Plant Disease Detection using Deep Learning

Nathan Abela

Supervisor: Frankie Inguanez

June - 2024

A dissertation submitted to the Institute of Information and Communication Technology in partial fulfilment of the requirements for the degree of BSc (Hons) Software Development

Authorship Statement

This dissertation is based on the results of research carried out by myself, is my own composition, and has not been previously presented for any other certified or uncertified qualification.

The research was carried out under the supervision of Mr. Frankie Inguanez.

07/06/2024 ....

Date Signature

Copyright Statement

07/06/2024 . . . .

Date Signature

Acknowledgements

I would like to express my deepest gratitude to Mr. Frankie Inguanez for his unwavering support and invaluable guidance throughout the process of writing this dissertation. His wealth of knowledge, insightful feedback, and commitment to excellence have been instrumental in shaping my research.

I would also like to extend my heartfelt appreciation to my family, including my parents and brother, whose continuous support and encouragement have been the cornerstone of my academic pursuits. Their constant motivation and belief in my abilities have provided me with the strength and determination to overcome challenges and reach new heights.

Abstract

The integration of Artificial Intelligence and Computer Vision technologies into Precision Agriculture represents a significant advancement aimed at enhancing crop productivity and addressing critical challenges in global food security. This study considers recent deep-learning-based algorithms for efficient and accurate detection of plant diseases, a critical concern given the estimated 20-40% global crop yield losses annually. Traditional detection methods, reliant on human expertise, are not only labour-intensive but also prone to errors, making the case for automated, more reliable solutions.

This research employs advanced Convolutional Neural Networks (CNNs), particularly EfficientNet and VGG16, to automate and refine the disease detection process. The study was structured into three distinct experimental phases. The first phase involved replicating established methodologies using the well-documented PlantVillage dataset to set a baseline and validate the effectiveness of existing models. The second phase focused on the development of a custom deep learning algorithm, where extensive hyperparameter tuning and architectural adjustments were applied to enhance model performance. This phase was critical in optimising the deep learning models specifically for plant disease detection tasks, leading to significant improvements in detection capabilities.

In the final phase, the model was applied to a custom dataset relevant to Malta's agricultural sector, which included key local crops, potatoes, strawberries, and tomatoes. This phase aimed to assess the practical utility and real-world efficacy of the model, providing a robust evaluation of its performance in detecting plant diseases under actual agricultural conditions.

The study's results demonstrate the effectiveness of deep learning models in plant disease detection across several phases. Initially, replication of existing EfficientNet and VGG16 models on the PlantVillage dataset yielded an accuracy of 95.64% and 94.22%, respectively, confirming their robustness with a difference of 2% from the original studies. Subsequently, the custom deep learning algorithm, by systematically adjusting the hyperparameters, achieved an accuracy of 81.01%. In the last phase, the application resulted in a strong overall performance, with an accuracy of 85.13%. These results highlight the practical utility and adaptability of optimised deep learning systems in real-world agricultural settings.

This study not only validates the effectiveness of deep learning applications in precision agriculture but also builds upon and strengthens previous research on the subject matter, thus laying a foundation for future research. The proven effectiveness of CNNs in practical agricultural settings builds upon existing research and paves the way for further innovations. Future research could expand the model's adaptability to diverse crop types and explore the integration of real-time data acquisition systems, moving towards fully automated, AI-driven disease management systems that could revolutionise global agricultural practices.

Keywords: Precision Agriculture, Convolutional Neural Networks, Computer Vision, Deep Learning, EfficientNet, VGG16, Hyperparameter Tuning.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ACC	Accuracy
AGRIHUB	Agriculture Research and Innovation Hub
AI	Artificial Intelligence
CNN	Convolutional Neural Network
CV	Computer Vision
DL	Deep Learning
ERA	Environment and Resources Authority
EU	European Union
F	False
FN	False Negative
FP	False Positive
Grad-CAM	Gradient-weighted Class Activation Mapping
mAP	Mean Average Precision
MCAST	Malta College of Arts, Science, & Technology
ML	Machine Learning
NN	Neural Network
PA	Precision Agriculture
PPV	Precision
SVM	Support Vector Machine
T	True
TN	True Negative
TP	True Positive
TRL	Technology Readiness Levels
TPR	Recall

Chapter 1: Introduction

1.1 Definition and Significance of the Problem

Plant cultivation has been a vital trade for centuries, driven by the need to feed millions of people. To achieve higher and healthier crop yields, the farming industry has embraced the combination of modern technology and agricultural practices, leading to the creation of Precision Agriculture (PA) [6]. This study aims to address these challenges through the development of an efficient and accurate plant disease detection system using deep learning techniques and investigating how such technologies can be beneficial for the local agricultural sector.

Traditional methods for plant disease detection rely on visual inspection by human experts. While effective, these methods are time-consuming, labour-intensive, and prone to human error. Current state-of-the-art systems utilise pre-trained Convolutional Neural Network (CNN) models, such as EfficientNet [2] and VGG16 [25].

In recent years, advancements in Artificial Intelligence (AI) and Computer Vision (CV) have not only revolutionised agricultural practices but have also extended to innovative methods for combating plant diseases. The use of laser-based technologies exemplifies this trend, showcasing the potential of AI-driven systems to accurately identify and target diseased plants [19]. By leveraging computer vision algorithms, these systems can precisely locate and selectively eliminate diseased plant parts, offering a promising solution to enhance crop quantity, quality, and safety in agricultural production.

1.2 Research Motivation

The motivation behind this research is multifaceted. Globally, the agricultural sector faces mounting pressure to enhance productivity while minimising environmental impact. According to the Food and Agriculture Organization (FAO), plant diseases are responsible for an estimated 20-40% of global crop yield losses, amounting to over $220 billion annually [7]. This staggering statistic showcases the impact of plant diseases on agricultural production and the urgent need for innovative solutions to mitigate these losses.

Locally and across the European Union (EU), farmers are voicing their concerns and engaging in protests, reflecting broader discontent within the agricultural community [21]. These protests signify the underlying concerns regarding government policies and other factors that are impacting farming practices, ultimately leading to a decline in farming participation. This situation highlights the challenges faced by farmers and the urgent need for policy frameworks that support sustainable agricultural practices.

Moreover, as the use of AI continues to increase rapidly, there is a growing interest in exploring its applications and understanding its underlying mechanisms. With recent research focusing more on the applied use of Generative AI, there is a risk of overshadowing the foundational principles of deep learning algorithms. This research endeavour serves as an opportunity to delve deeper into the workings of AI, particularly within the domain of deep learning, and to contribute to the advancement of knowledge in this field.

1.3 Hypothesis and Research Questions

This research hypothesises that recent advancements in deep learning models, particularly Convolutional Neural Networks (CNNs), permit a more holistic and accurate detection and classification of various plant diseases. To address this hypothesis, the following research questions were formulated:

1. How effectively can established deep learning methodologies from key research papers be replicated for plant disease detection? What insights can be gained from this process regarding the strengths and limitations of existing approaches?

2. Can computer vision techniques accurately identify key characteristics that differentiate healthy plants from those infected with various diseases?

3. How can an experimentation plan aid in the identification of model architecture variations for the advancement of plant disease detection?

4. How do variations in CNN architecture, such as the number of layers, filter sizes, and pooling strategies, influence the accuracy of plant disease detection?

5. What impact do different hyperparameters, including learning rate and batch size, have on the performance of deep-learning models for plant disease detection?

6. How does the proposed solution perform on the detection and recognition of key and important local crops?

1.4 Document Structure

The dissertation is structured as follows, Chapter 2 provides a concise literature review covering plant diseases, agricultural datasets, state-of-the-art CNN-based models, and local contributions. Chapter 3 outlines the research methodology and experimental procedures undertaken. In Chapter 4, the analysis of results and their subsequent discussions are presented, exploring their implications and relevance. Lastly, Chapter 5 offers conclusions based on the research findings, summarises key points, and suggesting recommendations for future research in the field.

Chapter 2: Literature Review

The integration of computer vision into agriculture, particularly in detecting plant health, marks a major progress in farming methodologies. Utilising AI and machine learning (ML), especially through deep learning techniques such as neural networks, plays a pivotal role in the identification and management of plant diseases. The importance of identifying plant disease promptly is very important, as it directly influences crop health, yield, and overall agricultural productivity. Undetected diseases pose a significant threat to food security, making early identification crucial for sustainable farming practices.

The use of AI and ML techniques offers a solution to overcome challenges in crop management. With the help of deep learning, these technologies enable accurate and efficient disease detection, reducing the need to rely on traditional methods and pesticides. This not only enhances crop productivity but also promotes environmentally friendly and sustainable agricultural practices. The PA field has become a key driver for optimising farming processes [6]. PA leverages advanced technologies, including computer vision and AI, to tailor agricultural practices with a high degree of accuracy.

2.1 Plant Diseases

Plant diseases pose significant challenges to agriculture globally. These diseases not only affect crop yield and quality but also have a considerable economic impact on the agriculture industry. Plant diseases are responsible for an estimated 20-40% of global crop yield losses, translating to over $220 billion annually, according to the Food and Agriculture Organisation (FAO) [7].

Plant diseases are caused by living organisms that attack and obtain nutrition from the plants they infect. These organisms are known as pathogens, which can take on various forms, these include fungi, bacteria, viruses, and nematodes. The plant targeted by a pathogen and used for its nutrients is referred to as a host [1].

In addition, the role of the environment plays a crucial role in disease development. Vulnerable plants exposed to significant pathogens will not develop disease unless environmental conditions are favourable. These environmental factors can include temperature, humidity, and soil conditions. Optimal conditions create a breeding ground for pathogens, emphasising the need to consider and manage environmental factors for effective disease control.

These three factors, often referred to as the plant disease triangle [18], form the cornerstone of understanding and managing plant diseases. An illustration of this disease triangle can be seen in Figure 2.1. In the plant disease triangle, the vertices represent the host plant, the pathogen, and the environment [24]. The occurrence of plant diseases is dependent on the interaction of these three elements. The host plant being prone to the pathogen plays a crucial role in the dynamics of the development of the disease. As such, the presence and activity of the pathogen, along with environmental conditions, contribute to the manifestation of plant diseases. Recognising and addressing each factor in this triangular relationship is essential for implementing targeted and effective strategies in plant disease prevention and control.

