HYPER-PARAMETER TUNING OF RANDOM FOREST ALGORITHM FOR MAIZE LEAF DISEASE DETECTION AND CLASSIFICATION

BY

TAOFIQ OMOTOWO ALABI

24/11006

JOSEPH SARWUAN TARKA UNIVERSITY, MAKURDI

FEDERAL POLYTECHNIC STUDY CENTER, BAUCHI

DEPARTMENT OF COMPUTER SCIENCE

AUGUST, 2025

Table of Contents

CHAPTER ONE

INTRODUCTION

1.1 Introduction

Plant disease detection is crucial in agriculture because farmers must constantly assess if the produce they're harvesting is of acceptable quality (Varshney et al., 2022). It is critical to treat this seriously because it can cause major difficulties in plants, affecting product quality, quantity, or productivity. Plant diseases have had an influence on society and history around the world. Several countries such as the United States affected with great economic losses due to plant diseases which can affect any type of plant. Plant leaves are the most sensitive and display disease symptoms earliest; from the beginning of their life cycle until they're ready to be picked, the crops must be monitored for illnesses (Rossman, 2009).

Maize (*Zea mays* L.) is one of the most vital cereal crops globally, serving as a staple food for over 1.2 billion people and a critical feedstock for livestock and industrial applications (Erenstein et al., 2022). However, maize production faces significant threats from foliar diseases such as Common Rust (*Puccinia sorghi*), Gray Leaf Spot (*Cercospora zeae-maydis*), and Northern Leaf Blight (*Exserohilum turcicum*), which collectively account for yield losses of 15–70% in sub-Saharan Africa alone (Nsibo et al., 2024). Traditional disease identification relies on manual scouting by farmers or pathologists, a process that is labor-intensive, time-consuming, and often inaccurate due to the visual similarity between early-stage symptoms (Bachhal et al., 2024).

Recent advances in machine learning (ML) and computer vision have demonstrated remarkable potential for automating plant disease detection. Notably, convolutional neural networks (CNNs) have achieved high accuracy in classifying diseases using benchmark datasets like PlantVillage (Ali et al., 2024) . However, CNN-based models face practical challenges in resource-constrained agricultural settings, including high computational costs, large training datasets, and limited interpretability

(Iftikhar et al., 2024). In contrast, Random Forest (RF) algorithms offer a compelling alternative due to their computational efficiency, robustness to overfitting, and inherent feature importance analysis.

A critical gap persists in optimizing RF models for plant disease tasks. Existing studies often use default hyperparameters, despite evidence that tuning parameters like tree depth (max_depth) and feature subsets (max_features) can improve accuracy by 10–20% (Salman et al., 2024). This study addresses this gap by systematically optimizing RF hyperparameters for maize disease classification, using the PlantVillage dataset as a benchmark. The study combines; transfer learning with ResNet50 for discriminative feature extraction and bayesian optimization (via Optuna) to efficiently explore hyperparameter spaces.

1.2 Statement of the Problem

Maize leaf disease detection remains a critical challenge in sustainable agriculture, where timely and accurate diagnosis can mean the difference between a bumper harvest and catastrophic yield loss. Despite significant advances in machine learning-based plant disease classification, fundamental gaps persist in both research and practical implementation. The gap found in literature include performance limitations of conventional approaches and neglect of hyperparameter optimization in hybrid models. The study by (Panigrahi et al., 2020) uses various machine learning algorithm to detect maize leaf disease and classify but it suffers from neglect of hyperparameter optimization by using the default scikit-learn parameters and also uses the default feature extraction technique. This study intend to optimize the random forest algorithm and also use Convolutional Neural Network (CNN) for feature extraction, thereby, solving the problem found in (Panigrahi et al., 2020).

1.3 Research Aim and Objectives

Aim of the Study

The aim of the study is to developed an optimized model for maize leaf disease detection and classification using Random Forest Algorithm.

Objectives of the Study

The following objectives is to achieve the aim:

i. To collect maize leaf disease datasets from the plant village dataset

ii. To use CNN for feature extraction in the model pipeline building

iii. To perform hyperparameter tuning on Random Forest algorithm

iv. To compare the results obtained against Panigrahi et al. (2020) using evaluation metrics like accuracy, recall, precision and f1 score.

1.4 Research Questions

The following research questions are what this study seek to answer:

i. What is the effect of using CNN for feature extraction against using conventional method?

ii. Does tuning the hyperparameters of Random Forest algorithm increase its accuracy and other evaluation metrics?

iii. How does this improvement compare to existing literature (Panigrahi et al., 2020)?

1.5 Method of Data Collection

This research would get its dataset from the repository of Plant Village dataset hosted with Kaggle and Github. The Plant Village dataset is a collection of images for plant disease detection and classification organized into training, testing and validation. It contains images for numerous crops diseases.

