model, YOLOv5, is used to train and test the dataset, and this model will be compared with the previous YOLO family, which is YOLOv4. In the end, the performance of these models is evaluated through the system's accuracy.

1.5 Thesis Outline

The first chapter provides an overview and explanation of a method for identifying the type of disease of solanaceous crops using the deep learning method. The straightforward statement of the issue demonstrates the project’s significance. The research issue, hypothesis, or assumption must be followed, then the subject of this chapter. This study discusses the advancements made in comparison to previous studies. The project aims to measure, plan, and implement a deep learning model to identify the different variations of disease from solanaceous crops using the CNN architecture based on the YOLOv5 model and get accurate system results. The scope of work is being clarified in detail to set the limitation of the project.

The second chapter discusses the literature, including the summary of papers and journals used as sources for this research. The topics related to the study are being collected and discussed in this chapter to achieve the project's objectives and solve the stated problems. This chapter covers the history of AI, methods used in the project, deep learning, and analyzing project-related journals. Journals related to this project have been summarized in this section.

The third chapter is a project methodology. This chapter explains how to implement the plan to meet the project's goals. This chapter explains the software and diagrams in detail. Furthermore, each step of the flow chart is broken down into phases. Each phase's techniques or algorithms are mentioned.

Next is the fourth chapter. This chapter will present the overall result, including the analysis of YOLOv5, and will be compared with the previous YOLO, YOLOv4. The discussion for the performance analysis will be interpreted to prove the accuracy of this model in detecting the diseases of solanaceous crops.

The last chapter, Chapter 5, will conclude everything about this project. The future works are also stated to improvise the project. These recommendations may help other researchers to explore further future system improvement.

CHAPTER 2

BACKGROUND STUDY

This chapter discusses the literature of several related journals, articles, and previous studies on Artificial Intelligence (AI), Deep Learning, and Convolution Neural Network (CNN). The previous work related to the YOLO family's performance has also been discussed in this chapter.

2.1 Introduction of Artificial Intelligence

AI is a computer science subfield. It includes implementing computer programs to perform tasks that would otherwise necessitate human intelligence. Learning, vision, problem-solving, language comprehension, and logical reasoning are all areas where AI algorithms might help [3]. AI is rapidly becoming prevalent due to its robust usability in situations that people cannot handle well, its rapid technological improvement, and the broad range of applications [4]. In 1950, Alan Turing claimed that a machine's intelligence might be determined by its ability to demonstrate intelligent behavior, which is different from average humans with good intelligence [5]. AI tries to duplicate human problem-solving techniques in the public sector to produce more efficient solutions. A unique feature in replicating human thinking, learning, and problem-solving capabilities is intended to enhance performances by incorporating AI technology into a computer application with human-computer interaction and data interaction, also referred to as AI application [6].

AI technology is now widely used in various industries worldwide and positively impacts manufacturing, healthcare, and agriculture. Agriculture is an extreme industry, with 30.7% contributing to economic progress [7]. In terms of accuracy and robustness, AI is at its best to support agricultural systems. Agriculture is a dynamic sector in which it is impossible to generalize situations to propose a standard solution. [8].

2.2 Deep Learning

Deep Learning extends traditional machine learning by adding more ‘depth’ (complexity) to the model and modifying the data using several features that allow data representation in a hierarchical form through multiple levels of abstraction [9]. If a large dataset describing the problem exists, these complex models used in Deep Learning can reduce errors in regression problems and improve classification accuracy.

Deep Learning has many different components like convolution layers, fully connected layers, pooling layers, memory cells, gates, encode/decode schemes, activation functions, etc., depending on the network architecture that implements (i.e., Convolutional Neural Networks, Recursive Neural Networks, Unsupervised Pre-trained Networks, Recurrent Neural Networks). Deep Learning models perform exceptionally well in prediction and classification because of their large learning capacity and highly hierarchical structure. They are also flexible and adaptable to various highly complex (from a data analysis perspective) challenges because of their highly hierarchical structure and large learning capacity [10].

2.3 Convolution Neural Network

CNN is an algorithm of deep learning which takes images as an input and then learns various aspects from the images. From the images, it can differentiate the output of the images of one from another [11]. A CNN mixes well-read features with a training dataset. Then it employs 2D convolutional layers, making this architecture more ideal for interpreting 2D data, such as images, than other types of architecture. CNNs will remove the manual feature extraction and the demand for image classification. CNN of its model extracts the data directly from images. The extracted features are not pre-trained; they are well-read when the network is trained on a small set of image groups. CNN architecture consists of numerous layers that perform image processing. These layers include input, multiple hidden, and output layers. The hidden layers, which are typically made up of numerous convolutional layers (CL), pooling layers, and a set of fully connected layers (FC), are where the last work is done [12].

Several researchers in this field have embraced the idea of implementing a deep learning method of developing neural networks to train the frameworks model to recognize types of crops and detect the disease that infected using images. In this paper, the researchers utilize deep learning for image-based disease identification. CNN oversees both feature extraction and objects classification. CNN is used to train the framework to analyze 26 diseases of 14 crop species using a dataset of 54,306