CABBAGE DISEASE DETECTION SYSTEM USING K-NN ALGORITHM

MOHAMAD AINUDDIN BIN SAHIMAT

FACULTY OF COMPUTING AND INFORMATICS

UNIVERSITI MALAYSIA SABAH

2022

CABBAGE DISEASE DETECTION SYSTEM USING K-NN ALGORITHM

MOHAMAD AINUDDIN BIN SAHIMAT

THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF BACHELOR OF 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 excerpts and summaries, each of which I have explained the source.

17 January 2022

MOHAMAD AINUDDIN BIN SAHIMAT

BI18110050

CONFIRMATION OF REPORT

NAME	**:**	**MOHAMAD AINUDDIN BIN SAHIMAT**
MATRIC NO.	**:**	**BI18110050**
TITLE	**:**	**CABBAGE DISEASE DETECTION SYSTEM USING K-NN ALGORITHM**
DEGREE	**:**	**ENGINEERING NETWORKING (HC05)**
VIVA DATE	**:**	**17 JANUARY 2022**

VERIFY BY;

SUPERVISOR	**SIGNATURE**
PN. SALMAH BINTI FATTAH	SALMAH FATTAH Lecturer Faculty of Computing and Informatics Universiti Malaysia Sabah

ACKNOWLEDGEMENT

All praise and gratitude I extend to Allah Subhanahu Wa Ta'ala because with His bounty I was able to be patient and given good health as long as I conducted this study. I am deeply indebted to everyone who, directly and indirectly, was responsible for this successful project. I am very grateful to Pn. Salmah Binti Fattah as my supervisor for her encouragement, positive criticism of my research in this project, and most importantly all her guidance and support for the whole project. Finally, many thanks to my parents, family and friends who have gone through this long process with me in their full guidance and support.

ABSTRACT

Identification of plant diseases is key to avoiding losses in agricultural yields and product quantities. Plant disease study means the study of disease patterns that can be visually seen on plants. The main objective of this research is to develop a prototype system with the help of machine learning to detect cabbage diseases which are Alternaria Leaf Spot disease, Mosaic Virus disease, Downy Fungus disease, Bacterial Soft Rot disease, and Black Rot disease . It is very difficult to monitor plant diseases manually because it requires a large amount of work, deep expertise in plant diseases, and also requires excessive processing time. The image sample pixel will need to convert first using an otsu method and histogram method in the image processing and segmentation technique. Then, the segmented cabbage sample will use the GLCM method for feature extraction. It is a method of extracting second-order statistical texture features to detect diseases more efficiently. Finally, the KNN algorithm will be used to classify the disease based on sample nature and a cabbage disease data set. Consequently, by employing the KNN technique, the cabbage diseases are recognized at average 90% percent accuracy rates. This prototype has a very great potential to be further improved in the future.

Keyword: Machine Learning, KNN Algorithm, GLCM, Otsu method, Histogram method

ABSTRAK

Pengenalpastian penyakit tumbuhan adalah kunci untuk mengelakkan kerugian dalam hasil pertanian dan kuantiti produk. Kajian penyakit tumbuhan bermaksud kajian corak penyakit yang boleh dilihat secara visual pada tumbuhan. Objektif utama penyelidikan ini adalah untuk membangunkan sistem prototaip dengan bantuan pembelajaran mesin untuk mengesan penyakit kubis iaitu penyakit Bintik Daun Alternaria, penyakit Virus Mosaik, Penyakit Kulat Downy, Penyakit Reput Lembut Bakteria dan Penyakit Reput Hitam. Sangat sukar untuk memantau penyakit tumbuhan secara manual kerana ia memerlukan jumlah kerja yang besar, kepakaran mendalam dalam penyakit tumbuhan, dan juga memerlukan masa pemprosesan yang berlebihan. Piksel sampel imej perlu ditukar terlebih dahulu menggunakan kaedah otsu dan kaedah histogram dalam pemprosesan imej dan teknik pembahagian. Kemudian, sampel kubis tersegmen akan menggunakan kaedah GLCM untuk pengekstrakan. Ia adalah kaedah mengekstrak tekstur statistik peringkat kedua untuk mengesan penyakit dengan lebih cekap. Akhir sekali, algoritma KNN akan digunakan untuk mengklasifikasikan penyakit berdasarkan sifat sampel dan set data penyakit kubis. Akibatnya, dengan menggunakan teknik KNN, penyakit kubis diiktiraf pada purata kadar ketepatan 90% peratus. Prototaip ini mempunyai potensi yang sangat besar untuk dipertingkatkan lagi pada masa hadapan.