1.6 Significance of Study

Creating an enhanced Random Forest model to classify maize leaf diseases marks a significant breakthrough in the fusion of agricultural technology and machine learning. This research goes beyond the traditional focus on accuracy to provide a practical, easy-to-understand, and efficient solution that caters to the constraints faced by real-world farmers. Through meticulous optimization of hyperparameters in a hybrid CNN-RF architecture, this study fills a crucial void in plant disease

diagnosis. While current deep learning models boast high accuracy rates, their impracticality for widespread field use due to computational requirements and lack of transparency has been a major obstacle. The impact of this research is far-reaching, offering valuable insights to both agricultural practices and machine learning techniques.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Recent advancements in the field of machine learning have brought about significant changes in the way plant diseases are detected, although there are still challenges to overcome in order to optimize models for practical use in agriculture. Researchers have primarily focused on two main approaches: deep learning methods that offer high accuracy but lack transparency and efficiency, and traditional machine learning techniques that provide interpretability but struggle with complex image patterns. Random Forest have shown promise in agricultural settings due to their robustness and ability to interpret feature importance (Panigrahi et al., 2020) . However, the performance of these methods heavily relies on the proper tuning of hyperparameters, a crucial aspect that is often overlooked. This section mention various literature related to the scope of this study.

2.2 Machine Learning Algorithm

Machine learning is a concept focused on enhancing future performance through the analysis of past experiences, specifically historical data. This field is dedicated to the development of automated learning techniques that allow algorithms to adjust and improve themselves without human interference. To tackle complex data issues, machine learning utilizes various algorithms tailored to specific problems. Experts emphasize that there is no universal algorithm that can solve all problems effectively. The choice of algorithm depends on factors such as the nature of the problem, the complexity of variables, the most suitable model, and other relevant considerations (Mahesh, 2020). Figure 2.1 shows various types of machine learning technique. Figure 2.2 shows a quick rundown of some of the most regularly used machine learning algorithms.

2.3 Classification Algorithm

In the realm of machine learning, supervised learning involves the process of teaching a model to understand the relationship between input and output data through labeled examples. By analyzing training data that contains a collection of input-output pairs, a function is derived to make predictions

(Mahesh, 2020). Some of the key supervised machine learning algorithms that focus on classification are Logistic Regression, Naïve Bayes, Perceptron, Support Vector Machine, Boosting, Decision Tree, Random Forest (RF), Neural Networks, Bayesian Networks, and more.

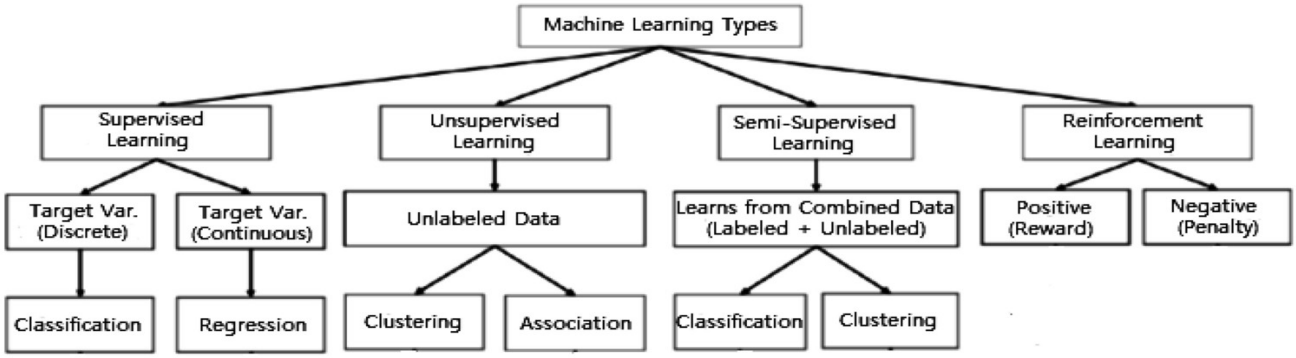

Fig 2.1: Various types of machine learning techniques (Sarker, 2021)

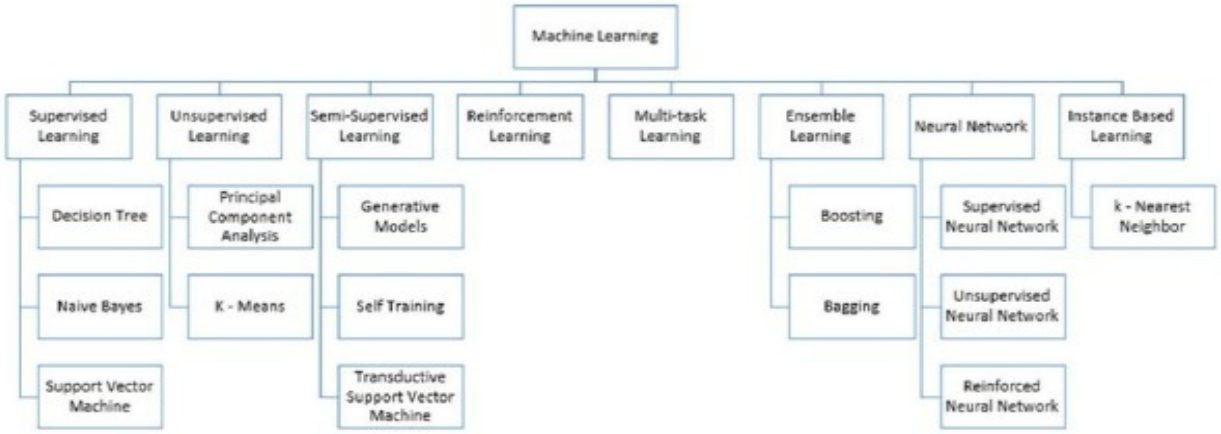

Fig 2.2: Machine Learning Algorithm (Mahesh, 2020)

2.3.1 Random Forest

This algorithm utilizes a collection of randomized decision tree classifiers as its learning methodology. During training, it generates numerous decision trees. When evaluating a testing dataset, it assigns class labels through a voting process conducted by each classification tree. The final classification is determined by the majority vote among the trees. To enhance accuracy, the algorithm incorporates bagging and feature randomness in the construction of each tree. The goal is to develop a diverse forest of trees that are uncorrelated, resulting in more precise performance predictions compared to a single tree (Panigrahi et al., 2020). A random forest is a machine learning model utilized in classification and forecasting. This technique utilizes the concept of decision trees, constructing a

collection of decision trees and aggregating their outcomes to generate the ultimate prediction. Every decision tree inside a random forest is constructed using random subsets of data, and each individual tree is trained on a portion of the whole dataset. Subsequently, the outcomes of all decision trees are amalgamated to derive the ultimate forecast (Salman et al., 2024) .

