CENTRAL UNIVERSITY SIERRA LEONE

FACULTY OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF TECHNICAL SCIENES

Topic: Crop Disease Detection Model

Design and Implementation of a Tomato Crop Disease Detection Model

Case study: Tomato Crop

Submitted By

1. **MOHAMED LAMIN KAMARA C/S 1920224**
2. **NUMU BAILOR BARRIE C/S 1920081**
3. **UNISA KAMARA BIT 1920285**

A project document submitted to Central University Sierra Leone in partial fulfillment of the requirement for the award of bachelor's degree in Computer Science and Business Information Technology.

CERTIFICATION

This is to certify that Mohamed Lamin Kamara, Numu Bailor Barrie and Unisa Kamara of Central University has successfully completed the research on the project **Crop Disease Detection Model** under the guidance of Mr. Mohamed Lamin Kamara during the year 2022-2023 in partial fulfillment for the award of bachelor's degree in Computer Science and Business Information technology.

Moreover, these students have consistently met the expectations of the course requirements and have completed all assessments and assignments to a high standard. They have also shown excellent collaboration skills, working effectively with their course mates to achieve common goals.

Mr. Mohamed Lamin Kamara (Supervisor) Date

Mr. Isaac Muckson Sesay (HOD) Date

Mohamed Lamin Kamara (1920224) Date

Numu Bailor Barrie (1920081) Date

Unisa Kamara (1920285) Date

DEDICATION

This work is dedicated to all those who have supported and encouraged us throughout our journey. To our families, who have always been our rock and our foundation, thank you for your unwavering love and support. To our friends, who have been source of joy and laughter, thank you for keeping us grounded and reminding us to enjoy life. To our lectures and mentors, who have inspired us and challenged us to be our best selves, thank you for your guidance and wisdom. And finally, to all those who have believed in us and abilities, thank you for your faith and trust in us. This work is a reflection of all of your love and support, and we are grateful beyond words for each and every one of you.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to all ourselves for the dedication and hard work in completing our project Crop Disease Detection Model. Our success would not have been possible without the outstanding collaboration and efforts of each member.

We would also like to express our gratitude to Mr. Mohamed Lamin Kamara his great guidance and support throughout the completion of our final project. Your directions and assistance have been irreplaceable in helping us achieve success. Throughout the process, we faced many challenges and obstacles that seemed undefeatable. But with your loveliness, we found the strength and perseverance to overcome them and complete the project successfully. Your presence and support have been providing us with the inspiration and motivation to do our best.

We also like to express our deep gratitude and appreciation to our lecturers for all the help and guidance they have provided to us throughout our course of study and final project. Their unwavering dedication, support, and encouragement have been invaluable to us, and we cannot thank you enough for all that you have done. Your passion for teaching and commitment to excellence has inspired us to strive for the highest standards and to push ourselves beyond our limits.

We give special thanks to God/Allah for His guidance and blessings throughout our course of study and final project. Your divine presence and support have been our constant companions, providing us with the strength and inspiration to overcome challenges and achieve our goals. Throughout our academic journey, there were times when we felt lost, confused, and uncertain about our abilities and the path ahead. But through your grace, we found the courage and resilience to persevere and succeed.

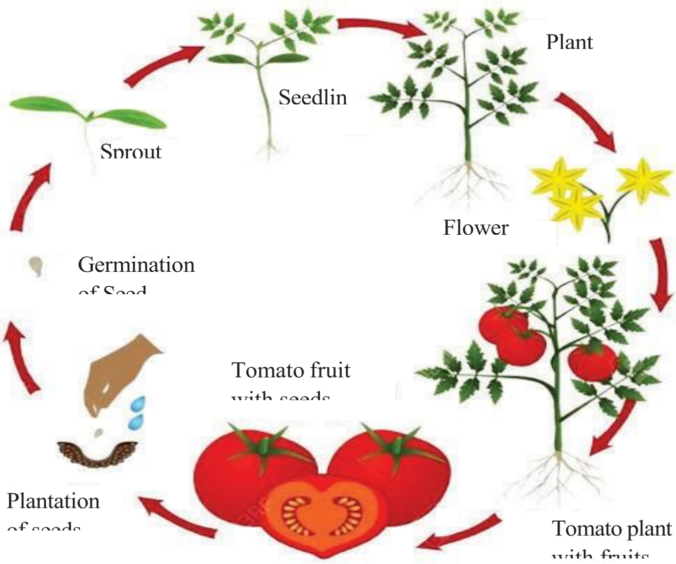
We express our heartfelt gratitude and appreciation to our donors for their generous financial support throughout our course of study and final project. Your kindness and generosity have been instrumental in helping me pursue our academic goals and achieve success. Without your support, we would not have been able to afford the tuition fees, books, and materials necessary for our education.

Table of Contents

CHAPTER ONE

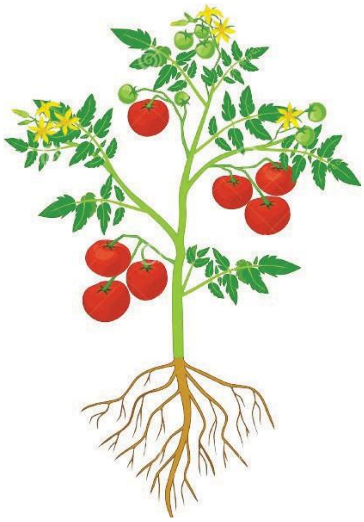
INTRODUCTION

Tomato is an essential popular vegetable crop grown in the world. The biological name for tomato is Solanum Lycopersicum. It is a relatively short-duration crop with a high yield and is economically attractive with an increase in area under cultivation day by day. Sierra Leone is the fifth-largest producer of tomatoes after China. Other significant players in the tomato market are the United States of America, the European Union, and Turkey. Together, these top five tomato producers supply around 70% of the global production. Tomato is an essential economic crop in the world. It is easily affected by many diseases, and this condition severely affects the quality and yield of tomatoes and causes substantial financial losses. Regarding total area and tomato production, Sierra Leone contributes approximately 19.0% and 11.1%, respectively, to the global total. Tomatoes account for 8% and 12% of total vegetable cultivation area and production in Sierra Leone, respectively (Gupta et al., 2020). The tomato plant has a lifespan of around 120 days. The flowering or fruiting stage occurs at approximately 45-50 days of the life of the tomato plant. The life cycle of the tomato plant is shown in Figure 1.1. The tomato life cycle starts from the plantation of seeds, then germination to growing into a sprout, seedling, and further into a plant. Once the plant matures, the flowering stage will be initiated, followed by fruiting. The seeds of the ripe fruits are further used for the next life cycle. The tomato plant in the flowering and fruiting stage is shown in Figure 1.2. The appearance of yellow flowers on the tomato plant indicates that the plant has begun the process of fruit production. The time required for ripe fruit after the flowers open on the tomato plant varies depending on the variety and several environmental factors. The disease in tomato plants occurs more often at the flowering or fruiting stage when the plant is mature. Predicting disease and time management will help reduce the loss in yield and help farmers produce better.



Source: Scopus Database

Figure 1.1. Life cycle of Tomato plant showing different stages.



Source: Scopus Database

Figure 1.2. Tomato plant in flowering and fruiting stage.

Plants are the foundation of all living things because they provide food. We must protect the plant from disease to have good food quality and quantity. Plant species are useful in medicine, agriculture, and industry. A few plants are on the verge of extinction; as a result, it is critical to establish a database for plant protection. Agriculture is one of the most significant factors in the economy of many countries. It is considered a way of life as well as a national benefit. Farming enables people with little or no farming experience to grow plants or crops (AlZu'bi et al., 2019). Sierra Leone is an agricultural country as agriculture is done on a large scale. We can say that we are a farming country. So, the focus of research is on how to increase the quality and amount of crops, as they are the most vital agricultural products, and reduce the loss due to diseases that humans can control. Because plant diseases can cause significant reductions in yield quality and quantity, identifying the disease is critical. If the crop is severely affected by the disease, there may be a substantial loss in output, putting the farmer out of pocket and threatening his livelihood.

Farmers should take preventive measures to protect their farms from diseases that can be avoided if the cause of the disease is known ahead of time. The priority is identifying the plant species, followed by the diseases affecting them. Plant disease is natural, so detecting plant disease is critical in agriculture. It is vital to take proper precautions to avoid severe effects on plants, which could affect product quality, quantity, or productivity. The rapid and precise diagnosis of disease severity will aid in the reduction of yield losses. Plant disease severity is an important parameter that can be used to forecast yield. Preventive measures must be taken to avoid significant effects on the plants; that impact product quality, quantity, or productivity.

1.1 PLANT DISEASE CATEGORY

Crop diseases are classified into two types: airborne and soil-borne. Fungal diseases are common in the airborne type. The affected plant's symptoms are visible in specific parts such as the leaves, stems, and fruit. The effect of soil-borne diseases is most visible on the plant's roots (Zhang et al., 2017). The plant is affected by a variety of viral and fungal diseases. Weather conditions and seasonal changes cause variations in temperature, humidity, wind speed, and so on. These changes have an impact on plants that are susceptible to certain diseases.

1.1.1 TOMATO PLANT DISEASE

Viral diseases: Plant diseases caused by infection are the most difficult to identify and diagnose; additionally, these symptoms are misinterpreted as signs of nutritional deficiency or injury because there is no preconceived indicator that can be constantly monitored. Whiteflies frequently transmit virus diseases, leafhoppers, aphids, and cucumber-crawling insects (Orchi et al., 2021).

Fungal diseases: Foliar diseases, such as downy mildew, anthracnose, and powdery mildew, are caused by fungi. It first appears on old lower leaves with gray-green spots or that have been soaked in water. As the parasite ages, these spots darken, and fungus grows on them (Orchi et al., 2021).

Bacterial diseases: Pathogens cause severe diseases in vegetables. They do not enter the vegetation directly but through crop injuries or openings. Crop injuries are caused by various pathogens, insects, and agricultural implements used during tasks such as picking and pruning (Orchi et al., 2021).

Table 1.1 Favorable conditions for tomato plant disease

Ref	**Disease**	**Pathogen**	**Viral/ fungal /bacterial**	**Environment**
Abrahamian et al. (2021)	Bacterial spot	Xanthomonas Campestris pv. Vesicatoria	Bacterial	High moisture, high relative humidity (≥80%), and warm temperatures (23-32°C).
Gupta et al. (2020)	Early Blight	Alternaria Solani	Fungal	Warm, humid (24-29°C) environmental conditions
Small et al. (2015)	Late Blight	Phytophthora Infestans	Fungal	the relative humidity is < 90% and temperature 18- 22°C
Lee (2020)	Leaf Mold	Pseudocercospora Fuligena	Fungal	Relative humidity greater than 85%. The Optimal temperature is between 24- 26°C
Mamode Ally	Target spot	Corynespora	Fungal	temperatures of 28-

et al. (2021)		cassiicola		32°C with high humidity
Ávila et al. (2020)	Septoria Leaf Spot	Septoria Lycopersici	Fungal	Extended periods of leaf wetness, high humidity, and warm temperatures
Hossain et al. (2010)	Tomato yellow leaf curl virus	Tomato yellow leaf curl virus	Viral	temperatures 19 - 29°C and relative humidity (73.-88%) with whitefly population
Imran et al. et al. (2013)	Tomato Mosaic Virus	ToMV	Viral	temperatures 20 - 31°C and relative humidity 55 to 70%

The favorable weather conditions for the eight tomato plant leaves diseases are shown in Table 1.1. Tomato plant leaves diseases are classified as viral, fungal, or bacterial. Favorable environmental conditions such as temperature and relative humidity are mentioned for each of the eight tomato plant leaf diseases. Maximum and minimum temperature, dew point, relative humidity, and rainfall are responsible for disease occurrence in tomato plants. Temperatures in the 22-38°C range and relative humidity levels in the 55-90% range are favorable for fungal diseases in tomato plants (Kamei et al., 2018; Chothani et al., 2017; Shternshis et al., 2021). The bacterial disease in tomato plants thrives in temperatures ranging from 24-32°C and relative humidity levels of more than 80% (Araújo et al., 2010). The mosaic virus in tomato plants occurs when the temperature is between 21 and 31°C with a relative humidity of 55 to 70 %. The yellow leaf curl virus occurs when the whitefly population, along with a temperature between 19 and 29°C with a relative humidity of 73 to 88% (Hossain et al., 2010).

