AgriHealth AI : Real-time Plant Health Monitoring System

Student ID: 250040722

Student Name : Mahmud Hasan

Supervisor : Dr. Dan Dai

Department : Artificial Intelligence with Business Strategy

Declaration:

I declare that I have personally prepared this assignment. The work is my own, carried out personally by me unless otherwise stated and has not been generated using paid for assessment writing services or Artificial Intelligence tools unless specified as a clearly stated approved component of the assessment brief. All sources, including citations, are recognized through appropriate citations and references. I declare that this work has not gained credit previously for another module at this or another University, save for permitted elements which formed part of an associated proposal linked directly to this submission.

I understand that plagiarism, collusion, copying another student and commissioning (which for the avoidance of doubt includes the use of essay mills and other paid for assessment writing services, as well as unattributed use of work generated by Artificial Intelligence tools) are regarded as offences against the University's Assessment Regulations and may result in formal disciplinary proceedings.

I understand that by submitting this assessment, I declare myself fit to be able to undertake the assessment and accept the outcome of the assessment as valid.

Student signature: Mahmud Hasan

Date: 23 January 2026

Abstract:

Plant diseases pose a significant threat to global agricultural productivity, particularly in fruit crops such as apples, where early diagnosis is essential to prevent widespread product loss. Advancements in image-based machine learning have enabled automated detection systems capable of supporting farmers and agronomists in real-time decision making. Real-time, automated disease monitoring systems offer a scalable alternative to manual inspection. However, the effectiveness relies heavily on the accuracy and robustness of the underlying image classification models. This study develops and evaluates the machine learning core of AgriHealth AI, a supervised learning system for a real-time plant health monitoring system by investigating how multiple image denoising techniques influence classical machine learning performance in apple leaf disease classification.

Using the Kashmiri Apple Plant Disease Dataset (four classes: Healthy Leaves, Scab Leaves, Leaf Blotch, Apple Rot Leaves), all images were preprocessed through RGB resizing (48x48), flattening, and standardisation. Five denoising methods were examined, Gaussian Blur, Median Blur, Bilateral Filtering, Non-Local Means(NLMeans) and Kernel Principal Component Analysis (Kernel PCA) to analyse how each affects downstream classification. Six classical machine learning algorithms were evaluated under a stratified 80/20 train-train-split. Support Vector Machine (SVM) with linear and RBF kernels, Logistic Regression, K-Nearest Neighbours, Random Forest, Gradient Boosting, and Naive Bayes. The RBF based SVM underwent Hyperparameter Optimisation using GridSearchCV.

Results show that Gaussian denoising combined with a tuned SVM (RBF kernel) produced the highest accuracy of 0.64, outperforming all baseline and alternative denoising methods, More complex and advanced approaches such as NLMeans and Kernel PCA did not enhance accuracy, likely due to loss of fine texture information relevant to disease pattern recognition. These findings demonstrate that lightweight smoothing methods offer a practical balance between computational efficiency and predictive performance, making them suitable for real-time agricultural applications.

This study provides a comparative evaluation of classical Machine Learning pipeline for leaf disease classification and highlights the importance of preprocessing in improving model reliability. Future work will integrate and extend the system towards a web based dashboard and a real-time mobile based deployment of AgriHealth AI.

Table of Content

1. Introduction:

1.1 Background:

Agriculture productivity is increasingly threatened by plant disease which affects both crop yield and the quality of product itself. Apple cultivation is particularly vulnerable due to its sensitivity to foliar diseases that primarily affects a plant's leaves, often caused by fungi, bacteria or viruses. Many of these diseases manifest in early stages through subtle visual patterns on leaves. Detecting these symptoms in time is crucial for preventing disease spread and mitigating losses. Conventional inspection practices rely on growers or agricultural specialists visually examining leaves, a process that is labour-intensive, inconsistent, and difficult to scale across large orchards [2].

Recent advancements in computer vision and machine learning have enabled automated plant disease detection systems based on image analysis. Image-based classification approaches allow disease symptoms to be identified efficiently and consistently, offering valuable decision-support tools for precision agriculture [4] [5]. However, the performance of such systems is highly dependent on the quality of input images. In real-world agricultural environments, images are frequently affected by noise, blur, variations in lighting conditions, and background interference. These factors can obscure disease-relevant visual features and significantly reduce classification accuracy if not properly addressed through appropriate preprocessing techniques [2] [6].