2.4 Deep Learning

Deep learning is a machine learning concept based on artificial neural networks. Deep neural networks typically consist of more than one hidden layer, organized in deeply nested network architectures. Furthermore, they usually contain advanced neurons in contrast to simple ANNs. That is, they may use advanced operations (e.g., convolutions) or multiple activations in one neuron rather than using a simple activation function. These characteristics allow deep neural networks to be fed with raw input data and automatically discover a representation that is needed for the corresponding learning task. This is the networks' core capability, which is commonly known as deep learning (Janiesch et al., 2021).

2.4.1 Convolutional Neural Network (CNN)

Before Convolutional Neural Networks gained popularity, computer recognition problems involved extracting features out of the data provided which was not adequately efficient or provided a high degree of accuracy. However in recent times, Convolutional Neural Networks have attempted to provide a higher level of efficiency and accuracy in all the fields in which it has been employed in most popular of which are Object Detection, Digit and Image Recognition. Since the beginning, the basic idea behind working of Neural Networks is that it is to mimic the working of human brain to the highest degree possible. Convolutional Neural Network contributes to this by working with the visual sensory organs of the living beings and in process recognizing various types of object be it Digit, Image or a particular action in any object using a string of various techniques followed in a particular order i.e. Convolutional Operation, ReLu Layer, Pooling, Flattening, and Softmax Cross Entropy (Ajit et al., 2020). Figure 2.3 shows the structure of CNN.

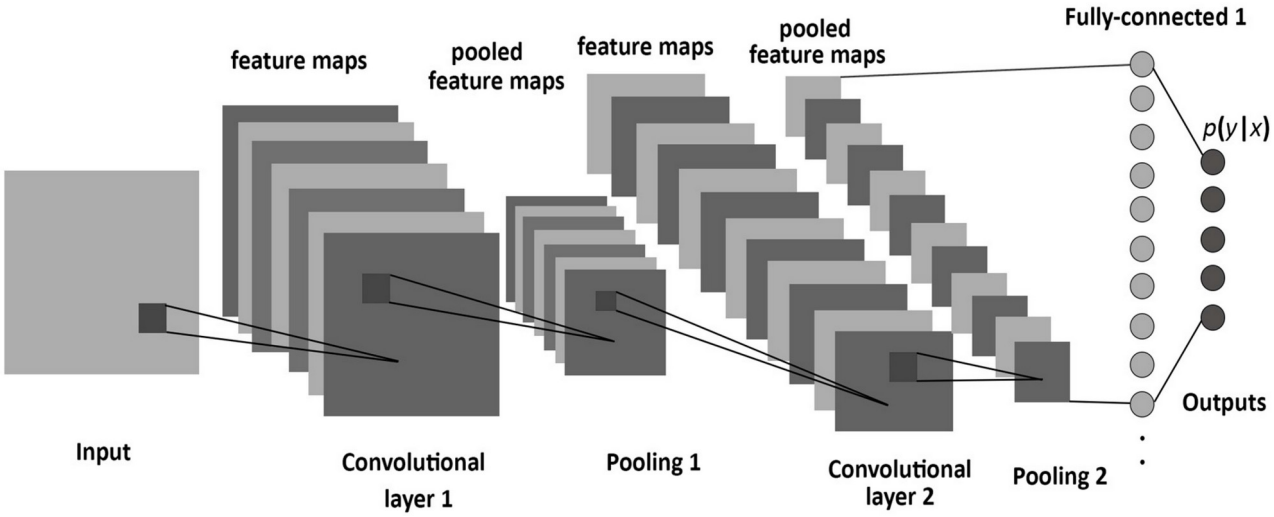

Fig 2.3: The structure of a CNN, consisting of convolutional, pooling, and fully-connected layers (Albelwi & Mahmood, 2017).

2.5 Dimensionality Reduction (DR) Technique

High-dimensional datasets present a challenge due to the presence of numerous noisy features, complicating tasks such as data processing, knowledge extraction, and pattern identification. Dimensionality reduction (DR) offers a solution by transforming the high-dimensional space into a lower-dimensional one, effectively eliminating noise and redundant information. This approach, as highlighted by Golay & Kanevski (2017), proves instrumental in addressing this issue. Various DR techniques have been developed, with feature selection (FS) and feature extraction (FE) methods standing out, as discussed by Velliangiri et al., (2019). FS, also known as variable selection or feature subset selection, involves choosing specific feature subsets for constructing models. On the other hand, FE involves creating new features based on the original ones, essentially mapping the original features to new dimensions. While FE offers more efficient feature compression, it may lead to a loss of meaning if the original feature set holds significant physical interpretations.