1.2 IMPACTS DUE TO THE SEVERITY OF DISEASE

Plant disease severity is a strong constraint for assessing disease stages and can thus be used to predict the produce. Recognizing plant disease using camera devices in mobile phone images proves a significant task. In this regard, we look for a reliable, fast, automatic, less costly and accurate method of plant disease detection to overcome this situation (Singh and Misra, 2017). When farmers discover a plant disease, they typically use only chemical fertilizers to prevent further disease growth. This may have harmful effects on the plant and the person contacting the plant. Sometimes, it helps solve a problem with essential products, like plucking and burning a diseased leaf or organic fertilizers. All this depends on how much the leaves of a crop have been affected by a disease.

1.3 TECHNIQUES USED FOR CLASSIFICATION OF PLANT DISEASE

The most basic or traditional method is to inspect the plant with your own eyes. This process requires continuous monitoring of a large farm area and must be carried out by experts. (Arivazhagan et al., 2013; Al Bashish et al., 2010; Kaur et al., 2019; Rehman et al., 2020). This is a lengthy and costly procedure. The quicker and more effective detection of disease severity will aid in the preventive and control measures and reduce yield. Recognizing plant disease using images captured by devices like mobile phone cameras or digital cameras is proving difficult. Expert disease diagnosis may not be sufficient to save or reduce diseases due to the large cultivating area. They must inspect and treat the plants accordingly (Ferentinos, 2018).

Image processing in the mechanism for disease detection is shown to be used more successfully. The current inclination to use different Machine Learning (ML) algorithms to classify plant diseases has demonstrated promising findings in a few specific conditions and crops (Rangarajan et al., 2018). A fast and accurate method will be developed based on computer image processing for plant disease grading. In agricultural engineering research, computer-based image processing technology was commonly used. The basic ML techniques were Artificial Neural Networks (ANN), Decision Trees, K-means, k neighbors, and Support Vector Machine (SVM). Developing Deep Learning (DL) techniques has demonstrated suggestively improved results than the shallow learning algorithm (Chen et al., 2018). LeCun et al. (1998) introduced DL model techniques in the field of classification and detection using Convolutional Neural Networks (CNN) as the primary DL tool. CNN is a vivid model that facilitates the classification of application fields that have large amounts of data. The DL network is used to detect disease on plant leaves, allowing faster and more accurate results. Botanists now benefit from scientific and technological advances with computer vision approaches in plant identification. In the literature, several perspectives on plant classification have been proposed. Recent advances in DL, particularly in CNNs, have resulted in significant breakthroughs in various applications, including classifying plant diseases (Aquil and Ishak, 2021). Robust and cost-efficient schedules for decision-making in multimode projects were discussed by (Schmidt and Hazır, 2019). The investigation of twenty Artificial intelligence techniques was discussed by (Elmousalami, 2020) for decision-making.

The classification of the plant is critical to achieving plant protection (Wu et al., 2007). Plant leaves, unlike flowers, are readily available for a limited time. As a result, the leaves are a fine choice for automatic plant classification. The leaves play an important role in understanding plant genetic relationships and development. How- ever, plant identification is difficult even for botanists because of the large number of species (Tiwari, 2020; Yang et al., 2020). Botanists classified specific plant species using leaf recognition technology. Plants differ in many ways, including texture, shape, color, and size; they are distinct. (Priya et al., 2012). Because of their high classification accuracy, other Computer-Aided Detection (CAD) methods have been used for leaf-based plant recognition in recent years (Bodhwani et al., 2019; Du et al., 2013)

1.4 INTERDISCIPLINARY APPROACH

Lee et al. (2017) suggested an integrative approach to plant classification that combines horticultural data and species concepts with computer remedies. Botanists will now use vision-based approaches to help them identify plants, thanks to recent advances in science and technology. Computer vision researchers used leaves as a comparison tool to classify plants (Kumar et al., 2012). ML can be used to solve the classification problem by implementing a new quick solution that brings together specialists, farmers, evaluators, and practitioners into a single chorus (Chen et al., 2020).

In recent years, many image processing and ML techniques have been used for classification. DL methods have demonstrated improved classification performance (Rangarajan et al., 2018). The current practices are more powerful due to the computer vision and ML computational systems, which can identify and diagnose plant diseases (Mohanty et al., 2016; Yang and Guo, 2017). DL networks are remotely operated methods for classifying plant leaves. A method automatically identifies plant leaf disease symptoms is beneficial because it reduces manual labor and saves time (Singh and Misra, 2017; Singh et al., 2019). When farmers detect a plant disease, they use chemical fertilizers to protect the crop from further disease growth. The usage could harm the yield and anyone who comes into contact with the crop. Simple methods such as plucking and scorching the diseased leaf or using organic fertilizers can often help solve the problem. The extent to which the disease has affected the crop's leaves determines this.

1.5 PARAMETERS IMPACTING PLANT DISEASE

The fast and effective diagnosis of disease severity will aid in reducing yield losses. Plant disease severity is a powerful constraint for measuring disease levels and can thus be used to forecast production. Plant disease prediction methods can help farmers prevent disease and warn them when their crops are at risk of infection.

The process of forecasting plant diseases is crucial. The step will help avoid further disease growth in plants, and the plant product will benefit the growing population's good health and well-being. The plan accounts for the possibility of plant infection caused by environmental factors that promote germination (Shivling et al., 2017). Temperature, relative humidity, absolute humidity, and rainfall create a favorable environment for plant disease. Abiotic factors significantly impact pest fluctuation (Deb and Bharpoda, 2017).

In this case, only plants will become infected with a disease if the environmental conditions favor breeding the pathogen and the host plant (Waqar, 2018; Berg et al., 2021; Baron and Rigobelo, 2022; Severns and Guzman-Martinez, 2021). A low-level disease in most plants is common, but sporadic epidemics can unacceptably decrease cultivation quality or yield. In some cases, genetic resistance in the host plant can prevent plant disease epidemics. Farmers often have to depend on the prudential use of crop guard in terms of chemical substances pre-empt conditions from getting severe enough to have an economic impact on quality and yield in the absence of genetic resistance (Shah et al., 2019).

Climate change is one of humanity's most challenging problems. Weather is crucial in agriculture. Precision crop disease prediction and continuous crop yield monitoring are critical for increasing crop yield in agriculture. Disease progression and weather parameters are well understood and used to create prediction models and decision support systems (Dar et al., 2021; Shah et al., 2019). Because of the varying meteorological conditions, a well-ordered technique is required to facilitate crop cultivation and assist farmers in producing and managing crops. Plant disease losses can be reduced with proper management, resulting in sustainability and food security. This management step may assist future farmers in improving agriculture. Farmers can be given suggestions to help with data mining in crop growth processes. This method is recommended for plants based on climatic factors and quantities (Vaishnavi et al., 2021; Derbile et al., 2022). The disease triangle conditions cause plant disease. When favorable environmental conditions breed the pathogen and the host plant, plants become infected with the disease. Plants are more resistant to disease in the remaining cases.

Weather factors such as temperature, rainfall, leaf wetness duration, and relative humidity drive plant disease forecasting. To identify infection conditions and disease cycle, hourly or daily data is required by the weather-driven models (Kim et al., 2020). Weather forecasting is a critical area of study in everyday life. Weather for the future is essential to forecast because agriculture and many industries rely heavily on weather conditions. Weather forecasting is required for future planning in agriculture and industry, as well as other fields such as defense, mountaineering, shipping, aerospace navigation, and so on (Fathi et al., 2021; Ganguli and Coulibaly, 2019; Hossain et al., 2018; Mehrkanoon, 2019). Weather forecasting is the use of science and technology to forecast the state of the atmosphere for a specific time and location in the future (Bushara et al., 2013). The variables defining weather conditions like temperature (maximum or minimum), relative humidity, rainfall, etc., vary continuously with time. For developing a forecasting model, the formation of a time series for each parameter can be used statistically (Torres et al., 2021) various approaches have been developed for weather forecasting and statistical analysis in the past decades. Regression models are still widely used approaches for predicting future events or values in these models Wang and Ma (2011); Zhang et al. (2019).

Climate change is essential because it can be beneficial and detrimental to crop production. After all, both the host and the pathogen can coexist and be the source of disease in crops. As a result, disease infection is most likely in favorable environmental conditions and between the host and the pathogen. The decrease in yield has an additional impact on seed quality and grain contamination. Appropriate care must be taken to avoid significant plant effects impacting product quality, quantity, and productivity. Rapid and accurate identification of disease severity will aid in the reduction of yield losses. Misinterpretation of climatic conditions can have an impact on disease risk. Skelsey and Newton (2015) developed data showing the spatial coverage of crops with probabilistic climate change in the future wheat disease prediction. The epidemics cause losses due to widespread use of reduced and minimum tillage, reduced crop diversity, increased acreage of host crops, and wet, humid weather conditions during anthesis and early grain fill stages (Andersen et al., 2015). Disease management programs benefit from multiple regression model equations based on temperature and vapor pressure in conjunction with epidemic climatic conditions (Manstretta and Rossi, 2015). The mean aggressiveness in the pattern of the epidemic climatic conditions was revealed using Amplified Fragment Length Polymorphism (AFLP) markers, demonstrating the pathogens' high genetic diversity. This suggests that the disease can spread from one season to the next (Laloi et al., 2016).

The time-series weather data consists of various parameters viz temperature, humidity, dew point, rainfall, and wind speed in an hourly, daily, and weekly for- mat. Numerous profound learning architectures have been designed to diversify data series in different fields (Lim & Zohren, 2021). The various models are linear regression, Support Vector Regression (SVR), Random Forest Regression (RFR), Auto-Regressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA), Prophet, CNN, Recurrent Neural Networks (RNN), Long Short-term Memory (LSTM) to name a few that can be applied to a time-series data. In predicting plant disease, time-series weather data analysis needs to be done for weather parameters like temperature and relative humidity that cause the favorable environmental condition for the disease triangle (Xiao et al., 2019). The hybrid model might give a better forecast result in the weather prediction that can predict plant disease occurrence in advance. The hybrid model can use post-processing techniques like model output statistics to improve the prediction handling errors with optimization methods (Jin et al., 2020; Khan et al., 2020).

1.6 MOTIVATION

It is said that the backbone of the Sierra Leone economy is agriculture. The agriculture sector's contribution to the national Gross Domestic Product (GDP) was 20.2 in 2020-21 (PIB Sierra Leone, 2021). Sierra Leone agriculture must modernize and use automation in agricultural activities to achieve a growth rate. Plant disease classification and prediction with computer vision and artificial intelligence are currently carried out and are at the research stage. Diseases harm the leaves of the plants and need to be detected. These harmful effects change the physical appearance of the leaf so that the cause of the harm can be seen from the images. An artificial Intelligence system offers the potential to automate manual grading practices and thus replace tedious human inspection tasks. It has proven successful for the objective of classification and prediction purposes.

The motivation of this work is to detect tomato plant diseases using the deep learning method and further predict the occurrence of diseases in tomato plants that will affect the plant. This will help reduce the loss due to disease if proper management is taken. Recent trends in machine learning and deep learning have achieved high accuracy in classification tasks. The application of deep learning methods started developing in the interdisciplinary area.

1.7 RESEARCH GAPS

Data is required in the proposed research on tomato plant disease detection. However, the required data is available in the Plant Village dataset, which can be used for detecting and classifying diseases in the tomato plant. This database is foreign data and does not suit the local region data of tomato plants due to different weather and soil conditions. The data used in the research is the segmented leaves of the tomato plant (Rangarajan et al., 2018). The segmented images are not readily available in real-time, so the work must be done on the raw images.

