Republic of Iraq

Ministry of Higher Education

And Scientific Research

University of Diyala

College of Science

Plant Diseases Recognition Using CNN, Naive Bayes And Random Forest Algorithms

A Thesis

Submitted to the Computer Science Department /College of Science /University of Diyala

In a Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer.

By

Saif Aziz Salman

Supervised by

ASST. PROF. Dr. Bashar Talib AL-NUAIMI

2021 A.D. **1442 A.H.**

بسم اللہ الرحمن الرحيم

(يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوامِنكُم ◌ْ

وَالَّذِينَ أُوتُوا الْعِلْم ◌َ

دَرَجَاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ)

صدق اللہ العظيم

سورة المجادلة

الآيه 11

Dedication

I would like to dedicate this Work to:

Who taught me that the champions will never be defeated, but they convert it to victory.

Our Prophet Mohammed

Peace be Upon Him (PBH)

My Father and my Mother, And All whom supported me to Achieve This Work

SAIF **2021**

Acknowledgment

In the name of Allah the Merciful. I am grateful to my Creator who blessed me for having completed this thesis. I thank **Ass. Prof. Dr .Bashar Talib AL-NUAIMI** for his guidance and ideas to accomplish this work, and I thank him for those hours and ideas that he shared with me. I would like to introduce my thanks to the Computer Science Department at the college of Science of Diyala University for their education and assistance to me. I would like to express my special appreciation and deep thanks to all those who has helped to allow me to bring this letter. Finally, I would like to thank my family, who have endured the difficulties of this stage, throughout their days and nights, without their presence I would not have arrived for this day.

SAIF **2021**

Supervisor's Certification

I certify that this thesis entitled **"Plant Diseases Recognition Using CNN, Naive Bayes And Random Forest Algorithms"**, *was prepared under my supervision at Department of Computer Science/ College of Sciences/ University of Diyala by* **"Saif Aziz Salman"**, *as partial fulfillment of the requirements for the degree of* ***Master of Science in Computer Science***

(Supervisor)

Signature:

Name: Assist. Prof. Dr. Bashar Talib AL-NUAIMI

Date: / / 2021

Approved by the University of a Diyala Faculty of Science Department of Computer Science.

Signature:

Name: Assist. Prof. Dr. Bashar Talib AL-NUAIMI

Date: / / 2021

(Head of Computer Science Department)

Scientific certification

This is to certify that this thesis entitled **“Plant Diseases Recognition Using CNN, Naive Bayes And Random Forest Algorithms”** *was prepared by "**Saif Aziz Salman**" under my scientific supervision. It has been evaluated scientifically, therefore; it is suitable for debate by examining committee*.

Signature :

Name :

Date: / / 2021

Linguistic Certification

This is to certify that this thesis entitled **“Plant Diseases Recognition Using CNN, Naive Bayes And Random Forest Algorithms “was** *prepared under my linguistic supervision. It was amended to meet the style of the English language.*

Signature :

Name :

Date: / / 2021

Abstract

Nowadays, several diseases affect plants and, it is able to reason great damage and financial losses to the agricultural economy. It can even lead to great ecological losses. Employed information technology in plants disease detection and diagnosis is become required. Therefore, there is a need for an accurate system for plant disease detection and diagnosis by using the most effective deep learning techniques to avoid such losses.

In this work, An Automatic System for Diagnosis Plant Disease Based on Deep Convolutional Neural Network (DPD-CNN) has been proposed. Furthermore, this work distinguishes itself from the previous by employing the CNN with 12 nested processing layer. The CNN with 12 nested processing layer increased the accuracy of plant disease detection and diagnosis. Also, it has the ability to detect and diagnose the most common and dangerous types of plant disease which are Bacterial and fungal viral infections in early stage. The most common and effective feature selection algorithm which it is called Linear Discriminate Analysis (LDA) is employed in this work to extract the images features. The LDA algorithm have been employed for feature extraction in the stage of the two machine learning techniques. On the other hand, the most effective machine learning techniques for plant disease detection and diagnosis which are Naive Bayes, and Random Forest have been implemented. Furthermore, this work is proved that the deep learning is most effective in plant disease detection and diagnosis from machine learning.

However, according to the obtained results it is observed that the DPD-CNN achieved an excellent result with accuracy of **99.5**% during the comparison with Naïve which achieved accuracy of 97% and Random Forest which achieved

accuracy of 98%. Also, the performance of the DPD-CNN has been compared with the related work and it is achieved the highest accuracy.

List of Content

List of Tables

List of Figure

List of algorithms

List of Abbreviations

ANN	Artificial Neural Networks
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DCNN	Deep Convolutional Neural Network
DFT	Discrete Fourier Transform
FN	False Negative
FP	False Positive
KDD	Knowledge Discovery Database
KNN	K-Nearest Neighbors Algorithm
LDA	Linear Discriminant Analysis
MLP	Multi-Layer Perceptron
MM	Mathematical Morphology
PCA	Principal Component Analysis
PNN	Probabilistic Neural Networks
RELU	The Rectified Linear Unit
RNN	Recurrent Neural Network
SEs	Structure Elements
TN	True Negative
TP	True Positive

Chapter One
Introduction

CHAPTER ONE

INTRODUCTION

1.1 General Overview

This chapter introduces a brief introduction about plant diseases detection and its general framework. Also, it presents the literature review of research activities dealing with plant disease detection techniques, then, explaining the statement of the problem, research objectives, and the organization of the thesis.