2.5.1 Feature Extraction

Image processing may be performed by extracting features for identification, classification, diagnosis, classification, clustering, recognition and detection. Feature extraction method are utilized to obtain much information as possible of image. The selection and effectiveness of feature chosen and extraction are a major challenge now. Many methods used to extract features, which may depend on

Geometric features, Statistical features, Texture features, and Color features. Each main type of feature divided into many subdivided types such as Color features divided into three types (Color moment, Color histogram and Average RGB) (Kavya & Harish, 2015). Figure 2.4 shows the most important features methods.

2.6 Related Work

In the work of (Singh et al., 2023), genetic algorithm was used for automatic leaf disease classification. The input image is preprocessed and segmented using a genetic algorithm to classify the diseases. The authors observed that the optimal result was obtained with a less computational cost. The author has recommended the use of fuzzy logic, ANN, and hybridization of several algorithms for the improvement of the recognition rate. Also, in a study by (Panigrahi et al., 2020), the author focuses on using supervised machine learning algorithms such as Naive Bayes, Decision Tree, K-Nearest Neighbor, Support Vector Machine, and Random Forest for maize plant disease detection and classification. The Random Forest algorithm achieved the highest accuracy of 79.23%. This very study form the basis of this research.

Another study reviews maize leaf disease classification using machine learning, highlighting methods like k-nearest neighbor, naïve bayes, decision tree, random forest, and support vector machine, emphasizing the need for image processing techniques such as preprocessing and feature extraction (Setiawan et al., 2022). In a study by (Bachhal et al., 2023), the authors reviews recent advancements in maize leaf disease detection using deep learning techniques, particularly convolutional neural networks (CNN). They also highlights the superiority of modified deep learning methods over traditional machine learning algorithms in terms of performance and accuracy.

In a study by (Vincent Mbandu Ochango et al., 2022), the authors focuses on detecting and classifying maize leaf diseases using feature extraction methods, particularly using histogram of oriented gradients, and machine learning algorithms, with the random forest classifier achieving the highest performance metrics in accuracy, precision, recall, and F1-score. Another study by (Pandey et al.,

2023), the authors presented a multilayer convolutional neural network model for maize leaf disease classification, achieving 93.28% accuracy. It automates the detection of diseases like blight, common rust, and gray leaf spot, aiding farmers in crop protection and yield improvement.

Another study by (Varshney et al., 2022) discusses the development of a new plant leaf disease detection technique using transfer learning with deep learning and SVM, achieving an 88.77% training accuracy. (Suthar et al., 2023) also evaluates various machine learning techniques, including CNNs, SVMs, random forests, and KNNs, for plant disease detection. It emphasizes the importance of feature extraction from images and demonstrates the effectiveness of these methods in accurately identifying plant diseases. (Prof. Vrushali Paithankar et al., 2023) proposes a plant disease detection system that uses image processing and convolutional neural networks (CNNs) for real-time detection of plant diseases. (Joshi & Panse, 2023) investigate the feasibility of machine learning models for early detection of plant diseases. According to the authors, machine learning techniques, particularly deep learning models like convolutional neural networks (CNNs), are highly effective for plant disease detection, achieving accuracies exceeding 90%. The study highlights their potential in identifying and diagnosing plant diseases, addressing global agricultural challenges.