1.8 SCOPE AND LIMITATIONS

SCOPE: The scope of this research proposal is to use the DL algorithms to detect and classify a tomato crop's healthy or diseased leaf with less time and accuracy. Further, it will also classify the disease from which the tomato plant is affected.

LIMITATIONS:

1. As far as limitations are concerned, if a large data set is used for analysis, then there is a need for a high computing facility for this purpose.
2. Time

1.9 Statement of the Problem

After reviewing the literature for plant disease classification, primary artificial neural networks like the feed-forward back propagation technique, k –means clustering, SVM, etc., were used. The use of the DL algorithm will significantly boost the results for plant disease detection and classification. The work proposes to detect and classify plant leaf diseases using the DL algorithm. The detection and classification of disease in the tomato crop are presented here. There are various airborne diseases like bacterial spot, early blight, late blight, leaf mold, septoria leaf spot, target spot, tomato mosaic virus, and tomato yellow leaf curl virus that affect the tomato leaves crop.

Detection and Classification of the tomato Plant Disease using Deep Learning Network

1.10 Objectives of the study

- To identify the tomato plant leaf from the data feed to the network with a mixed variety of leaves of different plants.
- To detect and classify disease in the tomato plant.
- To test and validate the algorithm.

CHAPTER TWO

LITERATURE REVIEW

This chapter comprehensively reviews existing and significant research from an earlier decade. The chapter focuses on image-processing classification techniques and their usage in various domains. Classifying a plant plays a vital role in implementing maintenance methods for that plant. Once the plant species is identified, the focus can then be shifted to the diseases in the plant and their prevention. The factors influencing plant disease are crucial to understanding practical steps during those conditions that can prevent the growth of the disease. This will have a positive impact on the produce from the plants. The study is broadly categorized into five sections.

Section 2.1 describes the study of bibliometric analysis for plant disease classification and prediction. The investigation is done with the data fetched from Scopus. This quantification will be helpful in understanding the current state of undertaken research in terms of the query posed to the publication databases through keywords related to an undertaken research problem statement.

Section 2.2 describes the classification using Image processing using different methods—overfitting problems and techniques to avoid it to improve the model's performance in classification. The modified models in the classification purpose enhance the performance and how classification gives the solution in various domains.

Section 2.3 discusses the classification of plant species. Plants are useful for medicine, food, and industrial applications. As a result, it is necessary to develop a method to identify them correctly.

Section 2.4 discusses classification of disease in plant leaves. The section focuses on the different techniques used in the classification of plant diseases. Machine learning and Deep Learning approaches and the related work in the classification of various plant diseases are showcased here.

Section 2.5 discusses the prediction of plant disease due to climatic conditions. The factors influencing the cause of disease in plants and the methods used by researchers are discussed here.

2.1 Current State of the undertaken research through Bibliometric Analysis

Bibliometric methods are used to investigate the impact of a specific field, the result of involved researchers, and the impression of a set of papers with respect to a particular area of research (Chaudhari et al., 2019). The quantification of published research in popular publication databases such as Scopus is known as bibliometric analysis. The current state of the undertaken study through Bibliometric Analysis through the Scopus database accessed on 9th January 2021 is done in this section. The search for the Plant Disease Classification (PDC) and Plant Disease Prediction (PDP) is made through the Scopus keyword search. The query keywords are shown in Table 2.1. The keywords that were used for the study of PDC and PDP are shown here.

Table 2.1 Query keywords for PDC and PDP.

Query	Keywords
PDC	"Plant disease" and "Classification" and "Deep Learning" or "Machine Learning"
PDP	"plant disease" and "disease prediction" and "regression model" or "viral dis- ease" or "bacterial disease" or "fungal disease" OR "plant disease" and "dis- ease prediction" or "regression model" and "weather condition" or temperature or humidity or rainfall

The correct choice of keywords will fetch the most exact search of the literature. Based on the keyword search, the number of publications in the area of PDC and PDP is shown in Figure 2.1. It is noticed that the trend of publication has increased in PDC and PDP since 2018, and there are more publications in the field of PDP.

The investigation of the query search through Scopus in PDC and PDP is shown in Table 2.2 to Table 2.5. These tables give statistics about details obtained from Scopus publications in terms of their data and metadata. Table 2.2 shows the overview of the papers published under the query PDC and PDP. There are around 48% more publications in PDP. The maximum of the research papers is written in English and are majorly published in Journals. There is also a good amount of rising in citing the literature in this field since

2018.

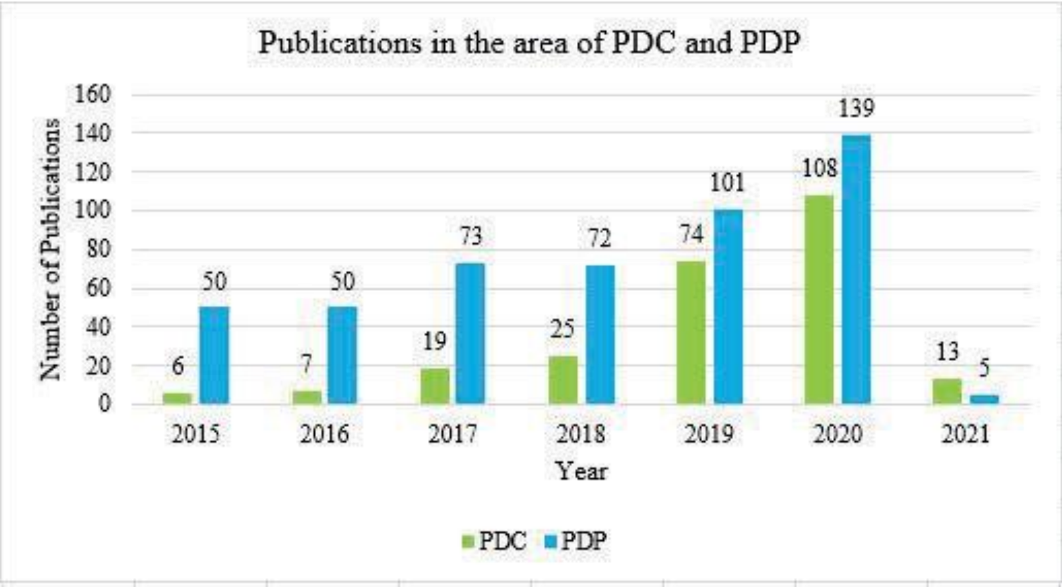


Figure 2.1. Publication trends in PDC and PDP.

Table 2.2 Overview about Scopus Publications

Sr. No	Publications Overview	PDC	PDP
1	Range of years	2015-2021	2015-2021
2	No of papers	252	490
3	Papers published in Article	115	408
4	Papers published in English	248	481
5	Citation increased from the year	2018 onwards	2018 onwards

Source: Scopus Database

The metadata of the publications for PDC and PDP are listed in Table 2.3. The top keywords used by the researcher are shown here. Sierra Leone, China, and the United States of America are the top producers of tomato crops. The influence can be seen through research as most of the authors belong to these countries. The aiguille affiliations and the subject areas of the PDC and PDP are detailed in the table.

The influential funding agencies providing research funds in PDC and PDP are listed here in Table 2.4. National Natural Science Foundation of China (NNSFC) is the apex funder in the research carried out under PDC and PDP and has funded 16 and 19 research works, respectively.

Table 2.3 Metadata for Scopus Publications

Sr. No	**Publications Metadata**	**PDC**	**PDP**
1	Top keywords	Deep Learning, Plant Disease, Machine Learning, Image Processing	Plant Disease, Climate Change, Microbiology, Regression Analysis
2	Top publishing countries	India, China, US	US, India, China
3	Influential authors	Chen, J. and Khan M.A	Rossi, V., and Makowski, D.
4	Aiguille affiliations	Vellore Institute of Technology, Vellore; Kalinga Institute of Industrial Technology, Bhubaneswar, Beijing Forestry University	University of Florida, Cornell University, USDA Agricultural Research Service, Washington DC
5	Subject area	Computer Science, Engineering, Agricultural and Biological Sciences	Agricultural and Biological Sciences, Environmental Science, Biochemistry, Genetics, and

			Molecular Biology

Table 2.4 Influential Funding details from Scopus Publications

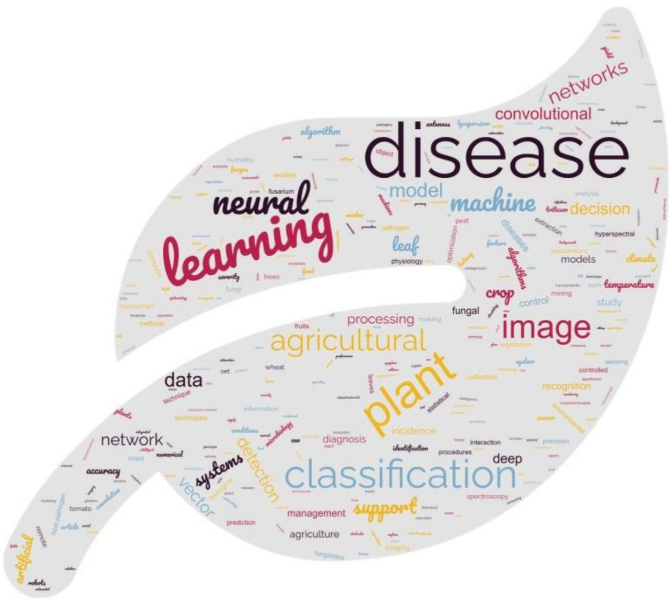
Sr. No	**Funding dtails**	**PDC**	**PDP**
1	Funding agencies	NNSFC, Fundamental Research Funds for the Central Universities, Empresa Brasileira de Pesquisa Agropecuária	NNSFC, National Institute of Food and Agriculture, U.S. Department of Agriculture
2	Research funded by NNSFC	16	19

Source: Scopus Database

Table 2.5 Leading Journals and Top cited articles from Scopus Publications

Sr. No	**Publication statistics**	**PDC**	**PDP**

1	Leading Journals	Computers And Electronics In Agriculture, Advances In Intelligent Systems And Computing, Communications In Computer And Information Science, Inter- national Journal Of Innovative Technology And Exploring Engineering	Plant Disease, Phytopathology, European Journal Of Plant Pathology, Remote Sensing
2	Top cited publications	Deep Neural Networks Based Recognition of Plant Diseases by Leaf Image Classification (Slado-jevic et al., 2016) Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning (Wang et al., 2017) • A deep learning-based approach for banana leaf dis- eases classification (Amara et al., 2017)	• Translating High-Throughput Phenotyping into Genetic Gain (Araus et al., 2018) • Estimation of winter wheat above-ground biomass using unmanned aerial vehicle-based snapshot hyperspectral sensor and crop height improved models (Yue et al., 2017) • Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: A review (Pu et al., 2015)



Source: Scopus Database

The leading journals focusing on the publication of PDC and PDP are Computers and Electronics in Agriculture and Plant Disease, respectively. Table 2.5 shows the details of the leading journals and the top-cited research papers. The potential papers showcasing benchmark research that other fellow researchers follows for their work in PDC and PDP are amongst the papers (Sladojevic et al., 2016; Wang et al., 2017; Amara et al., 2017) and (Araus et al., 2018; Yue et al., 2017; Pu et al., 2015) respectively.

The keywords used in the research are influential, and researchers search their related relevant literature with these keywords. The word cloud image of the most important keywords used in PDC and PDP is shown in Figure 2.2.

PDC and PDP are essential because it helps with crop disease management and leads to a healthy, high-quality, and plentiful yield. With global food demand in- creasing, this benefits society and the economy. This bibliometric survey helps the researchers have a focused set of literature to review, identify research gaps in PDC and PDP, and get closer to the niche area.

2.2 Classification using Image processing

Image processing has indeed been done for classification purposes in a wide range of domains. In recent years, researchers have been drawn to developmental neural networks since of their potential to provide better visual classification accuracy.