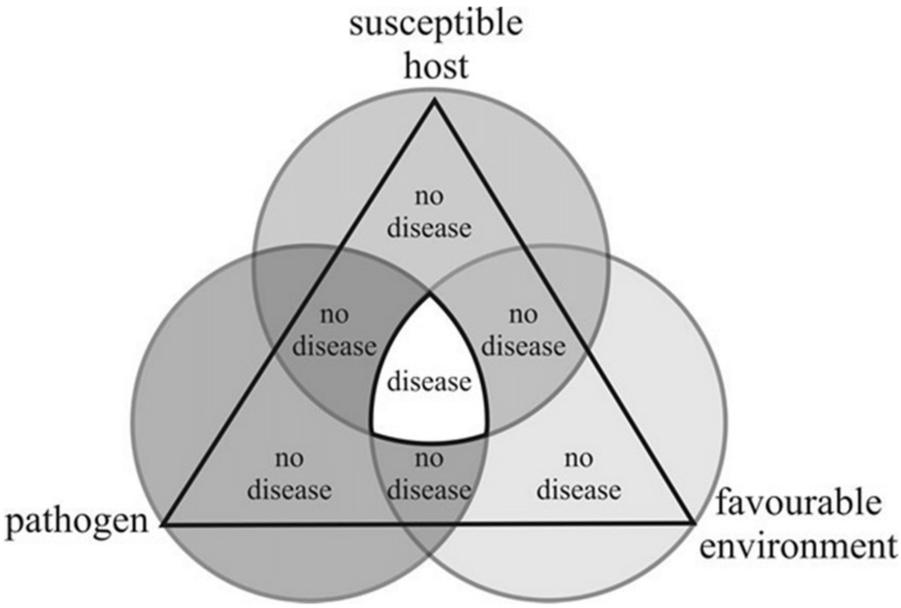

Figure 2.1: *The Plant Disease Triangle [18, p. 375]*

2.2 Datasets in Agricultural Research

2.2.1 PlantVillage

In agricultural research, the PlantVillage dataset [12] is a key resource, especially when it comes to computer vision-based plant disease identification. The dataset consists of 54,303 images of healthy and unhealthy leaves with a resolution of 256 by 256 pixels, which are categorised into 38 distinct classes, 14 different plant species in total, 12 of which are healthy and 26 of which are diseased. These classes represent a variety of plant species and diseases, as illustrated in

Figure 2.2. The leaves were removed from the plant, placed on a grey background, and photographed outdoors with a single digital camera on sunny and cloudy days.

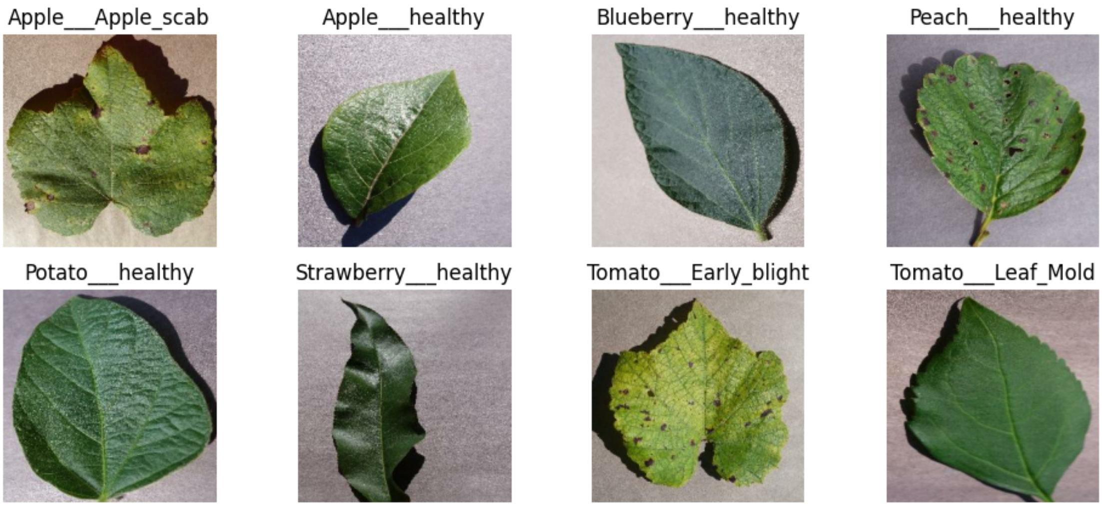

Figure 2.2: *Sample images of 8 classes within the PlantVillage dataset*

2.2.2 PlantDoc

The PlantDoc dataset [26], although smaller than the PlantVillage dataset, is another valuable resource in agricultural research using computer vision. This dataset is tailored towards the detection and classification of plant diseases using images. Unlike larger datasets, this dataset consists of 2,598 internet-scraped images categorised into 13 plant species and up to 17 disease classes. The motivation for the authors to create this dataset lies in addressing the lack of large-scale non-laboratory datasets for vision-based plant disease detection [26]. The unique aspect of PlantDoc is its emphasis on reflecting real-world scenarios.

2.2.3 PlantLeaves

The PlantLeaves dataset [8] uniquely targets 12 economically and environmentally significant plants in India. The dataset consists of 4,503 high-quality images, split into 2,278 healthy and 2,225 unhealthy plant leaves. This focused approach allows for a detailed examination of plant diseases within specific species. The images were captured in a closed environment with controlled lighting and camera settings, ensuring standardisation for image analysis tasks and minimising external noise. Unlike datasets with broader disease categories, PlantLeaves potentially offers insights into specific disease types within each selected plant species.

2.2.4 PlantNet-300K

The PlantNet-300K dataset [10] is another important resource in agricultural research, consisting of 306,146 images covering 1,081 plant species. Its size surpasses that of other plant datasets, offering a diverse collection of plant images suitable for extensive and deep learning models. Notably, the dataset's distinctive features include high intrinsic ambiguity and a long-tailed distribution, presenting challenges such as a strong class imbalance and visual similarity among species. These characteristics make PlantNet-300K well-suited for the evaluation of set-valued classification methods. While this dataset's primary focus is not on disease detection, it contributes valuable insights into agriculture and serves as a substantial resource for researchers exploring plant image analysis beyond disease-related applications.

A summary of the datasets mentioned above is shown in Table 2.1.

Dataset	No. of Images	No. of Classes	Resolution
PlantVillage [12]	54,303	38	256x256
PlantDoc [26]	2,598	17	Various
PlantLeaves [8]	4,503	22	6000x4000
PlantNet-300K [10]	306,146	1,081	600x600

Table 2.1: *Summary of commonly used datasets in plant disease detection research*

2.3 State of the art CNN based models

The current state of the art in plant disease detection is shaped by a range of studies, each contributing unique insights into the PA field. A comprehensive overview of key papers is essential for identifying trends, challenges, and opportunities and guiding the formulation of an effective research methodology.

2.3.1 Deep Learning

Mohanty et al. [17] trained a deep convolutional neural network on the PlantVillage dataset [12]. The models utilised in the study, namely, AlexNet and GoogLeNet, achieved high overall accuracy ranging from 85.53% to 99.35% respectively. The researchers evaluated their models using various train-test splits, such as 80:20, 60:40, 50:50, 40:60, and 20:80, to assess overfitting and evaluate their performance on unseen data. Among these, the 80:20 ratio performed the best in terms of split and model performance. These experiments demonstrated the feasibility and effectiveness of their proposed approach.

Similarly, Mohameth et al. [16] explored similar architectures and classifiers. The models used were VGG16, ResNet50, and GoogLeNet to extract meaningful features. For classification purposes, they utilised K-Nearest Neighbors and Support Vector Machine (SVM) classifiers. The researchers observed that SVM combined with VGG16 yielded superior performance, achieving an accuracy rate of 97.82%.

Atila et al. [3] used EfficientNet architecture to detect diseases in plants and demonstrated that EfficientNet outperforms the standard CNN models such as AlexNet, VGG16, and others, with an accuracy rate of 99.97% using EfficientNet. This model offers faster training times due to its reduced parameter count in comparison to other deep learning models. To enhance the dataset, the researchers increased the number of images to 61,486 by applying image augmentation techniques to labels with fewer than 1,000 samples. This approach followed the principles outlined by Gopal and Pandian [11]. The six image augmentation methods employed were flipping, gamma correction, noise injection, Principal Component Analysis (PCA) colour augmentation, rotation, and scaling.

As outlined above, recent researchers are focusing on popular deep learning architectures, including GoogLeNet [17], VGG16 [16], and EfficientNet [3]. These architectures have been used to tackle various machine learning problems. These CNN models have proven to be effective in extracting essential features, leading to impressive performance in numerous applications.

GoogLeNet

GoogLeNet is a renowned deep convolutional neural network with 22 layers, offering a powerful architecture for plant image classification [17]. Its interconnected layers, include convolutional layers, pooling layers, fully connected layers, and softmax layers for classification purposes. Mohanty et al. [17] opted for GoogLeNet due to its efficient structure, balancing depth and width with a manageable number of parameters (5 million parameters [13]). This efficiency reduces overfitting risks, which is especially valuable for smaller datasets.

VGG16

VGG16 is a CNN architecture that consists of 16 weight layers, including 13 convolutional layers and 3 fully connected layers [16]. The main feature of this architecture is that instead of having a large number of hyperparameters, it uses 3×3 convolutional filters throughout the network, followed by max-pooling layers to reduce spatial dimensions. This reduces the overall number of parameters and computational complexity, making it ideal for tasks with limited data or resources [25]. In addition, it helps capture localised features effectively, which is crucial for identifying subtle variations in plant morphology. Mohameth et al. [16] opted for VGG16 due to its established effectiveness in computer vision tasks and for achieving strong initial results. This model is considered state-of-the-art due to its accuracy rate of 97.82% in various image classification applications, showcasing its proven performance.

EfficientNet

Atila et al. [3] employed EfficientNet, a deep convolutional neural network architecture known for its efficiency and high performance. It uses compound scaling, where each dimension uniformly scales using a predetermined set of scaling coefficients rather than arbitrarily scaling width, depth, or resolution. A basic building block of EfficientNet-B0 with respect to Mobile Inverted Bottleneck (MBConv) layers is shown in Figure 2.3. The choice of EfficientNet aligns with the study's goal of achieving high accuracy while potentially operating with limited data or computational resources. The authors justify the choice of EfficientNet due to its state-of-the-art performance (84.3% accuracy with 66M parameters [27]) and scalability through pre-defined scaling coefficients. Ultimately, their experimental results confirm the success of this choice, with the EfficientNet B4 and B5 models achieving an accuracy rate of 99.97% and 99.39%, and a precision rate of 99.39% and 98.42%, respectively, on the PlantVillage dataset.