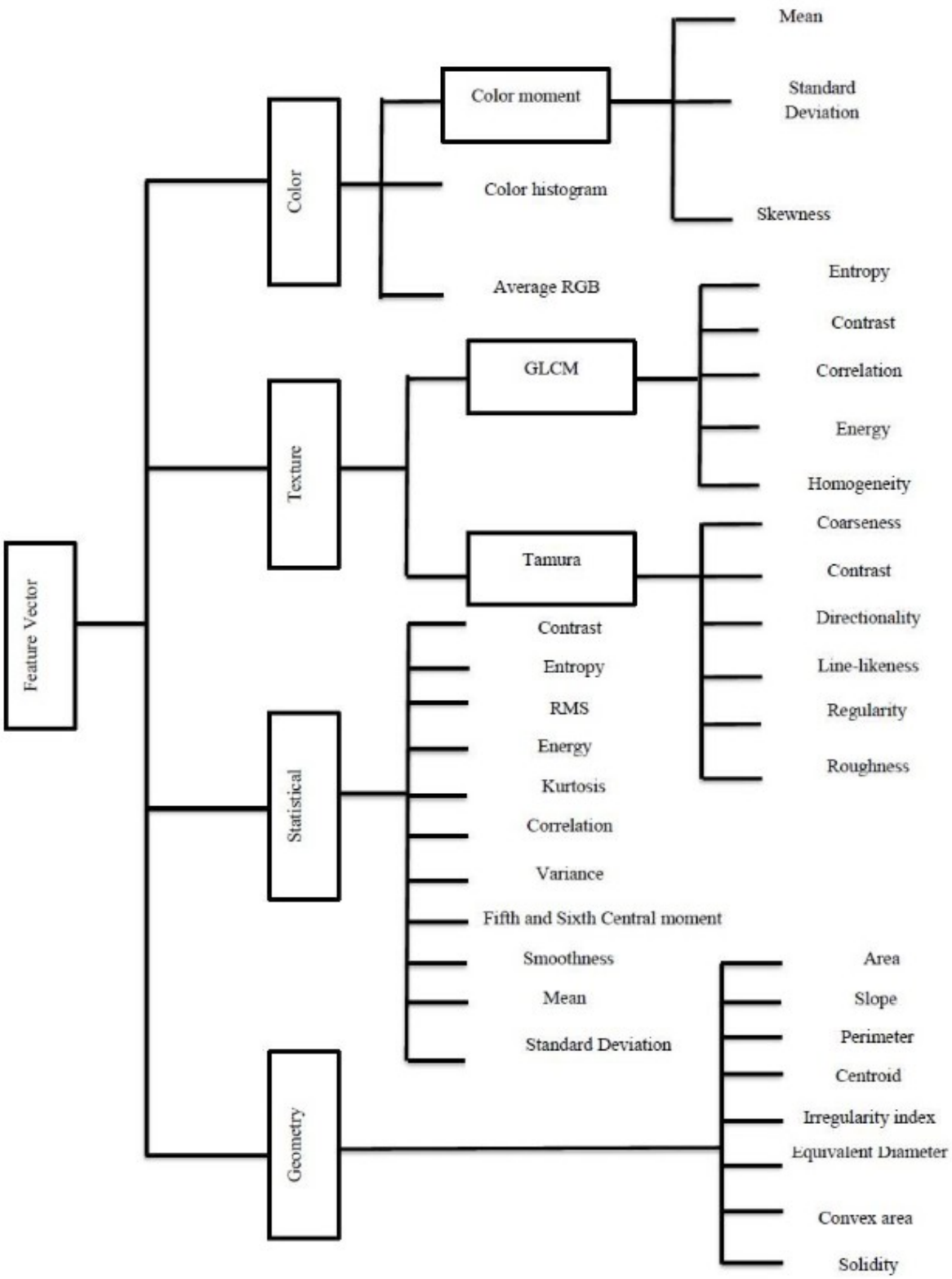

Fig 2.4: Feature Extraction Methods (Mutlag et al., 2020)

2.7 Summary of related work

S/N	Author(s)	Title	Finding(s)
1	(Panigrahi et al., 2020)	Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms	The paper focuses on using supervised machine learning algorithms such as Naive Bayes, Decision Tree, K-Nearest Neighbor, Support Vector Machine, and Random Forest for maize plant disease detection and classification. The Random Forest algorithm achieved the highest accuracy of 79.23%.
2	(Setiawan et al., 2022)	Machine Learning and Deep Learning for Maize Leaf Disease Classification: A Review	The paper reviews maize leaf disease classification using machine learning, highlighting methods like k-nearest neighbor, naïve bayes, decision tree, random forest, and support vector machine, emphasizing the need for image processing techniques such as preprocessing and feature extraction.
3	(Bachhal et al., 2023)	Maize Disease classification using Deep Learning Techniques: A Review	The paper reviews recent advancements in maize leaf disease detection using deep learning techniques, particularly convolutional neural networks (CNN). It highlights the superiority of modified deep learning methods over traditional machine learning algorithms in terms of performance and accuracy.
4	(Vincent Mbandu Ochango et al., 2022)	Feature Extraction using Histogram of Oriented Gradients for Image Classification in Maize Leaf Diseases	The paper focuses on detecting and classifying maize leaf diseases using feature extraction methods, particularly Histogram of Oriented Gradients, and machine learning algorithms, with the random forest classifier achieving the highest performance metrics in accuracy, precision, recall, and F1-score.
5	(Pandey et al., 2023)	Multilayer Convolutional Neural Network for Maize Leaf Disease Classification	The paper presents a multilayer convolutional neural network model for maize leaf disease classification, achieving 93.28% accuracy. It automates the detection of diseases like blight, common rust, and gray leaf spot, aiding farmers in crop protection and yield improvement.

6	(Singh et al., 2023)	Computer based Detection and Classification of Leaf Diseases using Hybrid Features	The research paper explores machine learning algorithms for leaf disease detection and classification, including SVM, KNN, SGD, XGB, and random forest, demonstrating their effectiveness in accurately identifying leaf diseases, which can be applied to maize and other crops.
7	(Varshney et al., 2022)	Plant Disease Detection Using Machine Learning Techniques	The paper discusses the development of a new plant leaf disease detection technique using transfer learning with deep learning and SVM, achieving an 88.77% training accuracy.
8	(Suthar et al., 2023)	An Extensive Evaluation of Plant Disease Detection Using Diverse Machine Learning Approaches	The study evaluates various machine learning techniques, including CNNs, SVMs, random forests, and KNNs, for plant disease detection. It emphasizes the importance of feature extraction from images and demonstrates the effectiveness of these methods in accurately identifying plant diseases.
9	(Prof. Vrushali Paithankar et al., 2023)	Plant Disease Detection using Machine Learning	The paper proposes a plant disease detection system that uses image processing and convolutional neural networks (CNNs) for real-time detection of plant diseases.
10	(Joshi & Panse, 2023)	Investigating the Feasibility of Machine Learning Models for Early Detection of Plant Diseases	Machine learning techniques, particularly deep learning models like convolutional neural networks (CNNs), are highly effective for plant disease detection, achieving accuracies exceeding 90%. This study highlights their potential in identifying and diagnosing plant diseases, addressing global agricultural challenges.