1.2 Motivation:

Image preprocessing plays a critical role in improving the robustness of image-based plant disease detection systems. Among preprocessing techniques, image denoising is commonly applied to reduce unwanted noise and enhance relevant visual structures. While denoising has the potential to improve classification performance, excessive smoothing may also remove fine texture details that are important for distinguishing between visually similar disease classes. This trade-off highlights the importance of systematically evaluating denoising techniques rather than assuming universal performance improvement.

The AgriHealth AI project aims to develop the machine learning foundation of a real-time plant health monitoring system with a focus on efficiency and practical deployment. Given the constraints of real-time applications, this study focuses on classical machine learning algorithms, which offer lower computational

complexity and faster inference compared to deep learning approaches. However, classical classifiers such as Support Vector Machine (SVM) and ensemble-based methods are particularly sensitive to feature quality, making effective preprocessing essential for reliable performance.

1.3   Aim of the Study:

The aim of this study is to investigate how different image denoising techniques influence the performance of classical machine learning classifiers in apple leaf disease classification. Rather than proposing a novel architecture, this research focuses on understanding the interaction between preprocessing strategies and classifiers behaviour within a controlled and reproducible experimental framework.

The primary contribution of this study lies in the systematic and comparative evaluation of multiple denoising techniques in combination with a diverse set of classical machine learning classifiers. While existing studies often evaluate a limited number of classifiers or apply a single, fixed preprocessing strategy, this work explicitly treats image denoising as a key experimental variable. Five denoising techniques - Gaussian blur, Median blur, Bilateral Filtering, Non-Local Means, and Kernel Principal Component Analysis (Kernel PCA) Reconstruction are evaluated under identical conditions to assess their impact on downstream classification performance.

This study conducts a broad comparison of classical classifiers, including linear and non-linear SVMs, Logistic Regression, KNN, Random Forest, Gradient Boosting, and Naive Bayes. By analysing classifier performance across both raw and denoised dataset, the study provides insight into how different learning paradigms respond to noise reduction and feature smoothing. Hyperparameter Optimisation is incorporated where appropriate to ensure fair and meaningful comparison.

Another distinguishing aspect of this is its exclusive focus on classical machine learning approaches, in contrast to the deep learning-centric direction of much recent literature. This design choice reflects a practical emphasis on computational efficiency, interpretability, and suitability for real-time agricultural deployment.

Overall, this research contributes an empirical analysis of preprocessing-classifier interactions for apple leaf disease classification, offering

evidence-based guidance on the effectiveness of denoising strategies within classical machine learning pipelines.

1.4 Research Objectives:

The objectives of this research are to preprocess apple leaf images for classification, apply and compare multiple image denoising techniques, evaluate the performance of classical machine learning classifiers under different preprocessing conditions, and identify an optimal configuration that balances classification accuracy and computational efficiency.

1.5 Research Questions / Hypothesis:

This study is guided by the following research questions:
1. How do different image denoising techniques affect the classification performance of classical machine learning models in apple leaf disease detection?
2. Does image denoising improve classification accuracy compared to using unprocessed images?
3. Which denoising technique provides the most effective balance between noise reduction and preservation of disease-relevant visual features?
4. Which classical machine learning classifier performs most effectively under varying preprocessing conditions?

Based on these research questions, the following hypothesis is proposed:
- H_1 : Applying appropriate image denoising techniques improves the classification performance of classical machine learning models for apple leaf disease detection compared to using non-denoising images.

2. Literature Review:

2.1 Image-based Plant Disease Detection:

Automated plant disease detection using leaf imagery has been widely studied due to the visible manifestation of disease symptoms on foliage, including discoloration, lesions, blotches, and texture irregularities. Early research commonly adopted classical computer vision pipelines consisting of image preprocessing, segmentation, handcrafted feature extraction, and classification using machine learning models such as Support Vector Machines (SVM) and K-Nearest Neighbours (KNN) (Elangovan and Nalini, 2017 [1]).

These approaches demonstrated that RGB images contain sufficient discriminative information for disease recognition under controlled conditions. However, several studies report that performance is highly sensitive to image quality, lighting variation, and background noise, limiting generalisability in real-world agricultural environments (Barbedo, 2018 [2]).

Comprehensive surveys highlight that while classical methods are computationally efficient and interpretable, their success depends heavily on preprocessing choices and feature stability, which are often insufficiently analysed in application-driven studies (Too *et al.*, 2019 [3]).