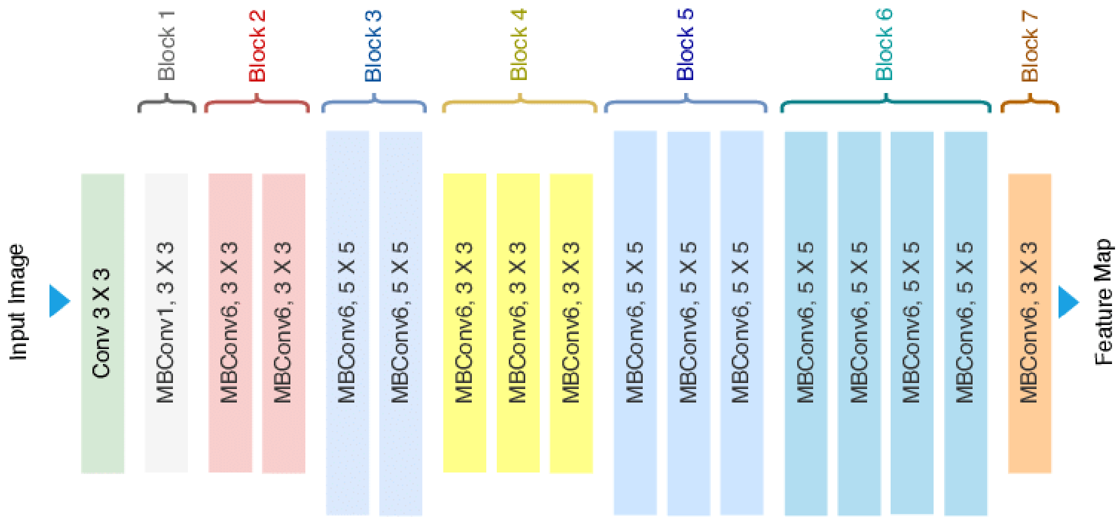

Figure 2.3: *Overview of the EfficientNet-B0 [2, p. 3] architecture, highlighting its layers and overall design*

This research was evaluated using accuracy, precision, recall, and the F1 score. The evaluation using accuracy is particularly important as it helps to gauge the overall correctness of the predictions, which is crucial in various applications and domains.

2.3.2 Summary Results

The performance of different machine learning models trained on the complete PlantVillage dataset is compared in Table 2.2. The evaluation metrics used in this comparison are based on accuracy, indicating that Atila et al. [3] achieved the highest performance. The authors utilised the EfficientNet architecture along with a combination of image augmentation and transfer learning techniques.

Study	**Classifier**	**Accuracy**
Mohanty et al. [17]	GoogLeNet	99.35%
Mohameth et al. [16]	VGG16	97.82%
Atila et al. [3]	EfficientNet	99.97%

Table 2.2: *PlantVillage Accuracy Evaluation Comparison*

2.4 Local Contributions

In the Galdes [9] study, the emphasis was on tomato plant disease detection, utilising the PlantVillage dataset. The dataset featured a subset with nine distinct tomato plant diseases and one class representing healthy plants. Coloured images were used, and various models were tested, exploring the impact of image quality and quantity. The study concluded that the XGBoost and VGG16 combination showed promise, achieving a 91% accuracy, but computational power limitations were acknowledged.

On the other hand, Busuttil's [5] research consisted of potato leaf disease detection using YOLOv5, using the PlantVillage dataset. The author compared the impact of including background information in training images on the model's accuracy. The results revealed that while a simpler configuration, which retained the background, achieved a Mean Average Precision (mAP) rate of 99.1%, excluding the background did not significantly impact accuracy, reaching 97.3%. The study also reports the accuracy values, where the configuration with background achieved 98.5% accuracy, compared to 91.75% without background. This finding further emphasises the potential benefit of retaining background information for improved model performance. In addition, the study explored the influence of image size on model performance. Scaling the images to 416x416 resulted in the highest accuracy and mAP rates, highlighting the importance of this factor for optimal model effectiveness. This suggests that YOLOv5 can be effectively utilised for plant disease detection with minimal configuration requirements.

Furthermore, Mercieca [14] addressed crop and disease classification using various ML and DL techniques. The research explored the application of CNNs, specifically utilising the VGG16 pre-trained model, along with transfer learning methods on both self-made and publicly available datasets. The self-made dataset was compiled by the researcher, who took photographs of various fruits in his home using a smartphone camera. This hands-on approach ensured a diverse and realistic dataset, capturing fruits in different lighting conditions and settings. For crop classification, the self-made dataset consisting of apples, grapes, and potatoes was used, while the disease classification task employed the PlantVillage dataset using the same plant types. The models achieved high accuracy of 98.67% and 96.33% for crop and disease classification, respectively. Additionally, Mercieca [14] developed a web application incorporating the best-performing models, demonstrating the potential for real-world implementation.

In another relevant study, Xuereb [30] explored the potential of CNNs for plant disease and healthy plant identification. Highlighting the threat these issues pose to agriculture and the environment, the study investigated the effectiveness of two CNN algorithms, AlexNet and GoogLeNet, in classifying 15 classes within the PlantVillage dataset. The subset of the full dataset considered includes segmented, coloured, and grey-scale versions of the 23,000 sampled images. Notably, the study employed fewer configurations than typically used for model testing. The research involved training, validation, analysis of generated graphs, post-training evaluation using F1-score, and layer visualisations for result generation and analysis. Through this process, Xuereb [30] found GoogLeNet to be more robust and effective than AlexNet, achieving an accuracy of 95% compared to AlexNet's 78%. This finding suggests that GoogLeNet's performance is superior in classifying the chosen plant disease and healthy plant categories.

The studies presented in this section highlight the research efforts in plant disease detection and crop classification at a local level. These studies, conducted by undergraduate students at Malta College of Arts, Science, & Technology (MCAST) as part of their final dissertations for the Bachelor of Science (Honours) in Software Development degree, contribute valuable insights to this field. These findings showcase diverse approaches and promising results, demonstrating the potential of local research efforts to contribute to advancements in this area. A summary of these findings is presented in Table 2.3.

Aspect	**Galdes** [9]	**Busuttil** [5]	**Mercieca** [14]	**Xuereb** [30]
Detection Type	Disease	Disease	Crop & Disease	Disease
Datasets	PlantVillage	PlantVillage	PlantVillage & Custom	PlantVillage
Plants	Tomato	Potato	Apples, Grapes & Potatoes	Various
Classes	10	3	11	15
Algorithm	XGBoost + VGG16	YOLOv5	VGG16	GoogLeNet & AlexNet
Accuracy	91%	98.5%	98.67%	95% & 78%

Table 2.3: *Summary of Local Research Contributions*

2.5 Key Takeaways

After comprehensively reviewing the diversity of datasets available on the subject matter and understanding the motivation and role that each dataset plays, this research has identified the PlantVillage dataset as the most suitable choice due to its extensive variety of labelled plant disease images across a diverse range of plant species, among the fact that it is commonly used in plant disease research. This allows us to effectively train and evaluate deep learning models for a broader range of plant disease detection tasks.

Furthermore, through the review of recent and local research on the subject matter a research gap has been identified in the study of the classification of all plant classes as well as in the utilisation of a key deep learning algorithm, EfficientNet, which is believed to be the more advanced and current state-of-the-art as per Atila et al. [3].

Building upon the insights gained from the incremental experimentation approach employed by Mohanty et al. [17], the next chapter will delve deeper into the research methodology adopted in this study. This approach allows for an evaluation of different deep learning models and facilitates the identification of the most effective configuration for plant disease classification. The evaluation metrics established in relevant research, such as accuracy, precision, recall, and F1-score, will be used to assess the performance of the proposed models and compare them to existing research findings.

Chapter 3: Research Methodology

3.1 Philosophical Positioning

In aligning with the foundational principles of research, it is imperative to set forth the philosophical positioning that underpins this study. Philosophical positioning serves as the cornerstone upon which the methodology, data collection, and analysis strategies are built. In this section, the axiology, ontology, and epistemology [23] stances are discussed to explain the fundamental perspectives of this research and ultimately establish the positioning of objectivism and subjectivism.

Axiology addresses the researcher's values and ethical considerations guiding the research process [23]. In this study, the values of accuracy, efficiency, and ethical responsibility are important. Accuracy ensures the reliability of the developed deep learning models in detecting and classifying plant diseases, thereby minimising misdiagnoses and facilitating effective disease management. Efficiency is essential to optimise computational resources and enable timely interventions in agricultural practices. Ethical responsibility encompasses considerations such as data privacy, algorithmic bias, and equitable access to AI-driven technologies, ensuring that the research prioritises societal well-being and environmental sustainability. The principal researcher in this study aims to pursue a path as a software engineer and sees this study as an opportunity to enhance their expertise in machine learning. Thus, the research questions and focus of this study centre around the machine learning pipeline and rigorous experimentation.

Ontology pertains to the researcher's understanding of the nature of reality within the agricultural domain [23]. This study operates under the assumption of an objective reality comprising various plant diseases, such as those of living organisms and the manifestations in agricultural ecosystems. Plant diseases are a physical phenomena that can be identified through computer vision techniques because they have distinguishable features. Thus, to effectively study and manage plant diseases, it is essential to adopt holistic approaches that consider the interactions between plants, pathogens, environmental conditions, and human interventions.

Epistemology addresses how knowledge is acquired, validated, and disseminated within the research domain [23]. Given the subject matter being investigated, that of plant species and plant disease identification, and the utilisation of a predefined dataset, this study incorporates objectivism. The effectiveness of the proposed solution is being evaluated with predefined formulas and compared with results obtained by other authors.

3.2 Research Aim, Hypothesis and Questions

Given the outlines of the philosophical position of the principal researcher and the subject matter being investigated, the research aim, hypothesis, and questions considered in this study are outlined hereunder.

The aim of this study is to develop an efficient and accurate plant disease detection system using deep learning techniques and to investigate how such technologies can benefit the local agricultural sector. By leveraging advancements in artificial intelligence and computer vision, this research aims to contribute to the advancement of precision agriculture practices within the local agricultural sector.

This research hypothesises that recent advancements in deep learning models, particularly CNNs, permit a more holistic and accurate detection and classification of various plant diseases.

To address the research aim and hypothesis, the following research questions have been formulated:

1. How effectively can established deep learning methodologies from key research papers be replicated for plant disease detection? What insights can be gained from this process regarding the strengths and limitations of existing approaches?

2. Can computer vision techniques accurately identify key characteristics that differentiate healthy plants from those infected with various diseases?

3. How can an experimentation plan aid in the identification of model architecture variations for the advancement of plant disease detection?

4. How do variations in CNN architecture, such as the number of layers, filter sizes, and pooling strategies, influence the accuracy of plant disease detection?

5. What impact do different hyperparameters, including learning rate and batch size, have on the performance of deep-learning models for plant disease detection?

6. How does the proposed solution perform on the detection and recognition of key and important local crops?

These research questions serve as guiding inquiries to explore and evaluate the effectiveness of deep learning techniques in the context of plant disease detection. By addressing these questions, this study will evaluate the effectiveness of deep learning in plant disease detection, explore potential improvements, and ultimately contribute to the development of a robust and practical solution for the local agricultural sector.

3.3 Pipeline and Experiments

Multifaceted in its approach, the research consists of replication experiments and a custom deep learning algorithm exploration. For the replication experiments, the methodologies of Atila et al. [3] and Mohameth et al. [16] were thoroughly analysed with a focus on understanding and replicating the approaches conducted by the researchers. The replication experiments involve altering the seed once precise values have been recorded, followed by averaging the results from a set number of runs. The results of the replicated experiments were analysed by comparing the performance metrics, such as accuracy, precision, and recall, as documented further on in Section 3.4.3, with those reported in the original papers.