Kata kunci: Pembelajaran Mesin, Algoritma KNN, GLCM, kaedah Otsu, kaedah Histogram

TABLE OF CONTENT

CHAPTER 4 : ANALYSIS AND DESIGN

CHAPTER 5 : IMPLEMENTATION

LIST OF TABLE

LIST OF FIGURE

LIST OF ABBREVIATIONS

KNN - K Nearest Neighbour

ROI - Region Of Interest

ML - Machine Learning

SVM - Support Vector Machine

CNN - Convolutional Neural Network

NN - Neural Network

GLCM - Gray Level Co-Occurrence Matrix

DFD - Data Flow Diagram

ERD - Entity Relationship Diagram

LIST OF APPENDIX

CHAPTER 1

INTRODUCTION

This chapter discusses an overview of the study conducted. The project title is Cabbage Disease Detection System Using K-NN Algorithm. It consists of problem statements and motivation, objectives, the scope of the study, and project impact on society. Problem statements describe the problems that arise and make the selected projects to be undertaken. The objectives are the goals list for the projects to be achieved. The scope of the study discusses the limitations of projects and users. Lastly, the impact, significance, and contribution of the proposed project to society.

1.1 BACKGROUND STUDY

Agriculture plays an important role in our daily life because it forms the basis for the food security of living things. Besides, agriculture also helps our country's economy. Based on an article from the nation's encyclopedia "Malaysia-Agriculture", agriculture remains an important sector of Malaysia's economy, contributing 12 percent to the national GDP and providing employment for 16 percent of the population. Although agriculture can contribute to our country's economy, there are various problems preventing the farmers from gaining maximum profit from their crops due to plant disease which can affect crops on their farm. Based on the CropLife website, the article states that annual crop losses due to pests and diseases amount to Rs.50,000 crore ($500 billion), which is significant in a country where at least 200 million Indians go to bed hungry every night. In the past, the farmer usually relied on their eye observation and waited until the disease symptom on the plant appears. Sometimes this method is not reliable. If this problem cannot be resolved, not only farmers will lose their income but our country also will suffer due to food problems and need to import more food from foreign countries.

1.2 PROBLEM STATEMENT

Plant diseases have turned into a dilemma as they can cause significant reductions in both the quality and quantity of agricultural products. Plant disease detection system is an important research topic as it may prove beneficial in monitoring large crop farms, and thus automatically detect disease symptoms as soon as they appear on plant leaves. The proposed system is a software solution for the automatic detection and classification of plant leaf diseases. The system consists of four main steps namely image processing, image segmentation, feature extraction, and classification.

1.3 PROJECT OBJECTIVE

There are three objectives to achieve in this project:

I- To design a disease detection system that can detect disease for the farmer to reduce the risk of losing their harvest from diseases.

II- To develop and deploy a k-nearest neighbors (K-NN) algorithm that can perform disease classification on Cabbage plants using image processing.

III- To evaluate the effectiveness of an early precaution through the deployed disease detection system.

1.4 PROJECT SCOPE

I- The users of this system are farmers.

II- The prototype will be developed by using MATLAB.

III- Cabbage disease samples of the Alternaria Leaf Spot disease, Mosaic Virus disease, Downy Fungus disease, Bacterial Soft Rot disease, and Black Rot disease are used in this project.

1.5 IMPACT, SIGNIFICANCE, AND CONTRIBUTION

In this project, the disease detection system provides a user an easy way to monitor their crop. In a traditional agriculture method, farmers usually monitor their crops manually. This method might be working fine with a small-sized farm but if in a large-sized farm, this method will consume a lot of time and a tedious task to be done and sometimes the farmers might miss some area of the farms. If there is a disease happening in the missed area of the farms, it might spread to the other area of the farms too. There might be a chance where the farmer could use the wrong cure method and this can cause them a big loss as the chemical solution is expensive. In conclusion, the proposed system can provide a much better and more convenient way to reduce the problems that the farmers have when monitoring their crops. By using this system, an accurate disease classification system is higher than eye observation and users can take early precautions before the disease spreads.