2.2 Transition to Deep Learning Approaches

With the emergence of deep learning, convolutional neural networks (CNNs) have become the dominant approach for plant disease classification. Mohanty *et al.* (2016) [4] demonstrated high classification accuracy using CNNs on the PlantVillage dataset, establishing deep learning as a powerful alternative to handcrafted feature pipelines.

Subsequent studies further expanded this approach across multiple crops and disease categories, with Ferentinos (2018) [5] reporting strong performance across a large-scale dataset. Despite these advances, later analyses revealed that CNN-based models trained on controlled datasets often suffer performance degradation when applied to images captured under field conditions (Barbedo, 2019 [6]).

These findings suggest that while deep learning excels in accuracy under ideal conditions, it introduces challenges related to computational cost, data requirements, and deployment feasibility, particularly in resource-constrained agricultural settings.

2.3 Apple Leaf Disease Classification

Apple crops are particularly vulnerable to foliar diseases such as apple scab, leaf blotch, and apple rot, which can significantly reduce yield if not detected early. As a result, apple-specific disease classification has attracted increasing attention.

Sharma, Padha and Bashir (2022) [7] introduced the Kashmiri Apple Plant Disease (D-KAP) dataset and proposed a deep learning-based framework for apple disease prediction. Their work provides a valuable, domain-specific dataset and highlights the feasibility of automated apple disease detection. However, the study primarily focuses on deep learning performance and does not extensively analyse classical machine learning pipelines or the influence of preprocessing variations.

Other apple disease studies similarly prioritise model accuracy while offering limited discussion on preprocessing strategies, leaving an open question regarding the role of noise reduction and feature stability in apple-specific datasets (Singh *et al.*, 2020 [8]).

2.4 Classical Machine Learning in Plant Disease Detection

Despite the dominance of deep learning, classical machine learning models remain relevant, particularly for applications requiring interpretability and low computational overhead. Support Vector Machines have been repeatedly shown to perform well on high-dimensional image features, especially when combined with appropriate preprocessing and kernel selection (Elangovan and Nalini, 2017 [1]).

Ensemble methods such as Random Forest and Gradient Boosting have also been applied to plant disease classification due to their robustness to noise and ability to model non-linear relationships (Too *et al.*, 2019 [3]). However, many studies evaluate only a single classifier, limiting insight into how different learning paradigms respond to identical data transformations.

This lack of systematic comparison restricts the ability to generalise findings across models and datasets.

2.5 Image Preprocessing and Denoising

Image preprocessing is widely acknowledged as a critical component of plant disease classification pipelines. Techniques such as resizing, normalisation, smoothing, and contrast enhancement are commonly applied to reduce noise and improve feature quality (Barbedo, 2018 [2]).

Recent studies have further highlighted that image preprocessing choices, including noise reduction and enhancement strategies, have a significant impact on classification performance, particularly when lightweight or classical machine learning models are employed in agricultural image analysis (Khan *et al.*, 2023 [9]).

However, denoising is frequently treated as an implicit or fixed preprocessing step rather than an experimental variable. Few studies systematically evaluate multiple denoising techniques or analyse how they interact with different classifiers. This is particularly relevant in plant disease classification, where disease symptoms often appear as fine-grained textures that may be inadvertently removed by aggressive smoothing (Li *et al.*, 2023 [10]).

Advanced denoising techniques such as Non-Local Means and dimensionality reduction-based reconstruction methods have been proposed in broader image processing literature, but their effectiveness in low-resolution agricultural datasets remains underexplored.

2.6 Identified Research Gaps

From the reviewed literature, several research gaps are identified:

1. Limited systematic evaluation of denoising techniques in plant disease classification pipelines.

2. Insufficient comparative analysis of classical classifiers under identical preprocessing conditions.

3. Over-reliance on deep learning accuracy metrics, with limited consideration of efficiency and interpretability.

4. Lack of apple-specific studies analysing preprocessing–classifier interactions, despite the availability of crop-focused datasets such as D-KAP.

2.7 Linkage to Project Aims

The identified gaps directly motivate this study's goals. Rather than proposing a novel deep learning architecture, this project focuses on systematically analysing how different denoising techniques affect the performance of classical machine learning classifiers on an apple-specific dataset.

By evaluating multiple classifiers, applying consistent preprocessing, and incorporating hyperparameter optimisation, this study addresses the underexplored interaction

between noise reduction and classification performance. The emphasis on lightweight and interpretable methods aligns with the practical goal of developing a real-time plant health monitoring system suitable for deployment in agricultural environments.