In the agriculture sector, one of the major problems in plants is diseases. Plant diseases can be caused by various factors such as viruses, bacteria, fungus, etc. Most of the farmers are unaware of such diseases [1]. That’s why the detection of various diseases of plants is very essential to prevent the damages that it can make to the plants themselves as well as to the farmers and the whole agriculture ecosystem [2].

Plant diseases are a threat to the yield and nature of agricultural creation of the world and reduce much of the cost. It is explained that the loss caused by plant infection represents at least a 10% decrease in world food production [3]. Much of the assessment and treatment of diseases is done by farmers in the field under the guidance of plant pathologists. False diagnosis and the use of insecticides are common [4]. Therefore, the prevention and control of plant diseases have always been widely considered, as plants are exposed to the external climate and very sensitive to diseases. Usually, accurate and rapid determination of disease plays an important role in the fight against plant infections, as useful measures are regularly implemented after proper diagnosis [3, 5].

Much recognition and search techniques have been proposed according to the method of image distribution in a pipeline, including sample extraction and recognition. Recognition strategies with pipeline techniques have gained ground. However, these strategies have two problems. First, the accuracy of these techniques greatly depends on the extraction and selection of highlights from visible diseases. In particular, the highlights of visible manifestations of infection must be carefully separated and legal highlights selected. Second, the techniques that follow the duct method are quite complicated [6]. The presence of noises is highly unavoidable in pathologies taken under field conditions, such as unbalanced lighting and the base of the interfering field. This can significantly reduce the nature of the highlights and influence the consequences of recognition. In this sense, efforts are being made to eliminate the disorders through common techniques to obtain concrete results [7].

The current models for improving computer vision and using various artificial intelligence calculations to classify plant infections has shown promising results in some selected diseases and crops. The development of structures based on the Convolutional Neural Network (CNN) has further improved the accuracy of the characterization, but it is not reaches to the best [8]. Therefore, increase the layers of CNN is required to increase the classification accuracy.

However, CNN is the main algorithm used to locate and determine plant diseases [3, 4, and 5]. This work involves the use of a deep CNN-based model to support plant leaf identification problems. Deep CNN is a deep learning algorithm. Deep learning expands conventional AI by adding a broader multifaceted nature and different representations of flat information to the model. Deep CNNs have extensive applications in image organization, object identification, speech recognition, recommendation frameworks, and general language management [4].

Deep CNN requires tremendous information preparation to get better results. Also, image magnification is usually needed to improve the model presentation to build the best Deep CNN model with a lack of preliminary information. Image magnification provides incorrect image preparation using some manipulation techniques or a combination of different preparation strategies, for example, Invert Image, Gamma correction, Noise Infusion, Examination Shadows, main segment (PCA), rotation and scaling [3, 4]. Figure 1 shows the general flow chart for identifying plant infections.

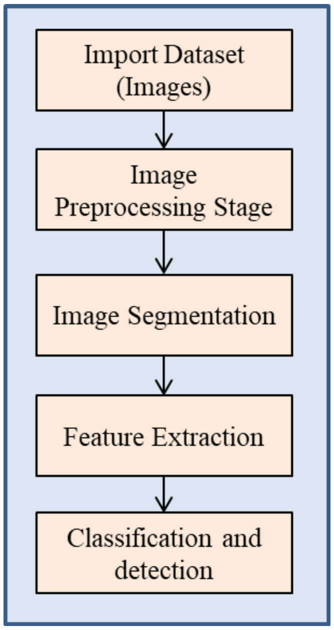

Figure 1.1: The general methodology diagram of plant disease detection [4]

1.2 Related Work

In recent years, different schemes have been proposed in automatic recognition of plant diseases.

A new model developed by **Geetharamani and Pandian [8**] **(2019),** has been proposed to identify a disease of plant leaves that depends on a deeply controversial neural network. The proposed model was constructed using an open data set containing 39 unique classes and original images of plant leaves. Six types of information enhancement techniques were used: image flipping, gamma management, infusion, magnification, rotation, and principal component analysis (PCA). Considering the results obtained, it was found that the application of information increases can extend the exposure of the model. In any case, the proposed model achieved good results with 96.4 accuracy. Finally, the proposed model did not cover the definition of plant leaf disease.

In the prior work of Anagnostis et al. [9] (2020), a model of a robust convolutional neural network (CNN) was developed. The proposed model can group images of leaves depending on whether they are contaminated by anthracnose and then determine whether a tree is contaminated. In addition, many images were used in grayscale and RGB modes, Fourier made a quick transition to highlight highlights, and chose CNN's design according to their display. In any case, the proposed model achieved excellent results with an accuracy of 98.7.