They use a neural network and computation to solve any problem. The methods conferred in the classification process are the neural network, supervised feedforward backpropagation, unsupervised self-organizing map, Probabilistic Neural Networks, Fuzzy logic, Genetic Algorithm, SVM, Principal Component Analysis, and k-Nearest Neighbor classifier. Due to the uncertainties of noise in low-level image processing and uncertainties in understanding during high-level image processing, fuzzy processing is desirable.

Krizhevsky et al. (2012) propose a DL model in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) named AlexNet network that can classify 1,000 classes successfully. In the 2012 Large Scaled Challenge for Visual Recognition, the authors achieved 10.9% more classification accuracy than the second-best entry in Im- ageNet. Image processing advancements provided various preprocessing techniques for image extraction. The LeNet-5 model was the original CNNs with a standard structure — stacked conversion layers (Szegedy et al., 2015), so it was named LeNet- like when the GoogLeNet model was developed at ILSVRC in 2014.

There are various CNNs models like AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2015), ResNet50, ResNet18, ResNet101 (He et al., 2016), VGG 16, VGG 19 (Simonyan and Zisserman, 2014), DenseNet (Huang et al., 2017), SqueezeNet Iandola et al. (2016), etc. for classification. The difference is the network's layer shallowness and nonlinear functions. DenseNet has feedforward layers, whereas ResNet has many residual blocks. SqueezeNet's squeeze layer is located inside the Fire module, containing several. The assembly consists of four major layers, viz. convolution layer, pooling layer, fully connected layer, and the output layer.

2.3 Overfitting problem and techniques to avoid it

Overfitting is a common pitfall in deep learning algorithms where a model attempts to fit all the training data and memorizes the data patterns, noise, and random fluctuations. Overfitting may occur as a result of the model's training time or architectural complexity. If the model trains on the training data for an inordinate amount of time or becomes overly complex, it learns the noise or irrelevant information within the dataset. The DL model can over fit if the statistical analysis shows random noise or error. The DL has a wide range of parameters to be trained, with high memory costs, time-consuming training and prediction, and the risk of overfitting (Zhao et al., 2018; Qian et al., 2020). This overfitting complication can be fixed at the training stage of the model, which increases the invasion of complex conditions, where the images are deformed slightly in the research phase. Remotely sensed images were classified with the pre-trained AlexNet (Han et al., 2017).

The techniques that are used to avoid overfitting are data augmentation, feature selection, adding noise to the data, early stopping the mode; when training, adding dropout layers in the model, etc. Position augmentation, color augmentation, and Generative Adversarial Networks (GAN) are used to augment the data. Data augmentation influences the average precision of the class (Fuentes et al., 2017, 2018; Mohammedhasan and Uğuz, 2020). The review done by (Hasan et al., 2020) shows most of the current training methods, dataset training, data augmentation, feature extraction methods, recognition of crops, enumeration of the plants, plant disease detection techniques, and performance of classifiers. Batch normalization or shortcut connections is among possible reasons for the strength of ResNet50 (Arechiga and Michaels, 2018). Transfer learning can use the pre-trained network and modify some parts of it as per the need for work (Bulla et al., 2020; Raghu et al., 2020; Syarief and Setiawan, 2020). The model by (Hu et al., 2020) proposed a less complex algorithm to achieve a precision of 94.26%. Başaran et al. (2020) obtained 75.85% precision with the augmented dataset for the DL model. The problem of overfitting can be solved with the augmentation of data.

2.4 Classification solution in various domains

The classification task is done in various domains to identify or classify the given data and give a reliable solution. One of the most critical problems in image recognition is medical image classification, which aims to categorize medical images to aid doctors in disease diagnosis or further research. Overall, the classification of medical images can be divided into two steps. The first step is to extract compelling image features. The features are then used to create models that classify the image dataset in the second step. Traditionally, doctors used their professional experience to extract features from medical images to classify them into different classes, which was a complex, tedious, and time-consuming task. This method is prone to producing inconsistencies or non-repeatable results. Medical image classification application research has yielded significant results compared to previous research.

Mesut To˘gac¸ar , Zafer C¨omert (2020) used AlexNet and VGG16 to classify brain tumors. Lenz et al. (2020) used the deep learning models AlexNet, GoogLeNet, and the inception model to determine adhesion strength. It was discovered that the classification of the implemented models indicates 85 to 90 % of accuracy compared to the assessment done by a human being. Rehman et al. (2020) used the VGG16, AlexNet, and GoogLeNet models with transfer learning to classify brain tumors and achieved an accuracy of 98.69% with the VGG16 model. Lu et al. (2019) used two deep learning models, VGG 16 and MobileNet, to classify Alzheimer's disease using patient MRI images. Khamparia et al. (2020) classified cervical cells in the case of cervical cancer using different CNN networks such as InceptionV3, VGG19, SqueezeNet, and ResNet50. Mehra et al. (2018) used pre-trained VGG 16, VGG 19, and ResNet50 to detect breast cancer using histopathological images.

The fine-tuned CNN outperforms the traditional handcrafted feature extraction methods in classifying the three types of gastric lesions, ResNet50 achieves 96% ac- curacy (Liu et al., 2020). For the breast cancer classification, (Salama and Aly, 2021) with the U-Net model and InceptionV3 with data augmentation achieved an accuracy of 98.87%. In the lung cancer detection, (Da Nóbrega et al., 2018) com- pared deep learning models of VGG16, VGG19, MobileNet, Xception, InceptionV3, ResNet50, Inception-ResNet-V2, DenseNet169, DenseNet201, NASNetMobile and NASNetLarge and discovered that ResNet50 outperformed with 88.41% accuracy. Hu et al. (2019) compares the classification of lumbar images from four different datasets with ResNet18. The authors compared the ResNet18 results to pre-trained networks and discovered that accuracy increases with transfer learning in pre-trained networks.

2.5 Classification of plant leaves

The main objective of this work is first to identify the plant leaf and further focus on the classification of disease in the tomato plant. Plant species are useful in medicine, agriculture, and industry. Some plants are on the verge of extinction. As a result, it is critical to establish a database for plant protection. The most basic or traditional technique is a naked-eye examination of the plant. This procedure entails experts constantly monitoring a wide range of farm areas (Rehman et al., 2020; Kaur et al., 2019). This is a lengthy and costly procedure. The classification of the plant is a critical step to achieving plant protection (Wu et al., 2007). There are different datasets of plant leaves, viz. Flavia, AyurLeaf, PlantVillage etc. Researchers have worked towards classifying plant leaves first to find the type of plant and further about the possible diseases occurring in them.

The different types of morphological features were selected for the plant leaf classification by (Ghaiwat and Arora, 2014). The classification of the plant leaf data with a CNN model achieved an accuracy of 86.2% (Dyrmann et al., 2016). Barré et al. (2017) used LeafSnap, Foliage, and Flavia datasets to classify different classes with their proposed model LeafNet, for plant leaf classification. A total of 184 LeafSnap classes were correctly classified with an accuracy of 86.3%, and 60 Foliage dataset classes were correctly classified with an accuracy of 95.8%. For the Flavia dataset, with 32 classes, achieved a performance accuracy of 97.9 %. For the MalayaKew dataset with 44 classes, a deep CNN model (Lee et al., 2017) with Multilayer Perceptron (MLP) classifier achieved 97.7% accuracy and improved to 98.1% accuracy with the SVM classifier. Haque and Haque (2018) presented work for plant classification that uses geometric features in preprocessing and achieved an accuracy of 90% for classifying ten plant species from the Flavia dataset. Gao et al. (2018) achieved an accuracy of 84.2% in a Life CLEF Plant Identification Task with their proposed 3SN Siamese network that learns from spatial and structural features for the leaf classification task. Zhu et al. (2019) used a two-way attention CNN model to recognize plant families and identify plant classes for the four datasets. The comparative analysis of plant classification work is shown in Table 2.6

Table 2.6 A comparative analysis of the work related to the classification of plants.

Ref No	Objective	Dataset	Number of classes	Model	Accuracy
Dyrmann et al. (2016)	Plant leaf classification	Six different datasets	22	CNN	86.20 %
Mohanty et al. (2016)	Identify 14 crop species	PlantVillage	38	AlexNet GoogLeNet	99.27 % 99.34 %
Barré et al. (2017)	Plant identification system	Leaf Snap	184 60 32	LeafNet LeafNet LeafNet	86.30 % 95.80 % 97.90 %
Lee et al. (2017)	Plant leaf classification	Malayakew	44	Deep CNN (D1) MLP Deep CNN (D1) SVM (linear)	97.70 % 98.10 %
Gao et al. (2018)	Leaf Identifica-tion	Life CLEF 2015	30	3SN	84.20 %

Dileep and Pour- nami (2019)	Medicinal plant classification	AyurLeaf	40	AlexNet Ayurleaf CNN	94.87 % 95.06 %
Duong-Trung et al. (2019)	Medicinal plant classification	Own data	20	MobileNet	98.50 %

Ref No	**Objective**	**Dataset**	**Number of classes**	**Model**	**Accuracy**
Liu et al. (2018)	Classification of 32 different plant leaves	Flavia	32	Ten-layer CNN model	87.92 %
Bodhwani et al. (2019)	Plant Identifi-cation	LeafSnap	180	ResNet	93.09 %
Tiwari (2020)	Plant leaf classification	Dataset col-lected by Silva et al. (2013)	30	DNN CNN	91.17 % 95.58 %
Yang et al. (202	Classification of plant leaf	Own data	15	VGG16 VGG19 Inceptio	91.50 % 92.40

0)				n-ResNetV2	% 89.60 %
Villaruz (2021)	Identification of berry plants	Own data	3	AlexNet	97.80 %

In the guidance of Ayurvedic drugs, the identity and category of medicinal vegetation play a critical role. In addition, it's far vital for farmers, botanists, practitioners, the forest department's offices, and people concerned with inside the preparation of Ayurvedic drugs for a precise class of medicinal vegetation. The classification of medicinal plants by (Dileep and Pournami, 2019) using the AlexNet model achieved an accuracy of 94.87%, and the accuracy of the Ayurleaf CNN model was 95.06%. Duong-Trung et al. (2019) achieved a classification accuracy of 98.5% on the MobileNet model of 20 self-collecting medicinal plant data types. For plant identification using the LeafSnap dataset, the ResNet model achieved a classification accuracy of 93.09% (Bodhwani et al., 2019).

Classification of plant leaves completed by (Tiwari, 2020) for images taken by (Silva et al., 2013) on an Apple iPad

The Deep Neural Network (DNN) model shows an accuracy of 91.17%, while the CNN model shows accuracy has been improved to 95.58%. Classification of plant leaves for the complex background of the photo taken with the mobile phone was completed with (Yang et al., 2020).

2.6 Classification of disease in plant leaves

When the species of the plant is classified, then one can focus on the classification of the disease in the plants. The machine learning and deep learning techniques in the classification of plant disease are discussed in sections 2.4.1 and 2.4.2.

2.7 Machine learning techniques

Machine learning in precision farming uses monitored SVMs, unsupervised networks such as k-means, and self-organizing maps for classification and clustering (Behmann et al., 2015). The CNN model developed by (Sladojevic et al., 2016) distinguishes plant leaves from the environment. It is also used to analyze whether the leaves of a plant are healthy or diseased. The model achieved an accuracy of 91% to 98%, with an average accuracy of 96.3% in individual class tests. Zhu et al. (2017) has developed various machine-learning algorithms for classifying health classes for tobacco mosaic virus disease and effective wavelengths, texture features, and data fusion. Performance is better with data fusion than texture features, reaching an accuracy of 95% with Back Propagation Neural Network (BPNN). Table 2.7 shows a comparative analysis of work related to plant classification using machine learning techniques.

Table 2.7 A comparative analysis of the work related to the classification of plants using machine learning techniques.