2.8 Research Gap

From the various study of related works in 2.6, it is seen that many machine learning algorithm and deep learning technique has been applied in the field of agriculture for maize leaf disease detection and classification, the results obtained in the various study looks promising, however the problem of hyperparameter tuning still exist in the literature likewise usage of basic feature extraction technique

still exist. Random forest has shown more accuracy across the various literature. This study seek to improve on the machine learning algorithm (Random forest) by finetuning the hyperparameters and also using CNN for feature extraction.

CHAPTER THREE
RESEARCH METHODOLOGY

3.1 Introduction

This chapter describes the structure and flow of the proposed framework that will be use in this study. They consists of the following stages:

i. Implementation Setup and System Specification

ii. Image Collection

iii. Image Preprocessing

iv. Image Segmentation

v. Feature Extraction

vi. Hyperparameter Tuning Using GridSearchCV

vii. Classification

viii. Performance Metrics

3.2 Implementation Setup and System Specification

Python programming language is mostly used for the data preprocessing, training, and implementation of the classification model. Python's advantages such as ease of use, huge framework library and flexibility make it the best choice for creating machine learning models (Hao & Ho, 2019). Python offers machine learning settings a remarkable level of power and customization. The language's straightforward syntax makes it easier to validate data and speeds the scraping, processing, refining, cleaning, arranging, and analyzing operations, which reduces the difficulty of working with other programmers (Srinath, 2017).

The following are the hardware and software specifications used for the implementation of the model:

Table 3.1: Summary of hardware and software specifications

Software specification	
Language	Python 3.11.0
Platform	Google Colab, RAM 13GB
Operating system	Ubuntu 24.04.2 LTS
Hardware specification	
Processor	Intel(R) Core(TM) i5-8350U CPU @ 1.70GHz x 8
Computer model	Dell Inc Latitude 5490 Intel(R) at 1.70GHz.
Memory	16.00 GB of RAM
Storage	256.1 GB of Solid State Drive

The overview of the proposed framework is shown below with each component of the framework explained in the following section.

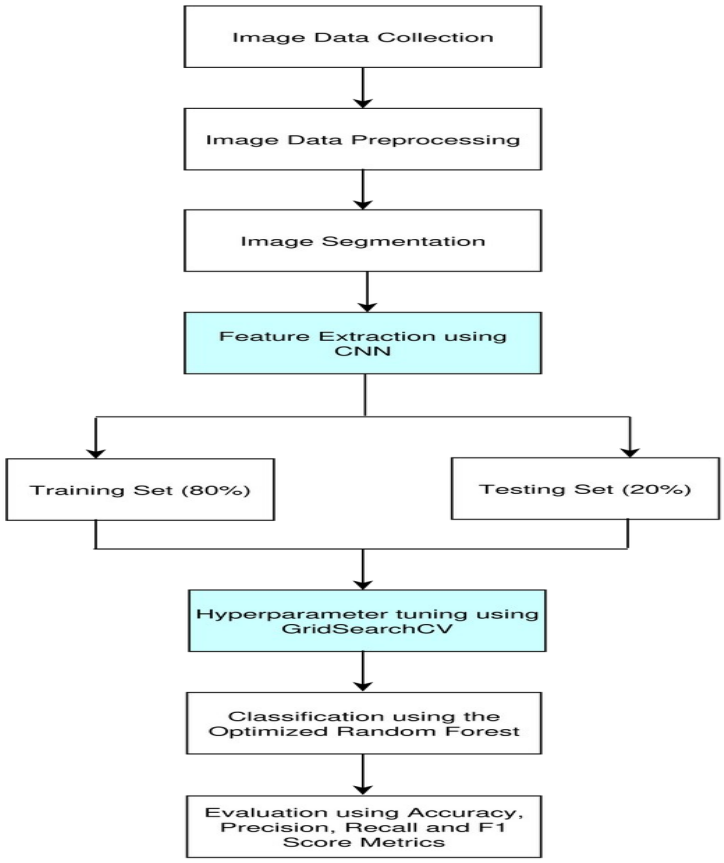

Fig 3.1: Overview of the proposed framework

3.3 Image Collection

The image dataset, particularly for maize disease pictures, is available at the plant village hosted with https://www.kaggle.com/datasets/abdallahalidev/plantvillage-dataset in Kaggle repository. The dataset is a public benchmark dataset used for plant disease detection and classification. Maize plants are the subsets that have the total number of 3823 images and four class labels of diseases such as common rust, gray leaf spot, northern leaf blight and healthy having 1192 images, 513 images, 956 images, and 1162 images respectively. These labeled images are considered for the training and testing of the disease classification.

3.4 Image Preprocessing

The image preprocessing is necessary for the realization of the superior results in consequence steps due to the presence of dewdrops, dust, insect excrements on the plants. These effects are considered as the noise of the maize image. To overcome these problems the input RGB photo is transformed into a grayscale image to provide accurate results. In this case, the size of the pictures is very large for which the reduction in the image size is necessary. This image reduction is also useful to reduce memory size.

3.5 Image Segmentation

Image segmentation plays a crucial role in plant disease detection and classification. It simply divides the image into several objects or regions. It analyzes the image data to extract useful information for further processing. This image segmentation can be carried out in two ways based on similarities and discontinuities. In similarities, the images are partitioned based on some specific predefined criteria. Therefore, the label edge detection method is used in image segmentation and also it calculates the gradient of photograph intensities at each pixel within the image. But in discontinuities, the images are partitioned based on the sudden changes in the intensity of values such as edge detection.