In addition, the custom deep learning algorithm consists of hyperparameter tuning by adjusting key parameters like learning rate, batch size, layer count, filter sizes, and pooling strategies. Each parameter undergoes systematic variation, while others remain constant, enabling the identification of optimal configurations through the analysis of performance metrics. This analysis involves assessing the impact of different hyperparameters on model performance and identifying optimal configurations for accurate plant disease detection.

3.3.1 Calibration Phase

The calibration phase of this study involves replicating methodologies from two key papers, Atila et al. [3] and Mohameth et al. [16], laying the groundwork for subsequent experimentation. These two papers serve as benchmarks for evaluating the effectiveness and feasibility of existing deep learning approaches in plant disease detection, specifically EfficientNet and VGG16. By replicating the methodologies outlined in these papers, the research aims to gain insights into the strengths and limitations of established techniques, thereby informing the development of a refined pipeline for plant disease detection.

Experiment 1

The following two experiments were conducted to address research question 1: *How effectively can established deep learning methodologies from key research papers be replicated for plant disease detection? What insights can be gained from this process regarding the strengths and limitations of existing approaches?*

The experimental setup for replicating the experiment conducted by Atila et al. [3] involved several key parameters. Firstly, the original image size of 256×256 pixels was reduced to 132×132 pixels to match the specifications outlined in the paper. This resizing step aimed to reduce computational complexity while preserving the essential features of the plant diseases. Additionally, the dataset was split into training, validation, and testing sets using a split ratio of 90%, 7%, and 3%, respectively. This split ensured that the model was trained on the majority of the data while allocating a small portion for validation and testing purposes. In addition, the learning rate was set to 0.001, and the optimizer used was Adam. The training process spanned 20 epochs, with early stopping implemented based on a patience value of 5 epochs. Lastly, the choice of model architecture was based on the findings of the original study, which utilised the EfficientNetB4 model as it achieved the highest performance for the researchers. These parameters were selected to replicate the experimental conditions outlined in the paper and facilitate a comparative evaluation of the proposed approach.

Experiment Parameter	**Atila et al. [3]**
Image Size	132×132
Dataset Split Ratio	90% - 7% - 3%
Model	EfficientNetB4
Learning Rate	0.001
Optimizer	Adam
Epochs	20
Patience	5

Table 3.1: *Summary of the Atila et al. [3] replication parameters*

Experiment 2

On the other hand, the steps required for replicating the experiment conducted by Mohameth et al. [16] involved resizing the original image size from 256×256 pixels to 224×224 pixels to align with the specifications detailed in the paper. This resizing aimed to standardise the input dimensions for the neural network model. Additionally, the learning rate parameter was set to 0.0001, consistent with the values reported in the original study. The dataset was divided using a split ratio of 70%, 15%, and 15%, training, validation, and testing, respectively. Furthermore, the learning rate was set to 0.0001, and the optimizer used was RMSprop. The training process spanned 25 epochs, with early stopping implemented based on a patience value of 5 epochs. Lastly, the choice of model architecture was based on the comparative analysis conducted by the researchers, which found VGG16 to achieve better results than GoogLeNet and ResNet50. Therefore, VGG16 was selected as the preferred architecture for its performance. These adjustments were chosen to replicate the experimental conditions outlined in the paper for a meaningful comparison with the proposed approach.

Experiment Parameter	**Mohameth et al. [16]**
Image Size	224×224
Dataset Split Ratio	70% - 15% - 15%
Model	VGG16
Learning Rate	0.0001
Optimizer	RMSprop
Epochs	25
Patience	5

Table 3.2: *Summary of the Mohameth et al. [16] replication parameters*

3.3.2 Novel Phase

In the novel phase, a refined pipeline for plant disease detection is proposed, offering a novel perspective on the application of deep learning in agricultural settings. Building upon the insights gained from the calibration phase, this research aims to develop an innovative approach that leverages advancements in computer vision techniques and deep learning architectures. The novel pipeline integrates state-of-the-art methodologies tailored specifically for the detection and classification of various plant diseases, thereby addressing the limitations of existing approaches and offering a more robust solution for agricultural practitioners.

Research Pipeline

As illustrated in Figure 3.1, a four-stage plan was devised to outline the pipeline of the proposed research. In the initial stage, two primary sources of data are utilised: the PlantVillage dataset [12] and open-source images. These datasets are analysed to verify annotations and labels, thereby laying the foundation for subsequent model training. The second stage involves the partitioning of the dataset into training and testing sets, followed by architecture definition and model training. In the third stage, various experimental setups are explored, encompassing the testing of custom models with different sample ratios, activation layers, seed and learning rates, and datasets. Finally, the fourth stage encompasses the computation of evaluation metrics, the generation and evaluation of confusion matrices, and the analysis of test results. Through these phases, the study aims to develop an effective plant disease detection system and provide insights into the performance and robustness of different experimental configurations.

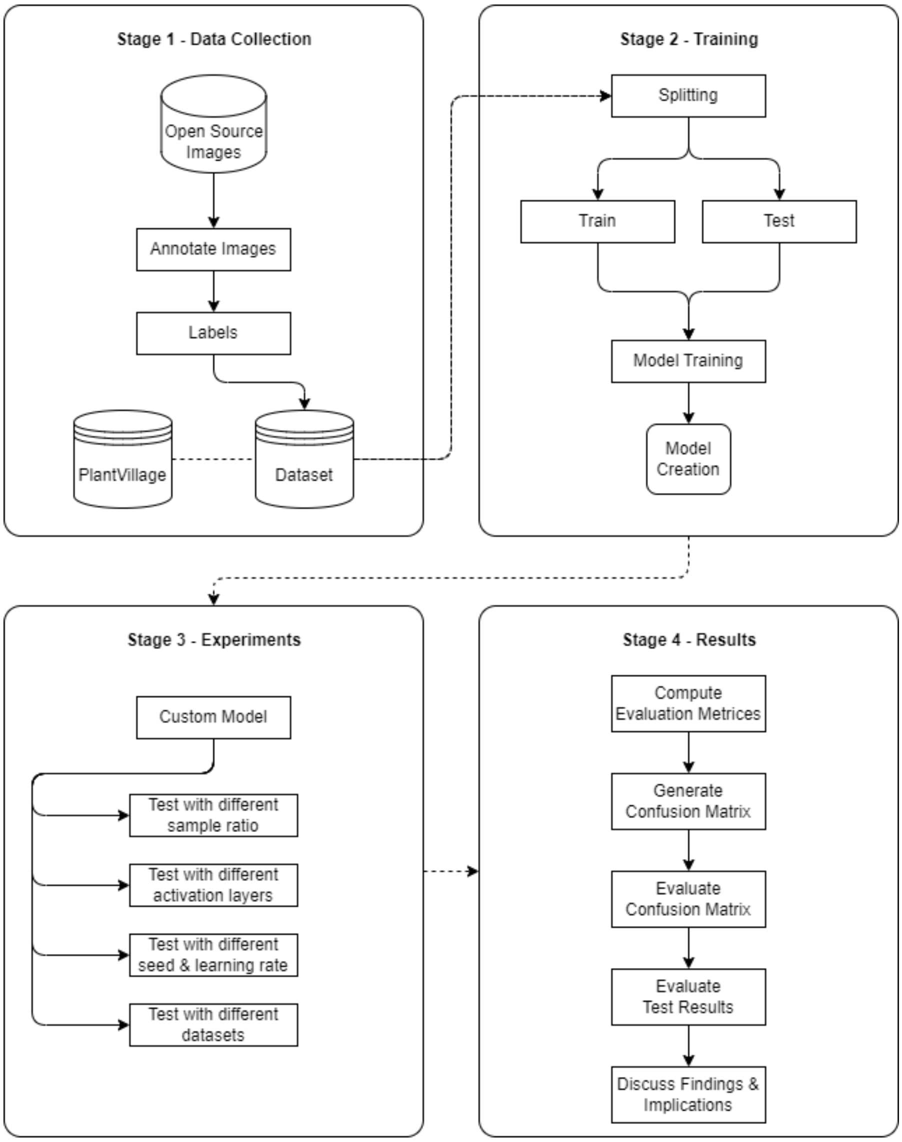

Figure 3.1: *Research Pipeline*

Experiment 3

For the third experiment it was aimed to address research question 4 and 5: *How do variations in CNN architecture, such as the number of layers, filter sizes, and pooling strategies, influence the accuracy of plant disease detection?* and *What impact do different hyperparameters, including learning rate and batch size, have on the performance of deep-learning models for plant disease detection?*

For this experiment, a custom deep learning algorithm was developed, incorporating several key specifications. The input images were retained as the original size of the images from the dataset, which is 256×256 pixels, without resizing. The dataset was split into training, validation, and testing sets with a ratio of 80%, 10%, and 10%, respectively. The learning rate was set to 0.001, and the activation functions used were ReLU for the first and second layers and softmax for the final layer. The optimizer employed was Adam, with a patience of 5 epochs and a total of 25 epochs for training.

The custom deep learning algorithm consisted of a sequential model architecture created with Keras. Firstly, a rescaling layer is used to normalise the pixel values of the input images. The initial stage included two convolutional layers, each with 32 filters and a kernel size of $(3,3)$. These were followed by a max-pooling layer with a $(2,2)$ pool size to reduce the feature map dimensionality.

Subsequently, the 2D feature maps were flattened into a 1D vector. This flattened vector was then passed through a dense fully connected layer containing 128 units with ReLU activation. The final step was the output layer, were a softmax activation function was applied to predict class probabilities, with the number of units corresponding to the desired output classes. This architecture enabled the model to process images, extract relevant features through convolution and pooling, and make predictions based on the dense and softmax layers.

This experiment systematically varied each parameter while keeping others constant, enabling the identification of optimal configurations through the analysis of performance metrics. This approach allowed for a comprehensive exploration of the impact of different architectural and hyperparameter choices on the accuracy and effectiveness of the deep learning model for plant disease detection.

Experiment Parameter	Custom Algorithm
Image Size	224×224
Dataset Split Ratio	80% - 10% - 10%
Learning Rate	0.001
Optimizer	Adam
Epochs	25
Patience	5

Table 3.3: *Summary of the custom deep learning algorithm parameters*

3.3.3 Real-World Application Phase

The real-world application phase of this study involves testing the developed system on a custom dataset, thereby validating its effectiveness and practical utility in real-world agricultural scenarios specific to Malta. The custom dataset, crucially relevant to Malta's agricultural landscape, comprises images of locally prevalent crops such as potatoes, strawberries, and tomatoes, along with an associated disease. Despite the relevance of these crops to Malta, due to resource limitations, the images were gathered from open-source sources. This dataset represents the diversity and complexity of agricultural ecosystems in Malta, encompassing both healthy and diseased instances of each crop. By evaluating the performance of the developed system on this custom dataset, insights are gathered.