Ref No	**Model**	**Advantages**	**Diadvantages**
Rumpf et al. (2010)	SVM is used for the classification of healthy leaves and disease class of sugarbeet	• It identifies dis- eases even before specific sympto ms became visible	Hyperspectral image data is re- quired
Arivazhagan et al. (2013)	Minimum Distance Classifier and SVM for classification of ten species of plants	• The textural features with multiclass SVM gives good accuracy	Less training data is used. The model may fail to predict if there is

			variation in the data.
Ghaiwat and Arora (2014)	The survey of classification techniques of feed-forward backpropagation, unsupervised self-organizing map, Probabilistic Neural Networks, Fuzzy logic, Genetic Algorithm, SVM, Principal Component Analysis, and k-Nearest Neighbor classifier.	Prediction accuracy of SVM is high. • SVM is robust • Its simple geo- Metric interpretation and a sparse solution. • Computational complexity of SVMs is indepen-	SVM requires more training time. It is laborious to realize the weight function of SVM. Classification requires multiple support vectors. Predictions using the k-NN method have time

		dent of input space dimensions • The k-nearest-neighbor method is the simplest algorithms for Prediction.	complexity.
Singh and Misra (2017)	Genetic algorithm with Minimum Distance Classifier and SVM for classification of plants	• Detection of dis-ease at an early stage.	• Hybrid algo-rithms can be used to improve the accuracy.

A machine learning framework with K-high-resolution feature maps can separate the visual symptoms of stress severity, intensity, and further stress quantification. (Ghosal et al., 2018). The authors assessed the stress caused by bacterial or fungal diseases and nutrient deficiency for 25,000 images of plant leaves. Griffel et al. (2018) used SVM to classify potato plants infected with potato virus Y. The multiclass SVM was used in automated disease detection with color space segmentation of citrus plant images (Sharif et al., 2018). The machine learning model performs well when the data is small. For large-size data, deep learning models perform well in the classification task.

2.7.1 Deep learning techniques

LeCun et al. (1998) introduced the deep learning models in the field of classification and detection with the basic DL tool of CNNs. In recent years, the DL models have to a limited extent, been used in agriculture. CNNs is a dynamic model that supports large amounts of data in classification applications. Amara et al. (2017) developed the LeNet CNN model's effectiveness in classifying background resolution, size, and orientation. Wang et al. (2017) compares the VGG16 model with transfer learning to shallow networks with fine-tuning and models created from scratch. The VGG16 model achieves a classification accuracy of 90.4%. A CNN model with a high success rate can be helpful as an early warning system. With an integrated supported system, this approach aids in the support of real-time conditions in cultivation (Ferentinos, 2018)

Rangarajan et al. (2018) detected plant leaf disease with fast and accurate re- sults. Botanists are now benefiting from advances in science and technology with a computer vision approach to plant identification tasks. Several perspectives on plant classification are proposed in the literature. Too et al. (2019) compares VGG 16, Inception V4, ResNet with (layers 50, 101, 152) and DenseNets with (layer 121) and classifies 38 different classes in the Plant Village dataset into healthy classes of disease. DenseNet worked well among other networks with an accuracy of 99.75%. Barbedo (2019) uses data segmentation using deep learning techniques to identify multiple diseases that affect the same plant leaves.

Adaptive model of 3D DeepCNN for hyperspectral images of soybean crops affect- ing economically significant diseases (Nagasubramanian et al., 2019). The two-level neural network proposed for the classification of plant diseases focused on the largest dataset of 79265 leaf images with inconsistent backgrounds and different weather con- ditions. The trained model achieved an accuracy of 93.67% (Arsenovic et al., 2019). In their study, Picon et al. (2019) used a customized deep residual neural network- based algorithm to address the detection of multiple plant diseases under real-time wheat disease detection conditions. With particle swarm optimization, CNN hyper- parameters have been improved to classify the model correctly (Darwish et al., 2020). Karthik et al. (2020) used residual learning to learn essential functions and atten- tion mechanisms as input for detecting the infection in tomato leaves. Deep learning processes are often used to detect and classify plant leaf diseases (Vaishnnave et al., 2020).

Liu et al. (2018) proposed a ten-layer CNN model for plant leaf classification and achieved an accuracy of 87.92% in 32 classes using the Flavia dataset. Li et al. (2019) achieved an accuracy of 95% using the same model in different pieces of training based on the CNN of remote sensing images. A small CNN was proposed by (Bharali et al., 2019) to categorize plants into the healthy or disease category and achieve an accu- racy of 96.6%. Verma et al. (2020) compared AlexNet and ResNet18 in classifying the three pathologies of vines from the PlantVillage database. Performance was mea- sured in terms of accuracy, precision, recall, F1 score, loss of validation, and Receiver Operating Characteristic (ROC) curve. ResNet18 outperformed AlexNet with an accuracy of 86.43%. Aravind and Raja (2020) uses six different deep learning mod- els: AlexNet, VGG16, VGG19, GoogLeNet, ResNet101, and DenseNet201 to use eggplant, hyacinth bean, ladies finger, and lime. The author achieved the highest ac- curacy of 97.3% on GoogLeNet compared to AlexNet, VGG16, VGG19, ResNet101, and DenseNet201. A comparative analysis of deep learning models and literature in the classification of plant diseases is shown in Table 2.8.

Table 2.8 A comparative analysis of the work related to the classification of plant disease using deep learning techniques.

Ref No	**Model**	**Database**	**Objective**	**Challenges/Future scope**
Wang et al. (2017)	VGG16	Plant Village	Apple Black Rot for three stages	Only one class of healthy and dis- ease of apple black rot is consid- ered
Lu et al. (2017)	CNN	Own data	Rice disease	More data is required to improve the accuracy of the model.
Pereira et al. (2019)	AlexNet	Own dataset, Flavia	Identification of grape plant	Fine tunning and more data from different geographical locations can improve the performance of the model

Türkoğlu and Han- bay (2019)	AlexNet, VGG16	Own dataset	Identification of eight vari- ety of plant disease and Pest	More variety of disease and pest can be considered
Liu et al. (2017)	AlexNet, VGG16, GoogLeNet, AlexNet-Inception	Own dataset	Classification of apple dis- ease	The apple leaves which are fallen due to biological growth issues were not included in the study.
Kamal et al. (2019)	MobileNet	Plant Village	Plant dis- eases classification for 55 classes.	Implementation of the model on a large database can be time-consuming, and outperforming the other models can be challenging.
Ayu et al. (2021)	MobileNetv2	Kaggle dataset	Cassava plant dis- ease	Model can be fine-tuned to improve the accuracy of the model
Mohanty et al. (2016)	AlexNet, GoogLeNet	Plant Village	Identify 14 crop species and 26 dis- eases	Image data from a smart- phone can be supplemented with location and time in- formation to improve accuracy even further.

The comparison of models in terms of accuracy and size of the models with existing models are shown in Table 2.9. Jeon and Rhee (2017) achieved an accuracy of 99.60% with GoogLeNet. In their work on plant classification, (Kaya et al., 2019) used PV and Flavia datasets with AlexNet and VGG16 models. Wang and Wang (2019) classified plants with an accuracy of 84.47% with VGG16 and ResNet50 with 92.64%. For the VGG16 and VGG 19 models, the accuracy achieved by of models is 81.3% and 96.25% respectively (Anubha Pearline et al., 2019). The combination of pruning and post-quantization was applied to VGG16, AlexNet, and LeNet model (Fountsop et al., 2020). The pruning step was responsible for reducing the model size. The performance of models is 91.49%, 96.59%, and 95.2%, respectively. The ten-layer CNN model by (Liu et al., 2018) achieved an accuracy of 87.92% with the Flavia dataset and 84.02% with the PV dataset.

2.7.2 Related work in the classification of plant diseases

Researchers have implemented machine learning and deep learning techniques in the classification of diseases in different plants. Once the disease is correctly known, the farmers can take proper management steps to get rid of the disease or avoid the occurrence of the disease by spraying particular pesticides. The section briefs about the work done in the classification of diseases in plants

Cassava plant disease

Cassava plant disease classification was performed using the MobileNetv2 model with (Ayu et al., 2021) and achieved an accuracy of 65.6%. Abayomi-Alli et al. (2021) achieved a cassava plant disease classification accuracy of 99.7% using MobileNet V2 using a data augmentation technique. Detection of cassava plant disease by (Oyewola et al., 2021) attained an accuracy of 96.75% using the Deep Residual Neural Network Model.

Rice plant disease

Shah et al. (2016) collected data for rice plant diseases such as brown spot, bacterial leaf blight, and leaf smut in the village of Shertha, Gujarat, India, and classified them using SVM. The Color characteristics, such as the mean and the standard deviation, have been extracted and used for the SVM classification. CNN’s deep-learning model has been applied to identifying rice disease and recognizing plant leaves by (Jeon and Rhee, 2017; Lu et al., 2017). The accuracy achieved by (Chen et al., 2020) is 84.25% with the enhanced VGGNet with the Inception module (INC-VGGN) model on the PV dataset for the classification of rice plant disease, and the model's performance was improved to 91.83% on their dataset. In the classification of four paddy leaf diseases by (Islam et al., 2021), ResNet-101 model achieved an accuracy of 91.52%. The dataset of paddy leaves consisting of brown spot, leaf blast, leaf blight, leaf smut, and a healthy class was collected from Kaggle and UCI repositories. A Faster R-CNN algorithm was used by (Bari et al., 2021) to diagnose rice plant disease and attained an average accuracy of 98.84% for healthy and three disease classes.

2.8 Prediction of plant disease

Climate plays a dynamic role in the produce from agriculture. Climate change can also be considered an essential aspect in the occurrence of disease in plants that will affect the quality and quantity of food production. The method to detect the disease in plants plays a vital role in reducing the loss due to the disease. The prediction of plagues and diseases has been problem-solved as a time series prediction, and an LSTM approach has been devised to address the problem (Xiao et al., 2019). Kim et al. (2020) devised a seven-day move average regression analysis for rice illness prediction. The anomaly-detection algorithm developed by (Skelsey, 2021) showed the highest forecast accuracy at 97.0% with the Gaussian mixture model and 96.9% for the one-class k- means method. A hybrid model was built by (Bhatia et al., 2020), the technique for predicting powdered mildew diseases in tomato plants with SVM and Logistic Regression (LR); where the exactness of the model was 92.37%. Weather forecast data would make it easier for weather-driven illness models to man- age diseases better.

A Spatio-temporal recurrent neural network was used by (Xu et al., 2018) to predict wheat crop disease severity. The prediction of leaf wetness duration also causes disease in plants. Wang et al. (2019), in their work, collected the greenhouse data of tomato plants from Almeria, Spain, and Beijing, China that consisted of prediction of relative humidity, dew temperature, transpiration, radiation, and a combination of all these factors. A neural network simulator was used for the prediction of pa- rameters. The relative humidity was predicted with 73% and 83% accuracy in Spain and China's greenhouse data, respectively. The overall performance was 82% and 98%, respectively, for Spain and China datasets. For seasonally repeated patterns, statistical regression models like Prophet are used (Taylor and Letham, 2018). In Myintkyina, Malaysia, temperature prediction was made by

(Oo and Sabai, 2020) with the Prophet model, which gave the root mean square error (RMSE) value of 5.7573.

Pusa Ruby variety of tomato plants was considered for the study of the occurrence of early blight in the plants by (Gupta et al., 2020). The disease intensity was evaluated with the temperature and relative humidity values. The performance of the stepwise regression model with an RMSE of 6.129 was achieved by them for the prediction of disease intensity. Dar et al. (2021) have used a stepwise regression model to predict late blight in potato plants for the four varieties of potato. The parameters of temperature, relative humidity, rainfall, and windspeed were considered for disease prediction. Table 2.10 shows the comparative study of a prediction model for disease occurrence.