3. Methodology:

3.1 Research Design and Rationale:

The dissertation follows a quantitative, experimental research design. The objective is to evaluate the impact of image denoising on the classification performance of classical machine learning (ML) models for apple leaf disease detection. The study adopts a comparative benchmarking approach, where multiple classifiers are trained and tested under consistent conditions, and performance is assessed using standard evaluation metrics.

The methodological rationale is based on two considerations. First, image-based plant disease classification performance is sensitive to image quality, and images captured in practical settings may contain noise, blur, and illumination variation. Second, classical machine learning models are computationally efficient and therefore suitable for integration into systems with real-time constraints. However, classical models are also sensitive to feature representation and image preprocessing quality. For these reasons, this study focuses on systematically evaluating multiple denoising techniques and multiple classical classifiers rather than relying on a single model configuration. This approach supports both reliability, robustness, and provides stronger evidence for selecting an appropriate preprocessing and classifier pipeline for the AgriHealth AI system.

3.2   Project Framework:

The overall project framework consists of a complete image-classification pipeline implemented in Python using Jupyter Notebook. The framework includes dataset loading, preprocessing, denoising, model training, evaluation, and comparative reporting.
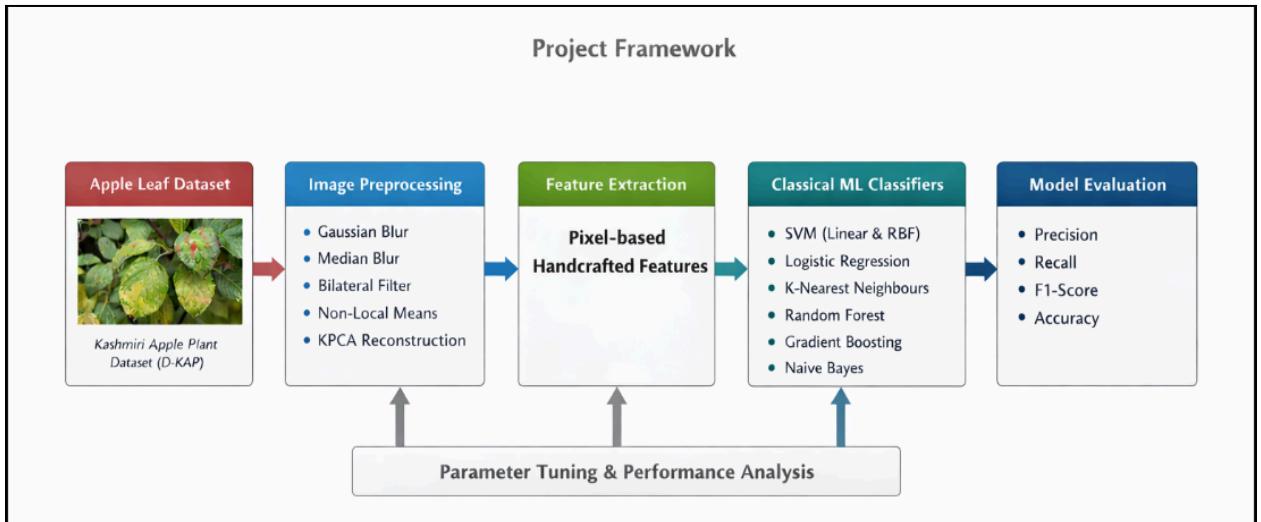
Figure 3.1: Project Pipeline Visualisation

3.3 Dataset and Class Labels:

This study uses the Kashmiri Apple Plant Disease Dataset obtained from Kaggle. The dataset contains four labelled classes:
- Healthy Leaves
- Scab Leaves
- Leaf Blotch
- Apple Rot Leaves

All images were organised in a directory structure where each class name corresponds to a folder containing relevant images. The dataset size in this study was 419 images.

3.4 Data Loading and Image Standardisation:

All images were loaded using OpenCV. During loading, each image was:
1. Read from disk,
2. Resized to a fixed resolution of 48x48 pixels in RGB format, and
3. Converted into a numerical representation suitable for machine learning.

To support model compatibility, images were flattened into one-dimensional feature vectors. This produced an input feature matrix of dimension:

$$X \in R^{N \times (48\times48\times3)} = R^{N \times 6912}$$

The corresponding class labels were stored as categorical strings. A separate copy of RGB images was retained for visual inspection and qualitative assessment of predictions.