The selection of potato, strawberry, and tomato as crops is based on their significant economic importance in Malta's agricultural landscape. According to the Census of Agriculture 2020 [20], these crops are among the most cultivated crops, indicating their relevance to local farming practices and the economy. Additionally, the National Agricultural Policy for the Maltese Islands [15] highlights the need for strategic interventions and support for such crops to enhance productivity, sustainability, and market competitiveness. By focusing on these crops, the study aims to address existing weaknesses in Malta's agricultural sector while exploring avenues for innovation and market responsiveness.

Experiment 4

The fourth experiment aimed to address research question 6: *How does the proposed solution perform on the detection and recognition of key and important local crops?*. This experiment involved the utilisation of a custom dataset tailored to reflect prevalent crops and associated diseases in Malta. This dataset consisted of three key crops: potatoes, strawberries, and tomatoes, making up six distinct classes. The original images had a size of 640x640 pixels. However, due to computational efficiency considerations, the images were resized to a uniform size of 384x384 pixels to standardise input dimensions while maintaining sufficient detail for accurate classification.

The dataset was split into training, validation, and test sets with a ratio of 70%, 15%, and 15%, respectively, ensuring an adequate distribution of data for training and evaluation purposes. A learning rate of 0.001 was employed with the Adam optimizer to facilitate gradient descent during training. The model was trained over 25 epochs, with early stopping implemented based on a patience parameter of 5 epochs.

These parameters were found by systematically adjusting the parameters one by one to assess their impact on model performance. Such a method allowed for the exploration of different configurations and their effects on the model's ability to classify crops and detect associated diseases. By systematically changing parameters and evaluating the model's performance, insights were gained into the optimal configuration for achieving the desired balance between computational efficiency and classification accuracy.

Experiment Parameter	**Custom Algorithm**
Image Size	384×384
Dataset Split Ratio	70% - 15% - 15%
Learning Rate	0.001
Optimizer	Adam
Epochs	25
Patience	5

Table 3.4: *Summary of the Custom dataset algorithm parameters*

3.4 Datasets and Evaluation

This section presents the datasets used in the study and outlines the evaluation metrics employed to assess the performance of the developed plant disease detection system.

3.4.1 PlantVillage Dataset

The PlantVillage dataset [12] is a widely recognised benchmark dataset for plant disease detection tasks. It consists of 54,303 high-resolution images of various healthy and diseased crop leaves, categorised into 38 classes. A summary of each species is found in Table 3.5, with a preview of the dataset in Figure 3.2. This dataset provides a valuable foundation for replicating established deep learning methodologies and establishing a baseline performance during the calibration phase.

Plant Name	Species Diseases	Species Healthy	No. of Images
Apple	3	1	3,171
Blueberry	0	1	1,502
Cherry	1	1	1,905
Corn	3	1	3,852
Grape	3	1	4,062
Orange	1	0	5,507
Peach	1	1	2,657
Bell Pepper	1	1	2,473
Potato	2	1	2,152
Raspberry	0	1	371
Soybean	0	1	5,089
Squash	1	0	1,835
Strawberry	1	1	1,565
Tomato	9	1	18,159

Table 3.5: *Summary of the PlantVillage dataset [12]*

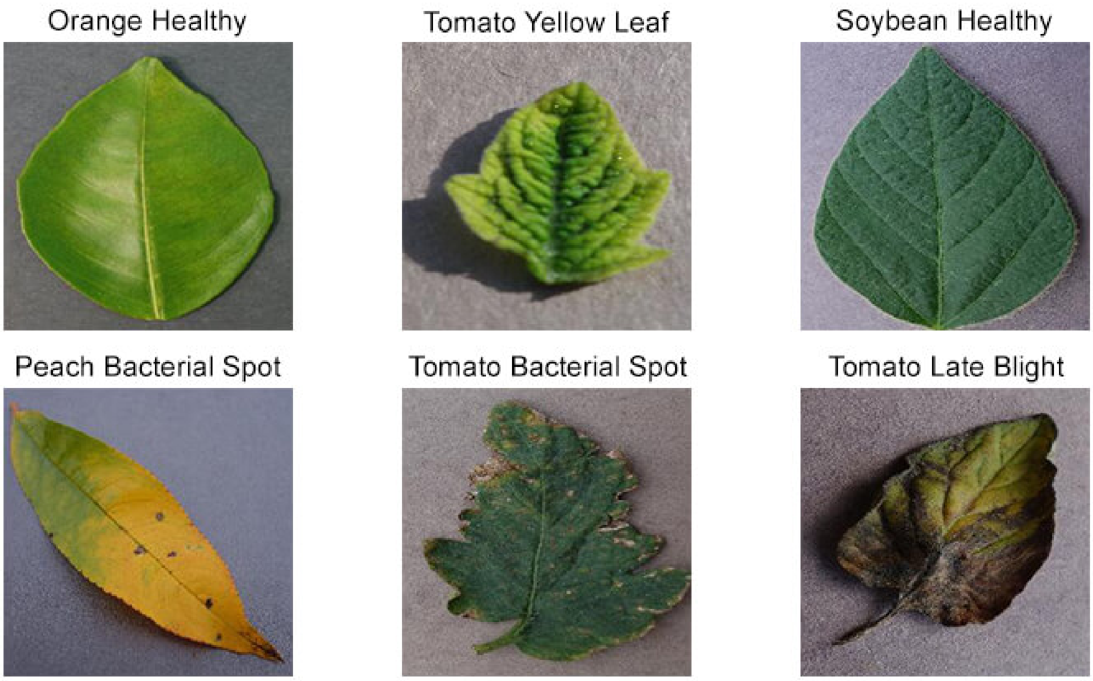

Figure 3.2: *Sample images of 6 classes within the PlantVillage dataset [12]*

In addition, a further breakdown of each class is shown in Table 3.6.

Class	Plant Name	Healthy or Diseased	Disease Name	Type of Disease	No. of Images
C_1	Apple	Diseased	Apple Scab	Fungal	630
C_2	Apple	Diseased	Black Rot	Fungal	621
C_3	Apple	Diseased	Cedar Rust	Fungal	275
C_4	Apple	Healthy	-	-	1,645
C_5	Blueberry	Healthy	-	-	1,502
C_6	Cherry	Diseased	Powdery Mildew	Fungal	1,052
C_7	Cherry	Healthy	-	-	853
C_8	Corn	Diseased	Cercospora Leaf Spot	Fungal	513
C_9	Corn	Diseased	Common Rust	Fungal	1,192
C_10	Corn	Diseased	Northern Leaf Blight	Fungal	985
C_11	Corn	Healthy	-	-	1,162
C_12	Grape	Diseased	Black Rot	Fungal	1,180
C_13	Grape	Diseased	Esca (Black Measles)	Fungal	1,383
C_14	Grape	Diseased	Leaf Blight (Isariopsis)	Fungal	1,076
C_15	Grape	Healthy	-	-	423
C_16	Orange	Healthy	-	-	5,507
C_17	Peach	Diseased	Bacterial Spot	Bacterial	2,297
C_18	Peach	Healthy	-	-	360
C_19	Pepper Bell	Diseased	Bacterial Spot	Bacterial	997
C_20	Pepper Bell	Healthy	-	-	1,476
C_21	Potato	Diseased	Early Blight	Fungal	1,000
C_22	Potato	Diseased	Late Blight	Fungal	1,000
C_23	Potato	Healthy	-	-	152
C_24	Raspberry	Healthy	-	-	371
C_25	Soybean	Healthy	-	-	5,089
C_26	Squash	Diseased	Powdery Mildew	Fungal	1,835
C_27	Strawberry	Diseased	Leaf Scorch	Fungal	1,109
C_28	Strawberry	Healthy	-	-	456
C_29	Tomato	Diseased	Bacterial Spot	Bacterial	2,127
C_30	Tomato	Diseased	Early Blight	Fungal	1,000
C_31	Tomato	Diseased	Late Blight	Fungal	1,909
C_32	Tomato	Diseased	Leaf Mold	Fungal	952
C_33	Tomato	Diseased	Septoria Spot	Fungal	1,771
C_34	Tomato	Diseased	Spider mites	Pest	1,676
C_35	Tomato	Diseased	Target Spot	Fungal	1,404
C_36	Tomato	Diseased	Mosaic Virus	Viral	373
C_37	Tomato	Diseased	Yellow Leaf	Viral	5,357
C_38	Tomato	Healthy	-	-	1,590

Table 3.6: *Class breakdown of the PlantVillage dataset [12]*

3.4.2 Custom Dataset

In addition to the PlantVillage dataset [12], a custom balanced dataset tailored to the local agricultural context was created. This dataset consists of images sourced from open-access repositories, specifically focusing on the three key crops in Malta: potatoes, strawberries, and tomatoes. Each crop category includes both healthy and diseased instances, resulting in a total of six distinct classes. A preview of these six classes can be seen in Figure 3.3. The dataset consists of 80 images for each class, amounting to a collection of 480 images.

The selection process for these images involved criteria aimed at ensuring the dataset's quality and relevance. Primarily, images were chosen with a focus on capturing the primary leaf of each crop. Given that the images were not obtained under controlled laboratory conditions, unlike the PlantVillage dataset [12], emphasis was placed on selecting images where the primary leaf was predominantly featured with minimal interference from the surrounding foliage. This approach aimed to minimise potential inaccuracies or biases in the classification process.

During the creation of this dataset, a process of validation and verification was undertaken to ensure the validity and accuracy of the data. This involved evaluating the images to verify that they accurately represented the specified crops and diseases. Additionally, the dataset underwent thorough validation to ensure its reliability and representativeness. Table 3.7 documents the acceptance and rejection criteria used during this validation process.