The performance of existing works in prediction work is shown in Table 2.11 The prediction of financial data (Siami-Namini et al., 2018) used ARIMA and LSTM model. In their work LSTM model outperformed the ARIMA model. Yang et al. (2019) predicted photovoltaic power with LSTM and BLSTM models achieving RMSE of 1.195 and 1.135, respectively. A stepwise regression model was used by (Gupta et al., 2020) in predicting the intensity of tomato early blight disease. They achieved the RMSE of 6.129 and R^2 of 0.832. A stepwise regression model with temperature, rainfall, relative humidity, and wind speed was considered to predict potato blight disease (Dar et al., 2021). The four varieties of potatoes were chosen for the study. The Kufri Jyoti variety shows the best prediction performance with RMSE of 0.58 and R^2 of 0.82. Bhardwaj et al. (2021) used the M1, M2, and M3 models to predict the temperature and relative humidity in the Ludhiana region. M1 model performed better than the M2 model in terms of RMSE and R^2 in temperature prediction. M3 model achieved R^2 of 0.12 and RMSE of 12.09 in the prediction of relative humid- ity. Temperature Forecasting in Myintkyina was done using the Prophet model by Oo and Sabai (2020), getting an RMSE of 5.7573. Air temperature forecasting was done by (Toharudin et al., 2020) using the LSTM model with Adam and RMSProp optimizer and Prophet model. The RMSE of the Prophet model is 1.03, making it a better performer in prediction.

2.9 Summary of Literature Review

This chapter presents a comprehensive review of the classification of plant leaves classification and the prediction of plant disease. This chapter presents a brief overview of strategies related to PDC and PDP from different publications falling under bibliometric analysis. The classification techniques with image

processing can serve as a solution in various research domains. The overfitting problems and methods to avoid them to improve the model's performance in classification are discussed. The classification of plant species is useful as they are for medicine, food, and industrial applications. As a result, it is necessary to develop a method to identify them correctly. The main focus is on the classification of diseases in tomato plant leaves. The chapter focuses on the different techniques used to classify plant diseases. Machine learning and Deep Learning approaches and the related work in classifying various plant diseases are showcased here. The disease in the plant can occur under the influence of climatic parameters. The prediction of plant disease due to climatic conditions and the challenges and future scope to improve the technique for better performance are discussed.

The survey highlighted the widespread use of machine learning and deep learning approach in the classification process. These approaches are studied to classify and predict disease in plant leaves. The research work carried out by the researchers drew attention to deep learning methods. This generated a scope to improvise the deep learning approach for the classification and prediction of tomato plant disease.

CHAPTER THREE
3. METHODOLOGY

This section elaborates the architecture of the proposed model to classify and to detect the diseases in tomato leaves. Automatic disease detection in plants is the need of day in order to increase the productivity of the plants. We have designed a deep learning based network to detect the diseases in plants. Our proposed work will use Tomato leaf images which will be given to deep learning network that identifies the type of disease. This work has focused on designing a suitable deep learning network for disease detection by tuning the hyper parameters empirically. The affected tomato plant leaf images have been collected from the Plant Village bench mark dataset for the classification. Totally we have 7166 tomato leaf images available in the dataset. These images are divided into training and testing of 80% and 20% each respectively. The proposed classifier is designed to detect 11 different diseases. The diseases that are to be classified along with the healthy leaves of tomato are bacterial spot, early blight, late blight, leaf mold, septoria leaf spot, two spotted spider mite, target spot, mosaic virus, and yellow leaf curl virus. Deep learning Convolutional neural network (DCNN) can be used for the classification of images. The CNN has three parts such as Convolutional, Pooling and Fully Connected layers. The convolution and pooling layer can act as feature extractors. The early convolutional layers with kernel will extract low level features from the leaf images, which will be represented in the form of feature maps or Convolution map. These feature maps are fed to pooling layers to reduce the dimensionality of the intermediate output, thereby reducing the parameters and the computation overhead. This is otherwise referred to as sub sampling. The feature maps are again given to the next set of convolutional layers, where the higher-level features are learnt from accumulating the low-level ones. These extracted features are then given to the fully connected layer for classification. In our classification, the training images are given to the convolutional layer for the feature extraction. This layer has several filters or kernels which are placed over the image and are used to extract the features from the input image. The output of each convolution layer is given to a pooling layer. We have used various filter sizes to identify the best possible architecture to detect the diseases in the tomato plant images. We have built 4 different variants of CNN architecture by varying the filter size. 3 x 3, 5 x 5, 7 x 7 and 11 x 11 are the various filter sizes used to build the models. All the variants have 3 convolutional layers paired with 3 maxpooling layers. All the convolutional layers are employed with Relu as the activation function. The fully connected layer is implemented through the Dense layer with the number of classes as 11 and softmax as the activation function. The architecture of the proposed model is given as a model plot in

Figure 1. W1 x H1 x D1 be the width, height and depth of the input image. Input image of size 256 x 256 with 3 channels is fed to the first convolutional layer as input, which has 32 filters, each of size 3x3. The feature map generated by this layer is W2 x H2 x D2. The size of the feature map is computed as given in Eqs. 1, 2 and 3. $W2 = (W1 - F + 2P)/S + 1$ (1) $H2 = (H1 - F + 2P)/S + 1$ (2) $D2 = K$ (3) F is the size of each filter; P is the padding; S is the stride and K being the number of filters. We have not used any padding to retain the same input and output size. Stride is the step size at which the kernel/filter is slide over the receptive area of the input image. With W1=256, H1=256, F=3, P=0, S=1, the size of the feature map will be 254 x 254 x 32. The same computation is followed by the consecutive two convolutional layers. Maxpooling layer follows the same convention with pool size 2 x 2 and stride 2.This layer takes 2 x 2 input and determines the maximum of all the four pixels and produces that as the output. The activation function Relu applies the function $f(x) = \max(0, x)$ to all of the values in the input volume. This activation function also helps to handle the vanishing gradient problem, which is the issue where the lower layers of the network train very slowly since the gradient decreases exponentially through the layers. To alleviate the over-fitting problem, the dropout layer drops out a random set of activations in that layer by setting them to zero. A fully connected layer expects a 1D vector of numbers, hence flattening is applied to simply arranging the 3D volume of numbers into a 1D vector. The fully connected layer gets the extracted features from the previous feature extraction part of CNN. The last fully connected layer employs the softmax activation function in order to compute the classes. The output of the last fully connected layer will be a probability that identifies to which the class does an image belongs to.

DATA LABELING: Assigning labels to images of tomato plants, indicating the presence or absence of disease. Data labeling can be done manually by human experts, or it can be automated using machine learning algorithms.

['Tomato_Bacterial_spot' 'Tomato_Early_blight' 'Tomato_Late_blight'

'Tomato_Leaf_Mold' 'Tomato_Septoria_leaf_spot'

'Tomato_Spider_mites_Two_spotted_spider_mite' 'Tomato__Target_Spot'

'Tomato__Tomato_YellowLeaf__Curl_Virus' 'Tomato__Tomato_mosaic_virus'

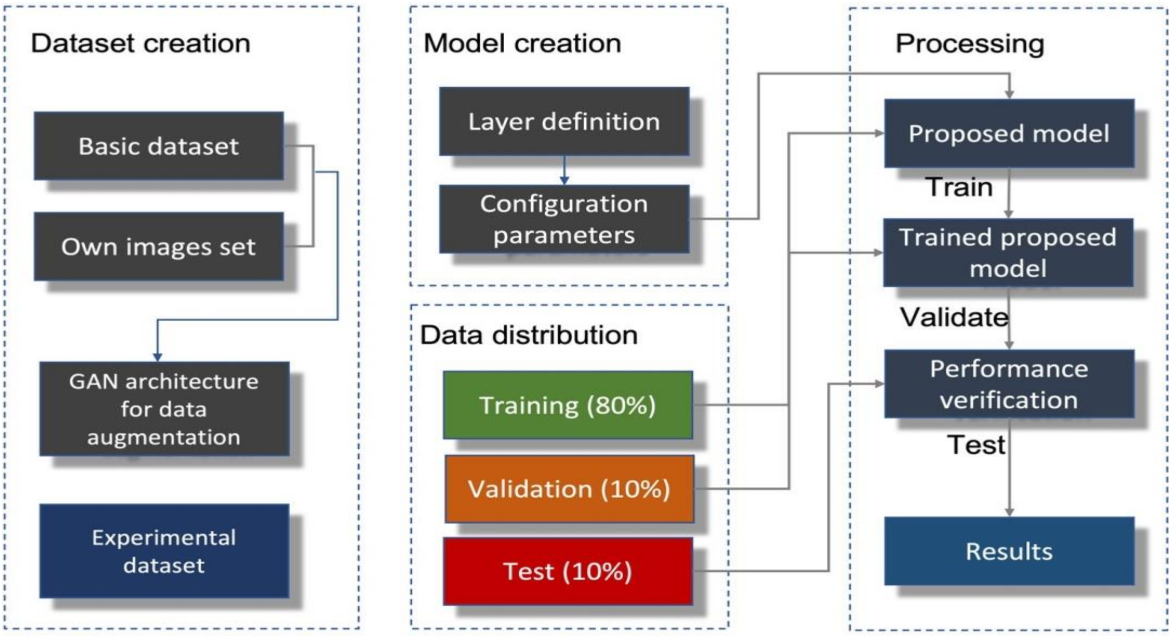
'Tomato_healthy' 'Unknown Crop']

IMAGE CAPTURE: Digital cameras can be used to capture images of tomato plants for analysis.

CHAPTER FOUR
SYSTEM DESGIN AND IMPLEMENTATION

4.1.1 Design

In this section, we explain in detail the proposed architecture for the detection of diseases in tomato leaves. In general, the proposed architecture takes tomato leaves as input images and the output is a set of labels indicating the type of disease in the image being analyzed or whether the leaf is healthy, Figure 2 shows the complete process of the algorithm that we applied for the process of detection and classification of diseases in tomato leaves. The global algorithm is composed of four stages: (a) creation of the experimental dataset, (b) creation of the proposed architecture, (c) distribution of the dataset, and (d) process of training and evaluation of the model.



Source: scopus database

Figure 1. Representation of the proposed architecture for tomato crop disease detection.

4.1.2 Dataset Creation

As a first step, we proceeded to create the experimental dataset that would be used for training, validation, and performance evaluation of the proposed architecture. The public dataset available in [21] consists of 11,000 images that were the basis of our dataset. The images represent 11 categories, including nine types of diseases (tomato mosaic virus, target spot, bacterial spot, tomato yellow leaf curl virus, late blight, leaf mold, early blight, two-spotted spider mites, septoria leaf spot) one category of healthy leaves and one category of Unknown crops. The dataset was complemented with 2500 images obtained from different crop fields in Mexico. The total number of images that made up our dataset was 13,500.

One of the problems that datasets face with deep neural network models is that when training the model, overfitting can occur, i.e., a model with high capacity may be able to “memorize” the dataset [22]. A technique known as data augmentation is used to avoid the problem of overfitting. The goal of applying data augmentation is to increase the size of the dataset, and it is widely used in all fields [23]. Commonly, data augmentation is performed by two methods. The first method, known as the traditional method, aims to obtain a new image, which contains the same semantic information but does not have the ability of generalization. These methods include translation, rotation, flip, brightness adjustment, affine transformation, Gaussian noise, etc. The main drawbacks of these methods may be their poor quality and inadequate diversity.

4.1.2 Model Creation

Figure 2 shows the proposed CNN architecture for disease detection in tomato. The network has 112 × 112 color images as input, which are normalized to (0, 1) values. The proposed convolutional network has four convolutional layers that use filters whose values were 16, 32, 64, and 128, respectively. These values were assigned in that orders since the layers closer to the beginning of the model learn convolutional filters less effectively than the layers closer to the result. In addition, the kernel size, which represents the width and height of the 2D convolution window, was set to a value of 3 × 3. This value was the recommended value for the number of filters to be used. Finally, rectified linear unit (ReLU) was used as the activation model for each convolved node.

After applying the convolutional layer, the maximum clustering layer was applied to down-sample the acquired feature map and condenses the most relevant features into patches. This process is repeated for each of the convolutional layers defined in the architecture.