In the previous work of Fujita and others. [10] (2016), the creators have developed a strategy to separate images under extreme and difficult conditions, in which cucumber leaves have seven kinds of disease in healthy. To obtain this effect, four-layer image preparation and CNN method were used, which showed an average accuracy of 82.3% with a quadruple cross-validation methodology.

Yang et al. [11] (2017), a new strategy for identifying rice disease based on a convolutional neural network (CNN) procedure. Using a data set of 500 regular images of infected, and healthy rice leaves and stems captured from a rice test field, CNN is ready to identify 10 underlying rice diseases. As shown by the results obtained, it should be noted that the proposed framework achieved satisfactory results with an accuracy of 95.48%.

Jayme Garcia A. B. et al. [12] (2016), present an automatic identification of a plant disease system based on a convolutional neural network. Also, it is utilizing the capability of the digital image-based algorithm. This algorithm is proposed to cope with various diseases and it is easy to retrain it to include new diseases. Its histogram-based structure enables it to be reasonably robust under different circumstances to capture images. The obtained results show that there is still a place for improvement. Many factors such as a considerable number of present disorders, heterogeneity of symptoms associated with the same disease, and symptom similarities between different disorders may need the adoption of hybrid techniques combining expert systems, image processing, and other information gathering techniques that may be the best hope to beat at least some of the disadvantages present in practice.

Balakrishna and Rao [13] (2019), proposed two methods for establishing and constructing a firm and unwanted tomato leaf. In the early stages, the tomato leaf is healthy or unwanted by the KNN method. Then, in the next step, the tomato leaf is unfortunately characterized by the PNN and KNN approach. Highlights similar to GLCM, Gabor, and shadow are used to sort objects. The experiment targets the authors 'database, which consists of 600 voices and unfortunate characters. According to the obtained results, it is observed that the PNN improve the accuracy to 91.88%, and these percentage is not good enough and there is a need to combine the PNN with other algorithm to give better results.

Shanwen and others. [14] (2019), created a three-channel convolutional neural networks model (CNN) that closely combined three-tone components to diagnose plant leaf disease. In this model, each channel protects one of the three shaded parts of the RGB disease foliage, the shrinking light of each CNN is learned. Thus, it sent to the next convolutional layer and pooling layer, at this point the most prominent parts of fully connected common layer is combined for deep confirmation of disease. Finally, the softmax layer uses the element to classify the information images into predefined categories. As a result, the proposed model can extract the active ingredient from complex diseased foliage and accurately diagnose plant diseases. The proposed system achieved satisfactory results with accuracy of 94.2%

1.3 Problem Statement

Plant diseases are a major threat to the productivity of crops, which affects food security and reduces the profit of farmers. Identifying the diseases in plants is the key to avoiding losses by proper feeding measures to cure the diseases early and avoiding the reduction in productivity/profit.

In addition, over-the-counter fungicides and insecticides can be relied upon to overcome this problem, but it also has a very negative impact on the weather. Whatever, it is necessary to diagnose plant diseases at an early stage to help farmers play it safe to protect the damaged grass. Later, a Deep Convolutional Neural Network (CNN) was proposed to detect and diagnosis plant disease.

1.4 Aim of Thesis

This research aims to propose a system that has the ability for early plant disease detection based on CNN. The following objectives are proposed to achieve the research aim:

1. To design an Automatic System for Early detection and diagnosis of plant diseases based on CNN.
2. To design two algorithms based on machine learning which are Naive Bayes, and Random Forest. Also, the LDA has been employed for feature extraction.
3. To test and evaluate the performance of the proposed model based on Deep learning with the two techniques which are based on machine learning according to the criteria of accuracy, then compare the obtained results with related work.

1.5 Contribution

Fast and accurate plant disease detection is required to increasing the agricultural productivity in a sustainable way. However, the work has three main contrbutions and they described as follows:

- An early and accurate system for plant disease detection and diagnosis based on Deep CNN has been proposed.
- The best two machine learning algorithms in this field have been designed for plant diseases detection which are Naive Bayes, and Random Forest.
- The LDA algorithm have been employed for feature extraction in the stage of the two machine learning techniques.

1.6 Outlines of Thesis

The rest of this thesis is organized as follows:

Chapter Two: (Theoretical Background); this chapter gives the background and review of some indirect and direct methods. Also, it focuses on the outline of the proposed system and the expected objectives.

Chapter Three: (The Proposed System); this chapter describes the proposed system of deep learning neural network- based recognition of plant disease with their design and implementation.

Chapter Four: (Implementation and Experimental Result); this chapter explains the results that have been gotten from the proposed system.

Chapter Five: (Discussion, Conclusion, and Suggestion for Future works);this chapter covers the conclusions observed from the system, and the applications in which the system can be used for. Finally, future work is recommended for the proposed system for higher disease, and health accuracy is considered.