1.6 ORGANIZATION OF THE PROJECT

This report contains 7 chapters. Chapter 1 talks about the overview of the project, objective, problem background, project scope, and also the project organization. Chapter 2 contains the literature review related to this project based on journals and articles. Chapter 3 reviews and discuss the method used and the overall framework taken in building this project. The project system development life cycle model and other related methods used will help in organizing the development phases more systematically. Chapter 4 contains system analysis and design. The design of the method discussed in chapter 3 is included here that would include some of the structures and coding in this chapter. Chapter 5 contains the implementation of the system to the actual workable system. Actual screenshot coding will be included in this chapter. Chapter 6 discusses the results of the project, where detail of how the experiment is run and tested on the system, for this research project detail of the data analysis and model development and stated clearly in this chapter. Lastly, chapter 7 will summarize the whole project and explanations on future works and concluded them as well.

CHAPTER 2

LITERATURE REVIEW

This chapter will be discussing vegetables used in the proposed project. Moreover, exploring the existing system and a concept used are related to developing this project.

2.1 CABBAGE OVERVIEW

This section will explain the definition of cabbage. After that, this subsection briefly discusses on type of cabbage disease, symptoms, and management of cabbage disease.

2.1.1 Definition of Cabbage

Cabbage is a vegetable that is an edible portion of a plant. Vegetables are usually grouped according to the portion of the plant that is eaten such as leaves, stems, roots, tubers, bulbs, and flowers. Cabbage is one of the most utilized food plants and widely grown originated in ASIA. It is an important crop worldwide and over half of the world, the population relies on it for food.

2.1.2 Cabbage Disease and Symptoms

Many factors make cabbage production slow and less productive. One of the main factors is cabbage disease. This section will show the type of cabbage disease and the symptom of cabbage disease. This research will focus on three types of diseases, which are mosaic virus disease, Alternaria Leaf Spot disease, Downy Mildew disease, Bacterial Soft Rot disease, and Black Rot disease that mostly occur in cabbage.

Table 2.1 cabbage disease

No.	Disease	Disease Symptoms
1	Mosaic Virus disease 	• The cabbage head begins to take on a mottled or "mosaic-like" appearance with a smattering of various colored rings and blotches, which in some cases turn black and necrotic. • The veins of the cabbage leaves could also show signs of chlorosis. • The head of cabbage starts to look very sticky and not edible.
2	Alternaria Leaf spot disease 	• yellow V-shaped lesions at the leaf margins. • A diseased area of the leaf expands and turns brown, the leaf veins in the affected area may appear black. • The leaf collapses
3	Downy Mildew disease 	• The whole vascular system can turn black. • Small, light green-yellow lesions on the upper leaf surface. • Spots turn yellow as they enlarge.

4	Bacterial Soft Rot 	• Cabbage looks small, water-soaked areas appear, and rapidly enlarge. • Cabbage leaf becomes soft and mushy, and within a few days, the affected plant part may collapse. • An offensive odor usually is present.
5	Black Rot 	• Yellowing of the leaf margins. • The vascular system turns black. • Yellow "V" at the midrib of the leaf.

2.2 MACHINE LEARNING

Machine learning (ML) is the study of computer algorithms that can improve automatically through experience and by the use of data. It is a part of artificial intelligence. Machine learning algorithms build a model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to do so. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks. This thesis will discuss three types of machine learning that is KNN, SVM, and Decision tree.

2.2.1 K-Nearest neighbors (KNN) algorithm

K-NN algorithm is a simple, easy-to-implement supervised machine learning algorithm that can be used to solve both classification and regression problems. K-NN algorithm assumes that similar things exist nearby. In other words, similar things are near to each other.