The result of the last MaxPooling layer is passed to a MaxAveragePooling layer to be converted to a column vector and connected to the dense layer of 10 output nodes (which represent the 10 categories) used as softmax activation. Each node represents the probability of each category for the evaluated image. Table 1 shows the information of the layer structure of the proposed model.

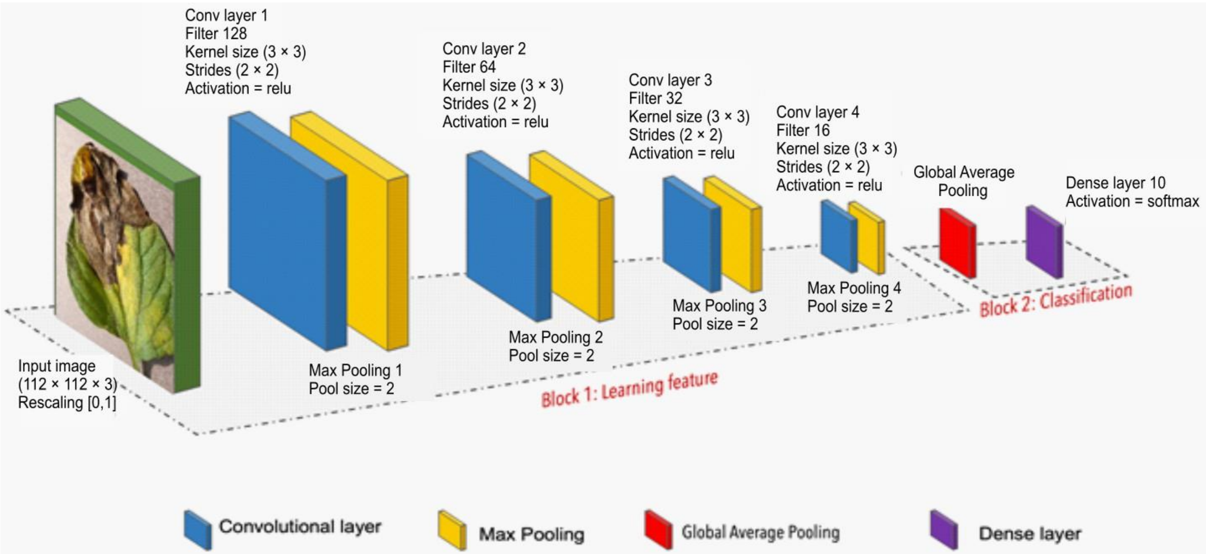


Figure 2. Representation of the proposed algorithm for tomato crop disease detection.

Table 1. Information on the layers structure of the proposed model.

Layers	**Parameters**
Conv2D	Filters: 128, kernel size: (3,3), activation: “relu”, input shape: (112,112,3)
MaxPool2D	Pool size: (2,2)

Conv2D	Filters: 64, kernel size: (3,3), activation: “relu”
MaxPool2D	Pool size: (2,2)
Conv2D	Filters: 32, kernel size: (3,3), activation: “relu”
MaxPool2D	Pool size: (2,2)
Conv2D	Filters: 16, kernel size: (3,3), activation: “relu”
MaxPool2D	Pool size: (2,2)
Dropout	Rate: 0.2
GlobalAveragePooling2D	
Dense	Units: 10, activation: “softmax”

4.1.3 Data Distribution

One of the most common strategies to split the dataset into training and validation sets is assigning percentages, for example, 70:30 or 80:20. However, one of the problems that can arise with this strategy is that it is uncertain whether high validation accuracy indicates a good model. When performing this division, it could happen that some information is missing in the data that are not used for training, causing a bias in the results.

We apply a k-fold cross-validation method to evaluate the performance of the model. The k-folds method tries to ensure that all features of the dataset are in the training and validation phases. The k-fold cross-validation method divides the dataset into subsets as k number. Therefore, it repeats the cross-validation method k times. Common values in machine learning are k = 3, k = 5, and k = 10. We use k = 5 to provide good trade-off of low computational cost and low bias in an estimate of model performance.

4.1.4 Model Creation

For the training process, we use Adam as the optimization algorithm. Adam up- dates network weights iterative based on training data. The loss function was categorical_crossentropy, one of the most used loss functions for multi-class classification models where there are two or more output labels. The number of epochs for the training and validation process was 200. The steps_per_epoch parameter was 12,000, and for the validation the parameter it was 3000. Table 2 shows a summary of some of the parameters used for the training and validation phase.

Table 2. Training Parameters for the Proposed Model.

Parameter	**Value**
Optimization algorithm	Adam
Loss function	Categorical cross entropy
Batch size	32
Number of epochs	200
Steps per epoch	12,000
Validation steps	3000
Activation function for conv layer	ReLu

4.1.5 Results

In this section, we describe the scenario setup and the results obtained in the performance evaluation process of the proposed model.

Environmental Setup

Our model was developed in Jupyter Notebook, a free Python development environment that runs in the cloud. Jupyter Notebook is widely used for the development of machine learning and deep leaning projects. In our project, we use the following libraries: Tensorflow, an open-source library used for numerical computation and automated learning; Keras, a library used for the creation of neural networks; numpy, used for data analysis and mathematical calculations; matplotlib used for graph management and TensorBoard to visually inspect the different runs and graphs.

The model was trained with 200 epochs. We applied early stopping to monitor the performance of the model for the 200 epochs on a held-out validation set during the training to reduce overfitting and to improve the generalization of the neural network. For the evaluation of the model, the validation accuracy scheme allowed early stopping to be activated during the process.

Since our problem is a multi-class classification model, we use the Adam algorithm as the optimizing algorithm. In addition, the cross-entropy categorical loss function was used due to the nature of the multi-class classification environment. During the training process, we implemented checkpoints to save the model with the best validation accuracy, and thus be able to load it later to continue training from the saved state if necessary.

Evaluation Metrics

To analyze the performance of our model, the following four metrics were considered. The first metric to evaluate was accuracy, which represents the behavior of the model across all classes. Accuracy is calculated as the ratio between the numbers of correct predictions to the total number of predictions (Equation (1)).

Precision was our second metric, which represents the accuracy of the model in classifying a sample as positive. This parameter is calculated as the ratio of the number of positive samples correctly classified to the total number of samples classified as positive (Equation (2)).

We also analyzed the recall parameter, which measures the ability of the model to detect positive samples and is calculated as the ratio of the number of positive samples correctly classified to the total number of positive samples (Equation (3)).

Finally, we analyzed the $F1$ score parameter. This metric combines the precision and recall measures to obtain a single value. This value is calculated by taking the harmonic mean between precision and recall (Equation (4)).

The following equations were used to calculate accuracy, precision, recall and $F1$ score:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

(3)

$$F1\ Score = 2 \times \frac{(precision \times recall)}{(precision + recall)} \tag{4}$$

4.2 Implementation

4.2.1 Developing the flask application

```
from flask import Flask, render_template, request

import numpy as np

import os

from tensorflow.keras.models import load_model

from tensorflow.keras.preprocessing import image

from werkzeug.utils import secure_filename



from flask import (Flask, redirect, render_template, request,

                   send_from_directory, url_for)



app = Flask(__name__)



# Load the model

model = load_model('/Project/cnn_model_2.h5')



def model_predict(img_path, model):
```

```
    test_image = image.load_img(img_path, target_size=(256, 256))  # Updated
size

    test_image = image.img_to_array(test_image)

    test_image = test_image / 255.0

    test_image = np.expand_dims(test_image, axis=0)

    result = model.predict(test_image)

    return result




@app.route('/', methods=['GET'])

def index():

    return render_template('index.html')


@app.route('/predict', methods=['POST'])

def predict():

    # Retrieve the uploaded image file

    image_file = request.files['file']


    # Save the file to the uploads folder

    basepath = os.path.dirname(os.path.realpath('__file__'))
```

```
    file_path = os.path.join(basepath, '/Project/uploads/', secure_filename(image_file.filename))

    image_file.save(file_path)



    try:

        # Make prediction

        result = model_predict(file_path, model)



        # Define your categories

        categories = ['Tomato_Bacterial_spot Disease', 'Tomato_Early_blight Disease', 'Tomato_Late_blight Disease', 'Tomato_Leaf_Mold Disease' , 'Tomato_Septoria_leaf_spot Disease' , 'Tomato_Spider_mites_Two_spotted_spider_mite Disease' , 'Tomato__Target_Spot Disease' , 'Tomato__Tomato_YellowLeaf__Curl_Virus' , 'Tomato__Tomato_mosaic_virus' , 'Tomato_healthy Leaf' , 'Unknown Crop']  # Update with your categories



        # Find the index of the class with the highest probability

        pred_class_index = np.argmax(result)



        # Map the index to the corresponding category

        pred_category = categories[pred_class_index]
```

```
        # Return the prediction category
        return pred_category
    except Exception as e:
        return str(e)  # Return the error message as a response



if __name__ == '__main__':
   app.run()
```

4.2.2 HTML Script

```
{% extends "base.html" %}

{% block content %}


<h2 style="text-align:center;">Tomato Crop Disease Detection</h2>

<div>
  <div class="containers">
    <div class="card p-3">
      <p style="text-align:center;">
```

```
        Welcome to our Crop Disease Detection model! Our system

        utilizes cutting-edge technology to identify diseases in

        tomato plants. With our image recognition technology and machine
learning

        algorithms, we provide fast and accurate diagnoses, helping you
maintain the

        health of your crop to prevent further damage."

      </p>

    </div>

  </div>

  <form id="upload-file" method="post" enctype="multipart/form-data">

    <label for="imageUpload" class="upload-label mt-3"> Upload... </label>

    <input

      type="file"

      name="file"

      id="imageUpload"

      accept=".jpg,"

    />

  </form>
```

```
  <div class="image-section" style="display: none">

    <div class="img-preview">

      <div id="imagePreview"></div>

    </div>

    <div>

      <button type="button" class="btn btn-primary btn-lg mt-3" id="btn-
predict">

        Detect

      </button>

    </div>

  </div>



  <div class="loader" style="display: none"></div>



  <h3 id="result">

    <span> </span>

  </h3>

</div>



<script src="https://code.jquery.com/jquery-3.6.0.min.js"></script>
```