Class	Plant Name	Healthy or Diseased	Disease Name	Type of Disease	No. of Images
C_1	Potato	Diseased	Late Blight	Fungal	80
C_2	Potato	Healthy	-	-	80
C_3	Strawberry	Diseased	Leaf Scorch	Fungal	80
C_4	Strawberry	Healthy	-	-	80
C_5	Tomato	Diseased	Yellow Leaf	Viral	80
C_6	Tomato	Healthy	-	-	80

Table 3.7: *Class breakdown of the Custom dataset*

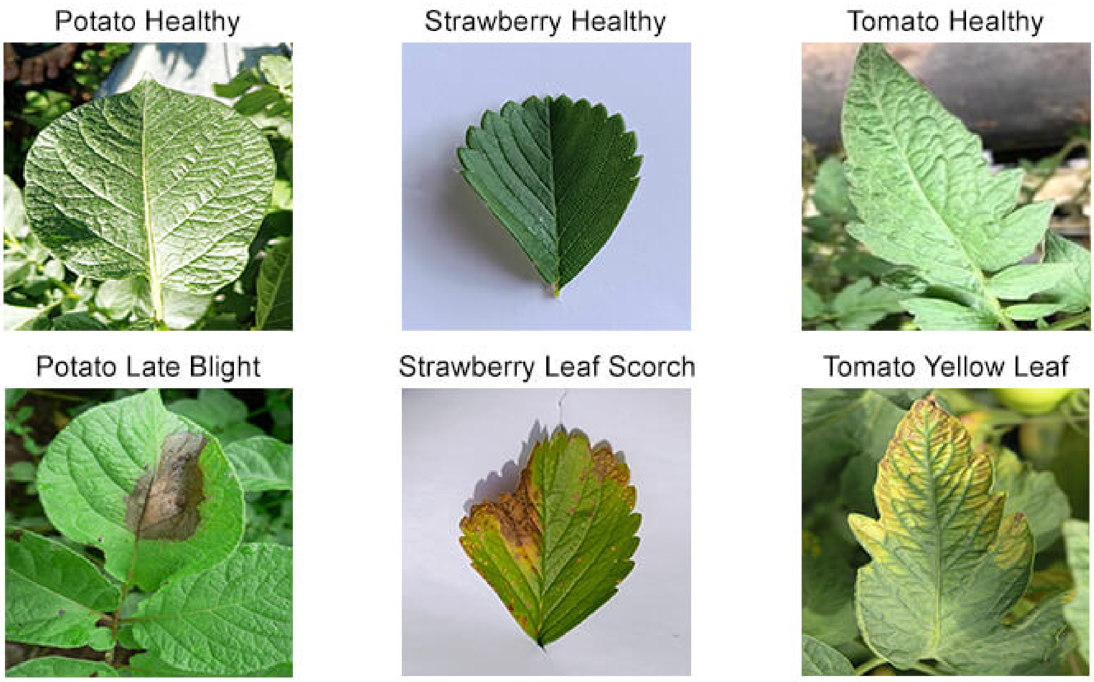

Figure 3.3: *Sample images of 6 classes within the Custom dataset*

3.4.3 Evaluation Metrics

The majority of the research was evaluated using accuracy, precision, recall, and F1-Score. The evaluation using accuracy is particularly important as it helps to gauge the overall correctness of the predictions, which is crucial in various applications and domains. Four types of outcomes could occur, being True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

True positives occur when an observation from the target class is correctly identified, while true negatives represent accurate predictions of observations not belonging to the target class. False positives occur when an observation is mistakenly classified as belonging to the target class when it does not, while false negatives happen when an observation from the target class is incorrectly classified as not belonging to it. The confusion matrix provides a tabular representation of these outcomes, facilitating a detailed analysis of the model's performance.

Accuracy

Equation 3.1 defines accuracy which refers to the accurate predictions over the total number of predictions.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.1}$$

Precision

Precision expressed by Equation 3.2 represents the ratio of true positives to all the examples predicted to belong to a specific class.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{3.2}$$

Recall

Recall denotes the ratio of examples predicted to belong to a certain class to all the examples that truly belong to that class, Equation 3.3.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3.3}$$

F1-Score

F1-Score combines precision and recall into a single metric defined by Equation 3.4. This measures the harmonic mean of precision and recall, providing a balanced assessment of model's performance.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision + Recall}} \tag{3.4}$$

3.5 Software and Hardware

The custom deep learning algorithm utilised and experiments conducted were trained and evaluated using an Nvidia GeForce GTX 1080 GPU with 8GB of dedicated memory, an Intel i7-4790 CPU, and 16GB of RAM. The training process leveraged NVIDIA CUDA[1] Development 11.2, cuDNN 8.1, and TensorFlow[2] 2.10.0 for enhanced performance. Additionally, Anaconda 3[3] 2022.10 was used to create a virtual environment with Python[4] 3.10.9 to manage and install all required packages. Several packages, such as Pandas, Keras, and scikit-learn, were installed. These libraries facilitated the implementation of the deep learning models and the analysis of experimental results.

[1] https://developer.nvidia.com/cudnn
[2] https://tensorflow.org
[3] https://anaconda.com
[4] https://python.org

3.6 Ethical Considerations

This research adheres to the ethical principles outlined by Tracy [28], which encompass procedural ethics, situational ethics, and relational ethics.

Procedural ethics ensure the protection and respect of research principles. While this study does not involve direct interaction with human subjects or personal data collection, principles such as voluntary participation, informed consent, and confidentiality are upheld. As such, ethical standards are maintained in the data handling and algorithm development processes to ensure privacy and integrity. All 3^{rd} party data and software utilised are properly and adequately documented and referenced.

Situational ethics guides ethical decision-making in response to specific circumstances. This study utilises public plant image datasets. These images do not contain any visual reference to business or individuals and thus do not pose any harm of any kind. Additionally, in the creation of the local dataset, the researcher undertook a selection of the produce that are key to the local context. Any omission of specific crops is not intended to diminish in any way the importance of said crop or the related business around it.

Relational ethics emphasises mutual respect, dignity, and reciprocity between researchers and participants. While this research primarily involves data analysis and algorithm development, ethical self-consciousness remains paramount. The researcher commits to respecting the communities and sources from which data is derived, ensuring that the work contributes positively and responsibly to the broader research community.

While the research does not involve conventional data collection methods, such as interviews, surveys, or questionnaires, nor does it pose risks to individuals or communities, it is still underpinned by ethical considerations. Adhering to principles of procedural, situational, and relational ethics ensures the integrity, credibility, and ethical conduct of the research endeavour. This research, at the proposal phase, was evaluated and approved by the Ethics Review Committee within the respective academic institution.

Additionally, it should be noted that limited use of AI was used to elevate the quality of the written text. However, all content was initially written by the primary researcher, ensuring that the intellectual contributions and original ideas remain authentic and unaltered.

Chapter 4: Analysis of Results and Discussion

This chapter aims to present the results achieved in the four experiments outlined in Chapter 3 while offering an in-depth analysis and discussion of each experiment. Comparisons are drawn with the key studies from Chapter 2 to highlight advancements. In alignment with the literature, performance metrics such as accuracy, precision, recall, and F1-score were utilised to evaluate the models. Additionally, this chapter will assess the strengths and limitations of each approach in detecting plant diseases, emphasising how these findings address the original research questions and contribute to the field of PA.

4.1 Calibration Experiments

This section presents the results obtained from replicating the methodology outlined by Atila et al. [3] and Mohameth et al. [16] in their papers. The experiments carried out aimed to reproduce their approach using the same dataset, model architecture, and training parameters. This phase aims to address research question 1.

4.1.1 Replication of Atila et al.

The first experiment aimed to assess the effectiveness of replicating the EfficientNet model architecture proposed by Atila et al [3]. The study provides a comprehensive evaluation of EfficientNet in comparison to other state-of-the-art CNN models. Using the PlantVillage dataset and their specified architecture, training parameters, and dataset split ratios, it is sought to confirm the validity of their findings while noting any limitations.

To maintain alignment with the original study, the EfficientNetB4 model was chosen for replication due to its high performance in Atila et al.'s experiments. Images were resized to 132×132 pixels, and the dataset was split into training, validation, and testing sets according to a 90%, 7%, and 3% ratio, which could introduce limitations in terms of the model's ability to generalise to new, unseen data due to the small size of the test set. The learning rate was set to 0.001 with an Adam optimizer, and early stopping was based on patience of five epochs, aligning with the original study's setup. Table 4.1 shows the results comparison of this study with that of Atila et al.

As previously mentioned, the choice of a 90%, 7%, and 3% ratio, while allowing for a large training set, does present a limitation. It restricts the volume of data available for validation and testing, which could skew performance metrics and potentially overfit the model to the training data. Future studies might consider employing cross-validation techniques instead of a fixed split ratio. Cross-validation allows for the use of varying train-test splits across different iterations, enhancing the model's evaluation against unseen data and providing a more comprehensive gauge of its generalisability.

Metric	**Atila et al. [3]**	**This Study**
Accuracy	99.84%	95.86%
Precision	97.24%	96.33%
Recall	96.82%	95.55%
F1-Score	N/A	95.94%

Table 4.1: *Results comparison with Atila et al. [3]*

The replication of Atila et al.'s [3] study highlights the effectiveness of the EfficientNetB4 model in plant disease classification, with results closely mirroring the original findings. The inclusion of the F1-score in this research provides an additional understanding of the model's performance, particularly in handling classes with varying degrees of representation in the dataset. It is worth considering that the minor disparities observed might be attributed to aspects of the original experiment that were not fully documented. For instance, subtle differences in data preprocessing, augmentation practices, or even slight variations in the computational environment might have contributed to these differences. While such factors are often overlooked, they can significantly impact the outcomes of highly sensitive deep learning models.

4.1.2 Replication of Mohameth et al.

The second experiment aimed to evaluate the effectiveness of the VGG16 architecture in plant disease detection, following the methodologies outlined by Mohameth et al [16]. The study involved the assessment of several CNN models on the Plant Village dataset, with a specific focus on deep feature extraction and transfer learning techniques. The choice of VGG16 is based on its effectiveness in previous studies, especially as it ranked the second highest performing model in the research conducted by Atila et al [3]. The replication followed the experimental setup detailed by Mohameth et al. [16].

The accuracy achieved for this replication was 95.64%, which is 2.18% lower than the 97.82% accuracy reported by Mohameth et al. However, precision and recall figures could not be directly compared, as Mohameth et al. did not provide specific values. The F1-Score was introduced to provide a balanced view between precision and recall, addressing potential class imbalances in the dataset, which was 96.42%. which is 0.76% higher than this study at 95.66%. A summary of the results comparison is shown in Table 4.2. Notably, the original study did not specify if the reported metrics were micro, macro, or weighted averages, complicating direct comparisons of these results. In this study, the macro averages were adopted, ensuring consistent and comparable metric reporting throughout.

Metric	**Mohameth et al. [16]**	**This Study**
Accuracy	97.82%	95.64%
Precision	N/A	96.64%
Recall	N/A	94.71%
F1-Score	96.42%	95.66%

Table 4.2: *Results comparison with Mohameth et al. [16]*

Mohameth et al.'s [16] original study did not detail certain methodological specifics, such as the train-test split ratio or the exact nature of the averaging method used for reporting performance metrics. For this replication, a train-test-validation split of 70%, 15%, and 15% was employed to ensure a more balanced distribution of data for training, validation, and testing. This split was chosen based on preliminary experiments to closely match the accuracy reported in the original paper. Metrics were calculated as macro averages to provide a comprehensive overview across classes, facilitating a better understanding of model performance in multi-class scenarios.

The experiment highlighted several discrepancies in the dataset description by Mohameth et al. [16]. For instance, the number of images for Apple leaves was incorrectly listed as 33,172 in the table when the actual number in the PlantVillage dataset is 3,172. Additionally, the classification of the squash crop was mistakenly reported as two distinct classes, whereas only one class exists. Such inaccuracies necessitate caution in interpreting the dataset's composition and its potential impact on the experimental outcomes.

This experiment supported the effectiveness of VGG16 as documented by Mohameth et al. [16] for plant disease classification and extended the insights by documenting additional methodological details. The experiment also highlighted the need for clearer documentation in original studies, particularly regarding data handling and metric calculations, specifically to enhance reproducibility and comparability in future research.

4.2 Novel Experiments

This section presents the results gathered from the development of the custom algorithm, designed to address research question 4 and 5.