3.5 Train-Test-Split Strategy

The dataset was divided into training and testing partitions using a stratified 80/20 train-test-split, with a fixed random seed for reproducibility. Stratification ensures that each class is represented proportionally in both the training test sets. The split sizes observed in the notebook were 335 training samples and 84 test samples, ensuring a consistent evaluation set across experiments.

3.6 Feature Scaling

Since the input features consist of numeric pixel intensity values, feature scaling was applied to improve optimisation behaviour in models that are sensitive to feature magnitude, particularly Support Vector Machines (SVM) and Logistic Regression (LR). Standardisation was performed using StandardScaler, transforming features to approximately zero mean and unit variance based on the training data. The same scaling parameters were applied to the test data to prevent data leakage.

3.7 Image Denoising Techniques

Image denoising refers to the process of reducing unwanted variation in pixel intensity that arise from factors such as lighting conditions, sensor noise, compression artefacts, or background inconsistencies. In image-based plant disease classification, such as noise can obscure subtle visual cues associated with disease symptoms and negatively affect the reliability of feature representations extracted from the images

In this study, denoising was investigated as a preprocessing step applied to resized RGB images prior to feature flattening and classification. The motivation for this approach lies in the nature of plant disease symptoms, which often manifest as texture-like patterns such as spots, lesions, discolorations, and blotches on leaf surfaces. An effective denoising technique should therefore reduce irrelevant noise while preserving discriminative visual information essential for classification.

To examine how different noise-reduction strategies influence classification performance, both simple and advanced denoising techniques were selected and evaluated under consistent experimental conditions.

1. **Gaussian Blur :** Gaussian blur is widely used smoothing techniques that reduces high-frequency noise by convolving the image with a Gaussian kernel. The degree of smoothing is controlled by the kernel size and standard deviation, which determine how strongly neighbouring pixels influence each other. Gaussian denoising is computationally efficient and tends to preserve the overall structure of the image while attenuating fine-grained noise.

   In the context of leaf disease classification, Gaussian blur was expected to stabilise pixel intensity distributions while retaining coarse texture patterns associated with disease regions. Its simplicity and low computational cost make it particularly attractive for real-time applications.

2. **Median Blur :** Median blur is a non-linear filtering technique that replaces each pixel value with the median value within a local neighbourhood. Unlike Gaussian blur, median filtering is effective at removing impulsive noise without excessively blurring edges.

   This method was included to assess whether preserving sharp boundaries, such as lesion edges or spot outlines, would improve classification performance. Median blur was therefore evaluated as an edge-preserving alternative to Gaussian smoothing, particularly for disease characterised by well-defined visual boundaries.

3. **Bilateral Filtering :** Bilateral filtering extends traditional smoothing by combining spatial proximity and pixel intensity similarity when averaging neighbouring pixels. As a result, it reduces noise while attempting to preserve edges and structural details.

   For this study, Bilateral filtering was evaluated to determine whether maintaining structural integrity at disease boundaries could enhance discriminative feature extraction. However, the non-uniform smoothing introduced by bilateral filtering can also alter local intensity distributions, which may affect classifiers sensitive to feature geometry.

4. **Non-Local Means (NL Means) :** Non-Local Means denoising is an advanced technique that reduces noise by averaging pixels with similar neighbourhood

patterns across the entire image rather than relying solely on local proximity. By exploiting redundancy in image textures, NLMeans aims to preserve fine details while removing noise.

This method was selected to examine whether patch-based denoising could better retain subtle disease textures compared to local filtering methods. However, NLMeans is computationally expensive and its effectiveness can be limited when applied to low-resolution images, where meaningful patch similarity is harder to establish.

5. **Kernel Principle Component Analysis (Kernel PCA) Reconstruction :** Kernel PCA reconstruction was evaluated as a feature-space denoising approach rather than a traditional image-domain filter. In this method, feature vectors are projected into lower-dimensional non-linear space using kernel function, and then reconstructed back into the original feature space. The reconstruction process is intended to remove noise by discarding low-variance components.

This approach was included to investigate whether dimensionality reduction and reconstruction could improve generalisation by filtering irrelevant variance from high-dimensional image features. Kernel PCA represents a fundamentally different denoising strategy compared to spatial filtering, allowing a border comparison of preprocessing techniques within the study.