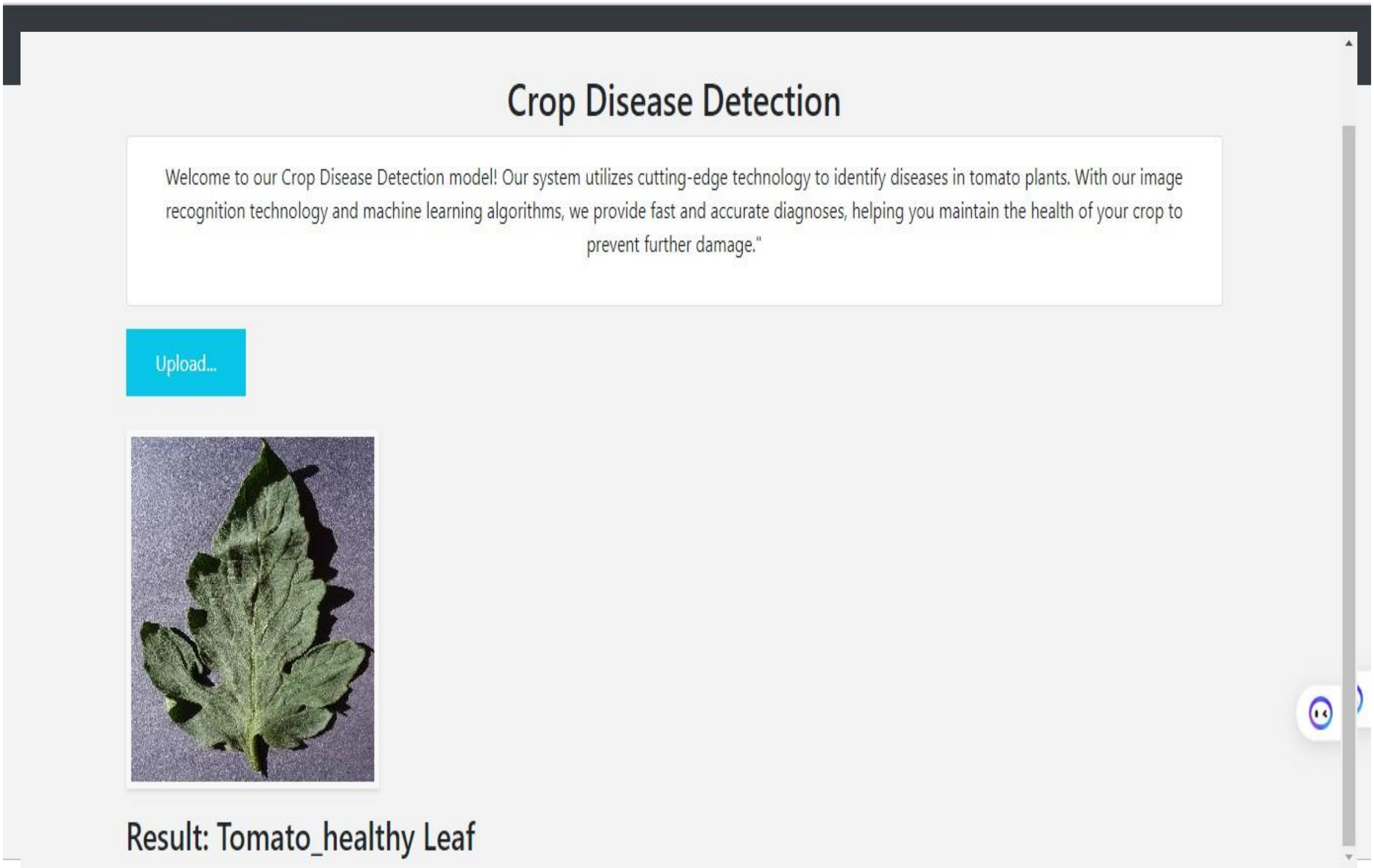
```
<script>
 $(document).ready(function() {
   // Display selected image preview
   $("#imageUpload").change(function() {
     var input = this;
     if (input.files && input.files[0]) {
       var reader = new FileReader();
       reader.onload = function(e) {
         $("#imagePreview").css(
           "background-image",
           "url(" + e.target.result + ")"
         );
       };
       reader.readAsDataURL(input.files[0]);
       $(".image-section").show();
     }
   });

   // Perform prediction on button click
   $("#btn-predict").click(function() {
```

```
    var form_data = new FormData($("#upload-file")[0]);
    $(".loader").show();
    $.ajax({
      type: "POST",
      url: "/predict",
      data: form_data,
      contentType: false,
      cache: false,
      processData: false,
      success: function(result) {
        $(".loader").hide();
        $("#result span").text("Result: " + result);
      },
      error: function() {
        $(".loader").hide();
        alert("Error occurred during prediction.");
      },
    });
  });
});
```

</script>

{% endblock %}

Home Page




Crop Disease Detection
Welcome to our Crop Disease Detection model! Our system utilizes cutting-edge technology to identify diseases in tomato plants. With our image recognition technology and machine learning algorithms, we provide fast and accurate diagnoses, helping you maintain the health of your crop to prevent further damage."
Upload...
Result: Unknown Crop

CHAPTER FIVE

SUMMARY CONCLUSION AND RECOMMENDATION

5.1 Summary

This project aims to develop a web-based application designed to detect diseases in tomato crop. The Crop Disease Detection Model will revolutionize farming practices by empowering farmers with essential knowledge on the precise disease affecting their tomato crop, leading to increased productivity.

The process begins with the collection of a diverse dataset comprising images of healthy and diseased tomato plants. These images serve as the training material for the CNN, which learns to differentiate between various disease symptoms. The convolutional layers in the network extract hierarchical features, recognizing patterns at different scales.

During the training phase, the model optimizes its parameters to minimize the difference between predicted and actual disease labels. Once trained, the CNN can be deployed to analyze new images and identify potential diseases. This automated detection system offers a rapid and reliable way for farmers to monitor the health of their tomato crops, facilitating timely actions to prevent widespread infections and reduce crop losses.

5.2 Conclusion

In this research, we propose architecture based on CNNs to identify and classify nine different types of tomato leaf diseases. The complexity in detecting the type of disease lies in the fact that the leaves deteriorate in a similar way in most of the tomato diseases. It means that it is necessary to develop a deep image analysis to judge the types of tomato leave diseases with a proper accuracy level. The CNN that we design is a high-performance deep learning network that allows us to have a complex image processing and feature extraction through four modules: the module dataset creation that makes an experimental dataset using public datasets and photographs taken in the fields of the country; model creation that is in charge of parameters configuration and layers definition; data distribution to train, validate and test data; and processing for the optimization and performance verification. We evaluate the performance of our model via accuracy, precision, recall and the F1- score metrics. The results showed a training accuracy of 90% and a validation accuracy of 90.76% in the leaf disease classification. The model correctly classifies the corresponding disease with a precision of 0.99 and an F1 score of 0.99. The recall metric has a value of 0.99 on the classification of the nine tomato diseases that we analyzed. The resulting confusion matrix describes that our classification model was able to predict half of the classes that were evaluated using the test dataset with 100% accuracy. For the rest of the classes, the model reached an accuracy level of 98%, thus obtaining better values than those of several of the works proposed in the literature

5.3 Recommendation

Based on the literature reviewed, several recommendations emerge for the development and implementation of the Crop Disease Detection Model:

- Conduct extensive field trials across diverse agro-ecological zones in Sierra Leone to validate the app's accuracy and suitability for local conditions.
- Incorporate user feedback during the app's development process to create a user-friendly interface and improve user adoption rates.
- Continuously monitor and update the app based on user inputs and data analytics to enhance its performance and relevance over time.

REFERRENCES

1. Gobalakrishnan, N.; Pradeep, K.; Raman, C.J.; Ali, L.J.; Gopinath, M.P. A Systematic Review on Image Processing and Machine Learning Techniques for Detecting Plant Diseases. In Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, Sierra Leone, 28–30 July 2020; pp. 0465–0468. [CrossRef]
2. Damicone, J.; Brandenberger, L. Common Diseases of Tomatoes: Part I. Diseases Caused by Fungi—Oklahoma State University. 2016. Available online: https://extension.okstate.edu/fact-sheets/common-diseases-of-tomatoes-part-i-diseases-caused-by- fungi.html (accessed on 19 October 2022).
3. Bock, C.H.; Parker, P.E.; Cook, A.Z.; Gottwald, T.R. Visual Rating and the Use of Image Analysis for Assessing Different Symptoms of Citrus Canker on Grapefruit Leaves. *Plant Dis.* **2008**, *92*, 530–541. [CrossRef] [PubMed]
4. Mugithe, P.K.; Mudunuri, R.V.; Rajasekar, B.; Karthikeyan, S. Image Processing Technique for Automatic Detection of Plant Diseases and Alerting System in Agricultural Farms. In Proceedings of the 2020 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 28–30 July 2020; pp. 1603–1607. [CrossRef]
5. Thangadurai, K.; Padmavathi, K. Computer Visionimage Enhancement for Plant Leaves Disease Detection. In Proceedings of the 2014 World Congress on Computing and Communication Technologies, Trichirappalli, India, 27 February–1 March 2014; pp. 173–175. [CrossRef]
6. Khirade, S.D.; Patil, A.B. Plant Disease Detection Using Image Processing. In Proceedings of the 2015 International Conference on Computing Communication Control and Automation, Pune, India, 26–27 February 2015; pp. 768–771. [CrossRef]
7. Li, L.; Zhang, S.; Wang, B. Plant Disease Detection and Classification by Deep Learning—A Review. *IEEE Access* **2021**, *9*, 56683–56698.

[CrossRef]

8. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep learning in agriculture: A survey. *Comput. Electron. Agric.* **2018**, *147*, 70–90. [CrossRef]
9. Lee, S.H.; Chan, C.S.; Wilkin, P.; Remagnino, P. Deep-plant: Plant identification with convolutional neural networks. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada; 2015; pp. 452–456. [CrossRef]
10. Zhang, Y.; Song, C.; Zhang, D. Deep Learning-Based Object Detection Improvement for Tomato Disease. *IEEE Access* **2020**, *8*, 56607–56614. [CrossRef]
11. Widiyanto, S.; Wardani, D.T.; Pranata, S.W. Image-Based Tomato Maturity Classification and Detection Using Faster R-CNN Method. In Proceedings of the 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 21–23 October 2021; pp. 130–134. [CrossRef]
12. Hlaing, C.S.; Zaw, S.M.M. Tomato Plant Diseases Classification Using Statistical Texture Feature and Color Feature. In Proceedings of the 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS), Singapore, 6–8 June 2018; pp. 439–444. [CrossRef]
13. Lu, J.; Shao, G.; Gao, Y.; Zhang, K.; Wei, Q.; Cheng, J. Effects of water deficit combined with soil texture, soil bulk density and tomato variety on tomato fruit quality: A meta-analysis. *Agric. Water Manag.* **2021**, *243*, 106427. [CrossRef]
14. Kaur, S.; Pandey, S.; Goel, S. Plants Disease Identification and Classification Through Leaf Images: A Survey. *Arch. Comput. Methods Eng.* **2018**, *26*, 507–530. [CrossRef]
15. Bhagat, M.; Kumar, D.; Haque, I.; Munda, H.S.; Bhagat, R. Plant Leaf Disease Classification Using Grid Search Based SVM. In Proceedings of the 2nd International Conference on Data, Engineering and Applications (IDEA), Bhopal, India, 28–29 February 2020; pp. 1–6. [CrossRef]
16. Rani, F.A.P.; Kumar, S.N.; Fred, A.L.; Dyson, C.; Suresh, V.; Jeba, P.S.

K-means Clustering and SVM for Plant Leaf Disease Detection and Classification. In Proceedings of the 2019 International Conference on Recent Advances in Energy-efficient Computing and Communication (ICRAECC), Nagercoil, India, 7–8 March 2019; pp. 1–4. [CrossRef]
17. Mokhtar, U.; Ali, M.A.S.; Hassenian, A.E.; Hefny, H. Tomato leaves diseases detection approach based on Support Vector Machines. In Proceedings of the 2015 11th International Computer Engineering Conference (ICENCO), Cairo, Egypt, 29–30 December 2015; pp. 246–250. [CrossRef]
18. Sabrol, H.; Satish, K. Tomato plant disease classification in digital images using classification tree. In Proceedings of the 2016 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 6–8 April 2016; pp. 1242–1246.
19. Molina, F.; Gil, R.; Bojacá, C.; Gómez, F.; Franco, H. Automatic detection of early blight infection on tomato crops using a color based classification strategy. In Proceedings of the 2014 XIX Symposium on Image, Signal Processing and Artificial Vision, Armenia, Colombia, 17–19 September 2014; pp. 1–5. [CrossRef]
20. Jones, C.; Jones, J.; Lee, W.S. Diagnosis of bacterial spot of tomato using spectral signatures. *Comput. Electron. Agric.* **2010**, *74*, 329–335. [CrossRef]
21. Tomato Leaf Disease Detection. Available online: https://www.kaggle.com/datasets/kaustubhb999/tomatoleaf (accessed on 24 October 2022).
22. Konidaris, F.; Tagaris, T.; Sdraka, M.; Stafylopatis, A. Generative Adversarial Networks as an Advanced Data Augmentation Technique for MRI Data. In Proceedings of the VISIGRAPP, Prague, Czech Republic, 25–27 February 2019.
23. Kukacka, J.; Golkov, V.; Cremers, D. Regularization for Deep Learning: A Taxonomy. *arXiv* **2017**, arXiv:1710.10686.