4.2.1 Custom Algorithm

Throughout this phase, experiments were methodically conducted by varying each parameter individually while keeping others constant to isolate the effects of each modification on the model's performance. The increased learning rate in the initial tests resulted in decreased accuracy, confirming the sensitivity of CNNs to optimal learning rate settings. A lower learning rate of 0.001 was subsequently identified as optimal, striking a balance between training speed and model stability.

Various dataset split ratios were also tested to determine the optimal distribution for training, validation, and testing. Experiments included testing the original 90%, 7%, and 3% split used by Atila et al. [3], among others. It was found that a split of 80%, 10%, and 10% for training, validation, and testing, respectively, offered a balanced approach, ensuring enough data for effective training and adequate evaluation without overfitting. It's evident from this experiment, as well as supported by Bichri et al. [4] and Racz et al. [22] that higher training ratios can indeed result in improved accuracy on the test dataset. This occurs because a larger training set provides more comprehensive coverage of the variability in data, while a smaller test set might not fully challenge the model's capabilities, potentially leading to an overestimation of its performance.

The performance of the custom model was quantified by averaging results across 10 seed values to stabilise outcomes against random initialisation effects. This approach yielded an average accuracy of 81.08%, precision of 85.21%, recall of 78.27%, and an F1-score of 81.59%. These metrics were calculated using macro averaging, which treats all classes equally, providing a balanced view of model performance across the various plant diseases, irrespective of their frequency in the dataset. A summary of these results is shown in Table 4.3.

Metric	**This Study**
Accuracy	81.01%
Precision	85.21%
Recall	78.27%
F1-Score	81.59%

Table 4.3: *Results summary of the Custom model using the PlantVillage dataset*

These findings highlight the critical influence of hyperparameters like learning rate and data splits on model performance, emphasising the importance of methodical parameter tuning in developing robust deep learning models for plant disease detection. This phase of the research contributes to a deeper understanding of the refined impacts of architectural and hyperparameter choices on the effectiveness of CNN models in practical applications. These insights provide a solid basis for future explorations into more complex models or datasets.

4.3 Real-World Application Experiments

This section presents the findings from testing the custom deep learning model on a novel dataset specifically collected to reflect prevalent crop diseases in Malta. This phase aimed to assess the model's effectiveness and its adaptability to conditions that closely mimic real agricultural environments. This experiment addresses research question 6.

For this phase, the model was adapted to classify images of three key crops in Malta, potatoes, strawberries, and tomatoes affected by healthy and various diseases, consisting a total of six distinct classes. The model was initially trained using the methodologies fine-tuned in earlier experiments from the novel experiment. The model achieved an accuracy of 85.13%, a precision of 91.94%, recall of 81.82%, and an F1-score of 86.13%. Table 4.4 shows these results. These metrics indicate that the model is capable of effectively identifying plant diseases under varied real-world conditions. A detailed analysis provides further insights into the model's performance across different dataset configurations. Notably, when the dataset was simplified to binary classifications, representing just healthy and diseased crops, the model demonstrated the expected superior predictive capabilities. This aligns with findings from other studies, where CNNs have shown high accuracy in binary classification tasks [29].

Metric	**Custom Dataset**
Accuracy	85.13%
Precision	91.94%
Recall	81.82%
F1-Score	86.13%

Table 4.4: *Results summary of the six-class classification of the Custom dataset representing healthy and diseased crops*

Comparatively, the model's performance on the PlantVillage dataset is superior due to the extensive size and diversity of the dataset which includes thousands of images per class, providing a more comprehensive learning environment for the model. The contrast in performance underscores the challenge of deploying machine learning models in real-world settings, where data may not be as ideal in terms of quantity and controlled conditions. Despite achieving similar results with the custom algorithm when tested on the PlantVillage dataset, which consists of images under laboratory conditions, this does not necessarily guarantee the same performance in more variable and complex real-world environments

To visually illustrate the model's diagnostic capabilities under real-world conditions, a Gradient-weighted Class Activation Mapping (Grad-CAM) visualisation was introduced. These images serve to interpret the model's focus during the disease detection process, highlighting areas deemed significant by the neural network. Figure 4.1 and 4.2 illustrate three panels for each scenario.

The first panel displays the original image of the leaf. The second panel shows the Grad-CAM heatmap, which highlights areas of the image that most influenced the model's output—warmer colours indicate regions with higher relevance to the model's decision. The third panel presents a superimposed view, merging the original image with the heatmap to contextualise the areas of focus within the natural appearance of the leaf. Figure 4.1 shows a healthy strawberry leaf against a plain background, while Figure 4.2 shows a diseased strawberry leaf among dense foliage, simulating complex environmental conditions.

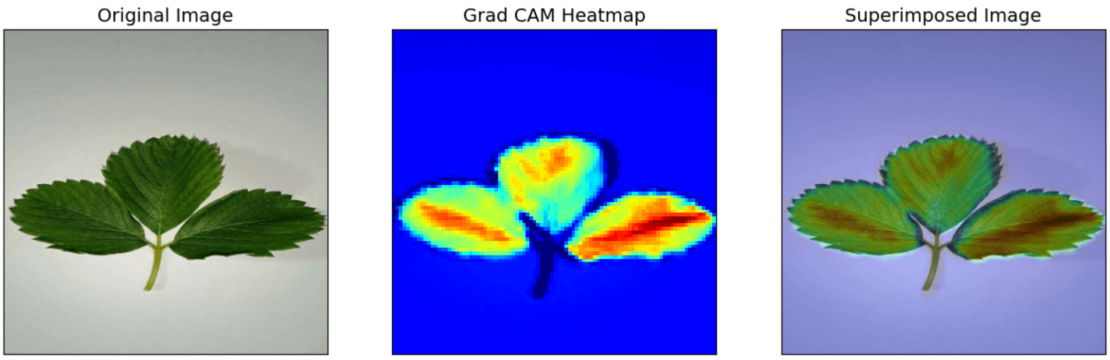

Figure 4.1: *Grad-CAM visualisation of a healthy strawberry leaf against a plain background*

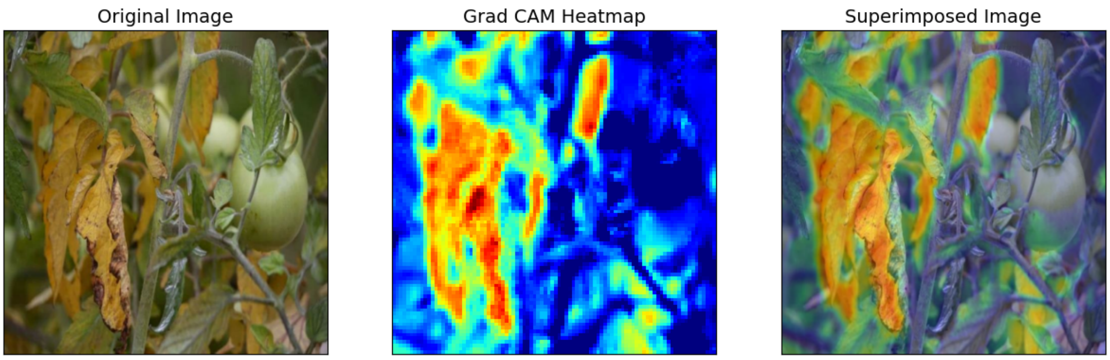

Figure 4.2: *Grad-CAM visualisation of a diseased strawberry leaf among dense foliage*

The integration of these visualisations is crucial, especially when considering the model's performance in distinguishing between healthy and diseased leaves among varying backgrounds and leaf densities. By comparing the focus areas in simple versus complex environmental conditions, it provides insights into the model's robustness and areas where further training or architectural adjustments may be necessary to improve accuracy and reliability in real-world settings.

These findings highlight the practical challenges of applying deep learning models outside of laboratory conditions, particularly in agricultural settings where data limitations are prevalent. While the model demonstrated promising results in controlled tests, its performance in real-world scenarios suggests that additional improvements are necessary. Such improvements is to include images of crops in their natural agricultural environments, rather than isolated on plain backgrounds, this could significantly enhance model performance. Such a dataset would offer a more realistic and background-diverse range of images, helping the model better generalise its detection capabilities in varied and realistic settings. This approach addresses the need for robustness and adaptability in practical applications of plant disease detection.

The insights gained from this phase are crucial for refining the model's deployment strategies in actual agricultural practices. This phase marks a significant advancement across Technology Readiness Levels (TRL), bringing the solution closer to practical use in the market. It represents a critical step towards deploying AI-driven solutions in the field, aiming to support agricultural professionals in making informed decisions based on reliable, automated disease detection tools.

4.4 Comparison with other studies

The findings from the calibration, novel, and real-world application phases are illustrated in the following section. Table 4.5 compares the performance metrics, accuracy, precision, recall, and F1-score, of the replicated methodologies across different studies.

Metric	Atila et al. [3]	Atila et al. Replication	Mohameth et al. [16]	Mohameth et al. Replication
Accuracy	99.84%	95.86%	97.82%	95.64%
Precision	97.24%	96.33%	N/A	96.64%
Recall	96.82%	95.55%	N/A	94.71%
F1-Score	N/A	95.94%	96.42%	95.66%

Table 4.5: *Results comparison of the replicated methodologies and the original studies using the PlantVillage dataset*

In addition, Table 4.6 shows the performance metrics of the model when utilising the PlantVillage dataset compared to other studies.

Metric	Atila et al. [3]	Mohameth et al. [16]	This Study
Accuracy	99.84%	97.82%	81.01%
Precision	97.24%	N/A	85.21%
Recall	96.82%	N/A	78.27%
F1-Score	N/A	96.42%	81.59%

Table 4.6: *Results comparison of the original studies and the custom model using the PlantVillage dataset*

Furthermore, Table 4.7 illustrates the performance differences of the model when tested with the PlantVillage dataset and a custom dataset.

Metric	PlantVillage Dataset	Custom Dataset
Accuracy	81.01%	85.13%
Precision	85.21%	91.94%
Recall	78.27%	81.82%
F1-Score	81.59%	86.13%

Table 4.7: *Results comparison of the experiments using the PlantVillage and Custom datasets*

4.5 Research Questions Evaluation

In the calibration phase, the primary focus was on replicating established methodologies as laid out by Atila et al. [3] and Mohameth et al. [16]. This phase addresses research question 1: ***How effectively can established deep learning methodologies from key research papers be replicated for plant disease detection? What insights can be gained from this process regarding the strengths and limitations of existing approaches?*** The replication efforts were notably successful, with the results exhibiting only a minimal difference of approximately negative 2% in performance metrics compared to the original studies. This close alignment not only confirms the reproducibility of the cited methodologies but also emphasises the robustness of these approaches when applied to similar datasets. This phase underscores the importance of precision in replicating experimental setups and provides foundational insights into the strengths and potential limitations of the established deep learning frameworks for plant disease detection.