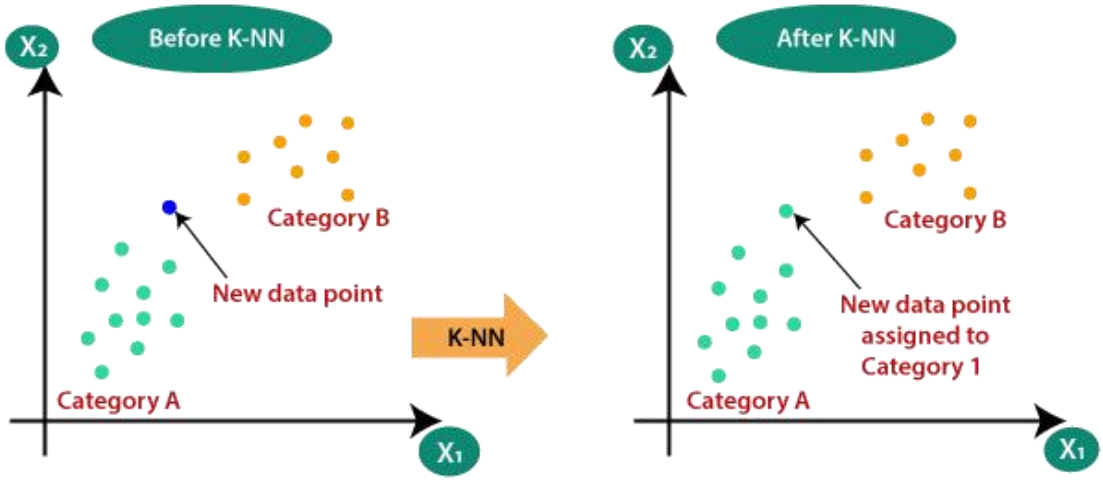

Figure 2.1 K-NN algorithm

Figure 2.1 shows that most of the time, similar data points are close to each other. The KNN algorithm hinges on this assumption being true enough for the algorithm to be useful. KNN captures the idea of similarity sometimes called distance, proximity, or closeness with some mathematics calculating the distance between points on a graph. KNN works by finding the distances between a query and all the examples in the data, selecting the specified number of examples (K) closest to the query, then voting for the most frequent classification or averages of regression.

2.2.2 Support Vector Machine (SVM) algorithm

Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The idea of SVM is simple, the algorithm creates a line or a hyperplane which separates the data into classes.

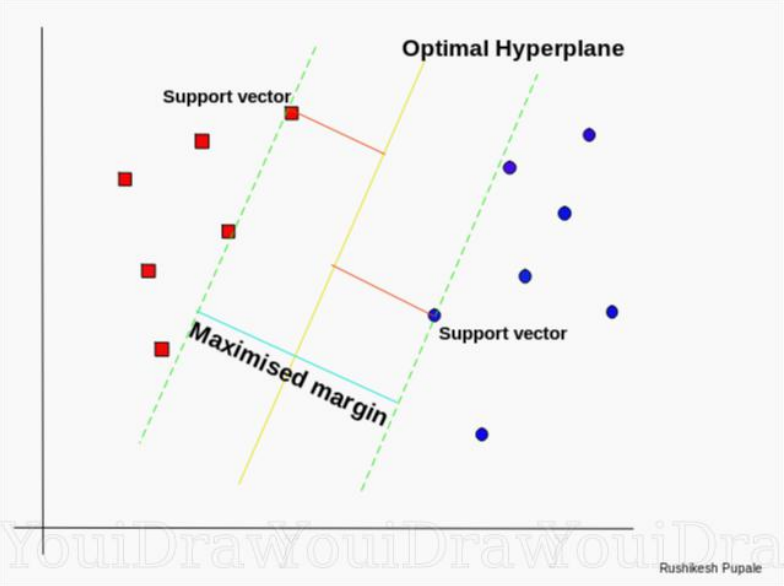

Figure 2.2 Hyperplane using SVM

Based on figure 2.2, SVM can classify data by adding an extra dimension to it so that it becomes linearly separable and then projecting the decision boundary back to original dimensions using mathematical transformation but finding the correct transformation for any given data set isn't easy.

2.2.3 Decision Tree

A decision algorithm is usually used to create a model that predicts the value of a target variable, for which the decision tree uses a tree representation to solve a problem where leaf nodes match class labels and attributes are represented on internal nodes. of the tree. It is made up of connected decision points. It has a starting point called the root, from the root it has a set of ongoing options that will be chosen. For every choice, the decision is made based on knowledge.