3.6 Feature Extraction

Feature extraction extracts the features of the objects that are present in the images. These extracted features are used to illustrate an entity. These features extracted and categorized into three categories such as shape, color, and texture. The diseases may vary their shapes into different several shapes of the image due to diseases. The model can easily identify the diseases from the shape of the features. These shapes of the features vary in their axis, areas, and angles. The second parameter, i.e., color is an important feature of these three features. It differentiates the diseases from each other. The third parameter, i.e., texture describes how the patterns of the color are sprinkled in the images. RGB feature extraction extracts the color information from the frequently used images for processing and identification of patterns. RGB is highly recommended for object detection in the image. It has the significant change in color that easily identifies the images in the leaves. The value of RGB color can determine all probable colors that can be made from the three colored lights such as red, green, and blue. The standard value of RGB varies from 1 to 255 and the tasks are normalized in the range of 0–1. This experiment considers the grayscale pixel values as features for analysis. This experiment uses CNN for the features extraction.

3.7 Hyperparameter Tuning using GridSearchCV

This experiment uses GridSearchCV function in the sklearn module to fine-tune the hyperparameters of the random forest model. Grid search is a hyperparameter tuning technique that involves testing a range of values for each hyperparameter to find the optimal combination of hyperparameters that works well with our datasets. Random forest has several hyperparameters that can be tuned to improve the model's performance. These hyperparameters include:

i. n_estimators: this refers to the number of trees in the forest. If it is too few, then it will lead to underfitting and if it is too many, it will lead to slower training (diminishing returns). It typical value range is between [50, 200, 500, 1000].

ii. max_depth: this refers to the maximum depth of each tree. Deeper trees is usually more complex and the model stand a risk of overfitting. The maximum depth of a tree usually range between [5, 10, 20, None (unlimited)].

iii. min_samples_split: this refers to the minimum samples required to split an internal node. Usually, higher values means simpler trees which prevents overfitting. It value range is between [2, 5, 10].

iv. min_samples_leaf:this is the minimum samples required in a leaf node. Usually, higher values means smoother predictions. The value ranges between [1, 2, 4].

v. max_features: this is the number of features considered for splitting ("sqrt", "log2", or integer/float). Lower values means more randomness and this reduces overfitting. The typical options are: "sqrt", "log2", 0.3 or 0.5.

vi. bootstrap: this is a boolean option which determine whether to use the entire datasets or a bootstrapped samples. The default value is usually True, if it is False, the whole dataset is used for each tree which means less randomness.

vii. oob_score: this is an option of whether to use out-of-bag (OOB) samples for validation. The default is usually False. It is Useful for estimating generalization without cross-validation.

viii. class_weight: it handles imbalanced classes ("balanced", "balanced_subsample"). It is usually for classification.

ix. criterion: this is the splitting criterion ("gini" or "entropy" for classification, "squared_error" for regression).

3.8 Classification

This experiment uses random forest for the classification task. Since CNN is use for the feature extraction, the last layer in the architecture of the CNN is removed and the random forest is used for classification.

3.9 Performance Metrics

Common evaluation metrics, such as accuracy, the number of features, sensitivity and specificity were used. The definition for some of the evaluation metrics are given below:

Accuracy: It is a measure that is defined as a total, correctly identified examples out of all the examples. Accuracy is determined as:

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

where TP refers to True Positive, which means correctly identified, (FP) refers to incorrectly identified or False Positive, TN is defined as True Negative or correctly rejected. FN refers to False-Negative, which means incorrectly rejected.

Recall: this is the True Positive Rate (TPR). It is the degree to which the learning algorithm is able to identify the data records that have been positively classified. It is calculated by:

$$Recall = \frac{TP}{TP + FN}$$

Precision: this is the True Negative Rate (TNR). It defines the proportion of actual negatives that are correctly identified. The learning algorithm's capacity to identify records of data with the negative class is demonstrated by the true negative rate, which is used to measure specificity. It's calculated by:

$$Precision = \frac{TN}{TN + FP}$$

REFERENCE

Ajit, A., Acharya, K., & Samanta, A. (2020). A Review of Convolutional Neural Networks. *2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE)*, 1–5. https://doi.org/10.1109/ic-ETITE47903.2020.049

Albelwi, S., & Mahmood, A. (2017). A Framework for Designing the Architectures of Deep Convolutional Neural Networks. *Entropy*, *19*(6), 242. https://doi.org/10.3390/e19060242

Ali, A. H., Youssef, A., Abdelal, M., & Raja, M. A. (2024). An ensemble of deep learning architectures for accurate plant disease classification. *Ecological Informatics*, *81*, 102618. https://doi.org/10.1016/j.ecoinf.2024.102618

Bachhal, P., Kukreja, V., & Ahuja, S. (2023). Maize Disease classification using Deep Learning Techniques: A Review. *2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, 259–264. https://doi.org/10.1109/InCACCT57535.2023.10141847

Bachhal, P., Kukreja, V., Ahuja, S., Lilhore, U. K., Simaiya, S., Bijalwan, A., Alroobaea, R., & Algarni, S. (2024). Maize leaf disease recognition using PRF-SVM integration: A breakthrough technique. *Scientific Reports*, *14*(1), 10219. https://doi.org/10.1038/s41598-024-60506-8

Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., & Prasanna, B. M. (2022). Global maize production, consumption and trade: Trends and R&D implications. *Food Security*, *14*(5), 1295–1319. https://doi.org/10.1007/s12571-022-01288-7