For the novel phase experiments were structured to explore the impact of architectural modifications and hyperparameter tuning on the effectiveness of CNN models in plant disease detection. This investigation addresses research questions 4 and 5: ***How do variations in CNN architecture, such as the number of layers,***

filter sizes, and pooling strategies, influence the accuracy of plant disease detection? and ***What impact do different hyperparameters, including learning rate and batch size, have on the performance of deep-learning models for plant disease detection?*** The systematic adjustments to the architecture, including alterations in the number of layers and filter sizes, directly correlated with noticeable improvements in model performance, achieving an optimised accuracy of 81.08% and precision of 85.21%. These results demonstrate the critical role of careful architectural tuning in enhancing model accuracy and handling more complex disease classification tasks.

Additionally, research question 3: ***How can an experimentation plan aid in the identification of model architecture variations for the advancement of plant disease detection?*** was addressed through the structured testing of various configurations. The experimentation plan proved to be invaluable, allowing for the precise identification of optimal configurations that enhanced the model's effectiveness. The plan's methodical nature facilitated a comprehensive understanding of the interaction between model architecture and performance, confirming the necessity for systematic experimentation in developing advanced detection systems.

In the real-world application phase, the model was tested against a custom-tailored dataset consisting of crop images, focusing on potatoes, strawberries, and tomatoes. This phase tackled research question 2: ***Can computer vision techniques accurately identify key characteristics that differentiate healthy plants from those infected with various diseases?*** and research question 6: ***How does the proposed solution perform on the detection and recognition of key and important local crops?***

The model achieved an accuracy of 85.13% and a precision of 91.94%. However, it struggled with the diversity of the real-world dataset, where more complex classifications resulted in a recall of 81.82%. These findings illustrate the model's robust capability in identifying disease characteristics under varied conditions, though it also highlight challenges in generalisation across multiple crop types. The results validate the effectiveness of computer vision techniques in real-world settings but also point to the need for further refinement to handle the complexities introduced by diverse agricultural environments.

4.6 Hypothesis Evaluation

This research hypothesises that recent advancements in deep learning models, particularly CNNs, permit a more holistic and accurate detection and classification of various plant diseases. The experiments conducted across the calibration, novel, and real-world application phases demonstrate that CNNs, when optimally configured and applied to suitably diverse and representative datasets, can indeed provide high accuracy in detecting and classifying plant diseases. The replication of established methodologies confirmed the robustness of CNN architectures like EfficientNetB4 and VGG16 in plant disease detection, while the novel and real-world application phases highlighted the model's adaptability and effectiveness in practical settings. These results collectively confirm that CNNs, supported by continuous advancements and appropriate training strategies, align well with the hypothesis, showcasing their potential to revolutionise plant disease detection in agricultural practices.

Chapter 5: Conclusions and Recommendations

This study has rigorously investigated the application of deep learning techniques for plant disease detection, utilising a three-phase approach consisting of calibration, novel exploration, and real-world application. Through these phases, the study has tested the robustness, adaptability, and practical utility of various CNN architectures and hyperparameters in detecting plant diseases across different datasets, including a custom dataset tailored to reflect prevalent crops in Malta, addressing the research questions in Section 4.5.

5.1 Research Process Reflection

The three-phase approach provided a structured pathway to assess both theoretical models and practical implementations. The calibration phase validated the reproducibility of existing deep learning models in plant disease detection, offering a solid foundation for further exploration. In the novel phase, systematic modifications to CNN architectures and hyperparameters yielded significant insights into optimising model performance. This phase highlighted the critical importance of architectural and hyperparameter tuning in achieving high accuracy. The real-world application phase highlighted the challenges of applying these models in uncontrolled environments, revealing the models' varying effectiveness across different crop types and disease manifestations.

Challenges in sourcing a relevant dataset that accurately represents the diverse conditions of local agriculture in Malta were significant. The reliance on a dataset primarily composed of printed materials and a limited number of real images posed constraints on the model's training and validation phases. Additionally, the computational demands of processing extensive image data and executing complex model architectures were consistently balanced against the need for computational efficiency.

5.2 Limitations

A notable limitation encountered specifically during the real-world application phase was the dependency on a custom dataset that, while inclusive of common crops found in Malta, primarily consisted of images sourced from online databases rather than directly from local fields. Despite efforts to collaborate with local agricultural entities, such as the Environment and Resources Authority (ERA) and the Agriculture Research and Innovation Hub (AGRIHUB) in Malta, and actively seeking field-specific images, the dataset did not fully capture the unique agricultural conditions or the specific manifestations of diseases in Maltese crops.

This challenge was also made up of the seasonal availability of the diseased leaves, which are not always present or detectable outside of their typical growth or infection periods. Consequently, while the dataset effectively facilitated model training and preliminary testing, the absence of locally sourced, field-specific images could limit the study's validity in directly reflecting the agricultural reality in Malta. This limitation highlights the need for a dataset that more accurately represents local crop varieties and disease conditions to enhance the model's practical applicability and reliability in Maltese agricultural settings.

5.3 Future Recommendations

To address the limitations identified in the real-world application phase, future research should prioritise the development of a more representative dataset that includes images directly sourced from Maltese fields. Collaborating with local farmers and agricultural organisations to collect and annotate images of local crops under various conditions could significantly enhance the dataset's relevance and utility. This targeted approach would ensure that the deep learning models are trained and validated on data that accurately reflect the specific plant diseases and agricultural practices prevalent in Malta.

Additionally, expanding the variety of crops and disease states within the dataset is crucial. By incorporating a broader array of local crop types and their respective disease manifestations, the models can be better tested and optimised for the unique challenges presented by Malta's agricultural landscape. Such enhancements could potentially improve the accuracy of disease detection and boost the models applicability and reliability for local farmers.

Regarding the utilisation of the PlantVillage dataset, while it has been valuable for calibration and initial testing, its generic nature and lack of specificity to Maltese agriculture limit its usefulness for novel or real-world applications. Future research should consider this as a calibration tool only and aim to develop more region-specific datasets for advanced experiments. Further exploration into more advanced neural network architectures that require less computational power and are more efficient in processing time should also be considered. These improvements could make the models more suitable for real-time applications in agricultural settings, where quick and accurate disease detection is essential.

List of References

[1] G.N. Agrios. *Introductory Plant Pathology*. Academic Press, San Diego, fifth edition edition, 2005.

[2] Tashin Ahmed and Noor Hossain Nuri Sabab. Classification and understanding of cloud structures via satellite images with efficientunet. *SN Computer Science*, 3:7, 07 2021.

[3] Ümit Atila, Murat Uçar, Kemal Akyol, and Emine Uçar. Plant leaf disease classification using efficientnet deep learning model. *Ecological Informatics*, 61:101182, 03 2021.

[4] Houda Bichri, Adil Chergui, and Mustapha Hain. Investigating the impact of train / test split ratio on the performance of pre-trained models with custom datasets. *International Journal of Advanced Computer Science and Applications*, 15(2), 2024.

[5] Jasmine Busuttil. Identifying plant diseases using yolov5. Bachelor of Science (Honours) in Software Development, 06 2022.

[6] Sebastian Candiago, Fabio Remondino, Michaela De Giglio, Marco Dubbini, and Mario Gattelli. Evaluating multispectral images and vegetation indices for precision farming applications from uav images. *Remote Sensing*, 7(4):4026–4047, 04 2015.

[7] Helen Canton. *The Europa Directory of International Organizations 2021*. The Europa Directory of International Organizations. Routledge, London, 07 2021.

[8] Siddharth Singh Chouhan, Ajay Kaul, Pratap Uday Singh, and Sanjeev Jain. A database of leaf images: Practice towards plant conservation with plant pathology, 2019.

[9] Dylan Galdes. Comparing image classification methods for tomato plant disease detection. Bachelor of Science (Honours) in Software Development, 06 2021.

[10] Camille Garcin, Alexis Joly, Pierre Bonnet, Antoine Affouard, Jean-Christophe Lombardo, Mathias Chouet, Maximilien Servajean, Titouan Lorieul, and Joseph Salmon. Pl@ntnet-300k: a plant image dataset with high label ambiguity and a long-tailed distribution. In J. Vanschoren and S. Yeung, editors, *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, volume 1, page 13. Curran Associates, Inc., 2021.

[11] Geetharamani Gopal and Arun Pandian J. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Computers & Electrical Engineering*, 76(C):323–338, 06 2019.

[12] David P. Hughes and Marcel Salathé. An open access repository of images on plant health to enable the development of mobile disease diagnostics through machine learning and crowdsourcing. *CoRR*, abs/1511.08060, 11 2015.

[13] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network, 2014.

[14] Robert Mercieca. Using convolutional neural networks to identify different plants and common diseases associated with them. Bachelor of Science (Honours) in Software Development, 05 2022.

[15] Ministry for Agriculture, Fisheries, Food and Animal Rights. National agricultural policy for the maltese islands 2018 – 2028, 2018.

[16] Faye Mohameth, Chen Bingcai, and Kane Amath Sada. Plant disease detection with deep learning and feature extraction using plant village. *Journal of Computer and Communications*, 08:10–22, 01 2020.

[17] Sharada Prasanna Mohanty, David Hughes, and Marcel Salathé. Using deep learning for image-based plant disease detection. *Frontiers in Plant Science*, 7:1419, 09 2016.

[18] David Moore, Geoffrey Robson, and Anthony Trinci. *21st Century Guidebook to Fungi*. Cambridge University Press, 2 edition, 05 2011.

[19] Mohammad Nadimi, Da-Wen Sun, and Jitendra Paliwal. Recent applications of novel laser techniques for enhancing agricultural production. *Laser Physics*, 31(5):053001, 03 2021.

[20] National Statistics Office. Census of agriculture 2020, 2020. Available online: www.nso.gov.mt.

[21] Janine Ryan. No farmers= no food= a broken society. *Farmer's Weekly*, 2024(24003):4–67, 01 2024.

[22] Anita Rácz, Dávid Bajusz, and Károly Héberger. Effect of dataset size and train/test split ratios in qsar/qspr multiclass classification. *Molecules*, 26(4):1111, 02 2021.

[23] Mark N.K Saunders, Philip Lewis, and Adrian Thornhill. *Research Methods for Business Students*. Pearson Education, 8 edition, 06 2019.

[24] Karen-Beth G. Scholthof. The disease triangle: pathogens, the environment and society. *Nature Reviews Microbiology*, 5(2):152–156, 12 2006.

[25] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, pages 1–14, 2015.

[26] Davinder Singh, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, and Nipun Batra. Plantdoc: A dataset for visual plant disease detection. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, CoDS COMAD 2020, pages 249–253, New York, NY, USA, 01 2020. Association for Computing Machinery.

[27] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *International Conference on Machine Learning*, 05 2019.

[28] Sarah Tracy. Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry*, 16:837–851, 10 2010.

[29] Jai Vardhan and Kothapalli Sai Swetha. Detection of healthy and diseased crops in drone captured images using deep learning, 2023.

[30] Melchior Xuereb. Plant diseases: Investigating how they can be detected and identified using ai. Bachelor of Science (Honours) in Software Development, 06 2020.