Golay, J., & Kanevski, M. (2017). Unsupervised feature selection based on the Morisita estimator of intrinsic dimension. *Knowledge-Based Systems*, *135*, 125–134. https://doi.org/10.1016/j.knosys.2017.08.009

Hao, J., & Ho, T. K. (2019). Machine Learning Made Easy: A Review of *Scikit-learn* Package in Python Programming Language. *Journal of Educational and Behavioral Statistics*, *44*(3), 348–361. https://doi.org/10.3102/1076998619832248

Iftikhar, M., Kandhro, I. A., Kausar, N., Kehar, A., Uddin, M., & Dandoush, A. (2024). Plant disease management: A fine-tuned enhanced CNN approach with mobile app integration for early

detection and classification. *Artificial Intelligence Review*, *57*(7), 167. https://doi.org/10.1007/s10462-024-10809-z

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Joshi, G., & Panse, P. (2023). Investigating the Feasibility of Machine Learning Models for Early Detection of Plant Diseases. *2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, 289–294. https://doi.org/10.1109/ICSSAS57918.2023.10331640

Kavya, R. & others. (2015). Feature extraction technique for robust and fast visual tracking: A typical review. *International Journal of Emerging Engineering Research and Technology*, *3*(1), 98–104.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*(1), 381–386.

Mutlag, W. K., Ali, S. K., Aydam, Z. M., & Taher, B. H. (2020). Feature Extraction Methods: A Review. *Journal of Physics: Conference Series*, *1591*(1), 012028. https://doi.org/10.1088/1742-6596/1591/1/012028

Nsibo, D. L., Barnes, I., & Berger, D. K. (2024). Recent advances in the population biology and management of maize foliar fungal pathogens Exserohilum turcicum, Cercospora zeina and Bipolaris maydis in Africa. *Frontiers in Plant Science*, *15*, 1404483. https://doi.org/10.3389/fpls.2024.1404483

Pandey, G., Sharma, R., & Kukker, A. (2023). Multilayer Convolutional Neural Network for Maize Leaf Disease Classification. *2023 7th International Conference on Computer Applications in Electrical Engineering-Recent Advances (CERA)*, 1–5. https://doi.org/10.1109/CERA59325.2023.10455141

Panigrahi, K. P., Das, H., Sahoo, A. K., & Moharana, S. C. (2020). Maize Leaf Disease Detection and Classification Using Machine Learning Algorithms. In H. Das, P. K. Pattnaik, S. S. Rautaray, & K.-C. Li (Eds.), *Progress in Computing, Analytics and Networking* (Vol. 1119, pp. 659–669). Springer Singapore. https://doi.org/10.1007/978-981-15-2414-1_66

Prof. Vrushali Paithankar, Ajinkya Awari, Akash Raskar, Shrirameshwar Patil, & Namrata Jamdar. (2023). Plant Disease Detection using Machine Learning. *International Journal of Advanced Research in Science, Communication and Technology*, 267–272. https://doi.org/10.48175/IJARSCT-9297

Rossman, A. Y. (2009). The impact of invasive fungi on agricultural ecosystems in the United States. In D. W. Langor & J. Sweeney (Eds.), *Ecological Impacts of Non-Native Invertebrates and Fungi on Terrestrial Ecosystems* (pp. 97–107). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9680-8_7

Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random Forest Algorithm Overview. *Babylonian Journal of Machine Learning*, *2024*, 69–79. https://doi.org/10.58496/BJML/2024/007

Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Setiawan, W., Rochman, E. M. S., Satoto, B. D., & Rachmad, A. (2022). Machine Learning and Deep Learning for Maize Leaf Disease Classification: A Review. *Journal of Physics: Conference Series*, *2406*(1), 012019. https://doi.org/10.1088/1742-6596/2406/1/012019

Singh, S., Roy, Y., Bhan, A., & Sah, S. (2023). Computer based Detection and Classification of Leaf Diseases using Hybrid Features. *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, 788–793. https://doi.org/10.1109/ICSCSS57650.2023.10169167

Srinath, K. R. (2017). Python – The Fastest Growing Programming Language. *International Research Journal of Engineering and Technology*, *04*(12), 354–357.

Suthar, F., Padhya, K., Joshi, R., Parikh, S., Panthakkan, A., & Mansoor, W. (2023). An Extensive Evaluation of Plant Disease Detection Using Diverse Machine Learning Approaches. *2023 6th International Conference on Signal Processing and Information Security (ICSPIS)*, 151–155. https://doi.org/10.1109/ICSPIS60075.2023.10343735

Varshney, D., Babukhanwala, B., Khan, J., Saxena, D., & Singh, A. K. (2022). Plant Disease Detection Using Machine Learning Techniques. *2022 3rd International Conference for Emerging Technology (INCET)*, 1–5. https://doi.org/10.1109/INCET54531.2022.9824653

Velliangiri, S., Alagumuthukrishnan, S., & joseph, S. I. T. (2019). A Review of Dimensionality Reduction Techniques for Efficient Computation. *Procedia Computer Science*, *165*, 104–111. https://doi.org/10.1016/j.procs.2020.01.079

Vincent Mbandu Ochango, Geoffrey Mariga Wambugu, & John Gichuki Ndia. (2022). Feature Extraction using Histogram of Oriented Gradients for Image Classification in Maize Leaf Diseases. *International Journal of Computer and Information Technology(2279-0764)*, *11*(2). https://doi.org/10.24203/ijcit.v11i2.204