The inclusion of both simple smoothing techniques and more advanced methods enables a comprehensive assessment of how varying levels of denoising complexity affect classification performance. By evaluating these techniques under identical conditions, this study aims to determine whether increased preprocessing sophistication leads to improved predictive accuracy or whether simpler methods provide a more effective balance between noise reduction and feature preservation.

3.8 Classification Models

To evaluate the effectiveness of classical machine learning approaches for apple leaf disease classification, multiple classification models were implemented and compared under consistent experimental conditions. The selection of models was guided by the objective of examining a diverse range of learning paradigms, including linear classifiers, distance-based methods, ensemble techniques, and probabilistic approaches. This diversity allows for a comprehensive assessment of how different algorithms respond to raw pixel-based images features and denoising strategies.

All classifiers were trained using standardised feature representations to ensure fair comparison and to prevent scale-related bias.

3.8.1 Support Vector Machines (SVMs)

Support Vector Machines were selected due to their effectiveness in high-dimensional feature spaces and their strong theoretical foundation in margin-based optimisation. SVMs aim to identify an optimal decision boundary that maximises the margin between classes, which often leads to improved generalisation on unseen data. This property is particularly valuable when working with flattened image representations, where the number of features is large relative to the number of samples.

To comprehensively evaluate the suitability of SVMs for apple leaf disease classification, three different SVM configurations were implemented and analysed in this study.

1. ***Support Vector Machine with Linear Kernel:*** The linear SVM uses a linear decision boundary to separate classes in the feature space. This variant was included as a baseline model to assess whether the disease classes were linearly separable when represented as flattened RGB pixel vectors.

    Given the visual complexity of plant disease symptoms, which often involve non-linear patterns such as irregular lesions and texture variations, the linear SVM was not expected to achieve optimal performance. However, its inclusion provides an important reference point for understanding the necessity of non-linear modelling in this study.

2. ***Support Vector Machine with Radial Basis Function (RBF) Kernel:*** The RBF kernel extends the SVM framework by enabling non-linear decision boundaries through implicit mapping of the input features into a higher-dimensional space. This allows the model to capture complex relationships between features that cannot be represented using linear separation.

    In the context of leaf disease classification, the RBF kernel was expected to perform substantially better than the linear kernel SVM due to its ability to model non-linear texture patterns and variations in colour distribution. As such, the RBF-based SVM represents a more expressive classifier suited to the characteristic of image-based data.

3. ***Support Vector Machine (RBF with Hyperparameter Optimisation) :*** While the RBF kernel provides increased modelling capacity, its performance is highly dependent on the selection of hyperparameters, particularly the regularisation parameter and kernel width. Inappropriate values can lead to underfitting or overfitting, resulting in poor generalisation performance.

   To address this, hyperparameter optimization was applied to the RBF-based SVM using GridSearchCV. This approach systematically evaluated combinations of hyperparameter values using cross-validation on the training data, selecting the configuration that achieved the best cross-validation performance. The tuned SVM was then restrained on the full training set before evaluation on the test data.

   This optimised SVM configuration was included to assess the extent to which careful hyperparameter selection can improve classification performance beyond that of default kernel settings.

3.8.2 Logistic Regression

Logistic Regression was included as a linear baseline classifier due to its simplicity, interpretability, and widespread use in classification tasks. Despite its name, Logistic Regression performs classification by modelling the probability of class membership using a linear decision boundary in the feature space.

This model was expected to provide a reference point for evaluating the benefits of non-linear classifiers. While Logistic Regression is computationally efficient and relatively robust, its linear nature limits its ability to capture complex relationships within high-dimensional image data. Nevertheless, it remains valuable for benchmarking and interpretability.

3.8.3 K-Nearest Neighbours (KNN)

K-Nearest Neighbours is an instance-based learning algorithm that classifies samples based on the majority class among their nearest neighbours in the feature space. Unlike parametric models, KNN does not learn an explicit model during training, instead relying directly on distance computations at inference time.

KNN was included to examine how local similarity-based classification performs on flattened image features. However, due to the high dimensionality of the feature vectors and the sensitivity of distance-based methods to noise, KNN was expected to be more susceptible to performance degradation in noisy or sparsely structured feature spaces.

3.8.4 Random Forest

Random Forest is an ensemble learning method that constructs multiple decision trees using random subsets of features and samples, and aggregates their predictions through majority voting. This approach reduces overfitting and improves robustness compared to individual decision trees.

Random Forest was selected for its ability to model non-linear relationships and handle noisy features effectively. In the context of this study, it was expected to perform well on denoised dataset by exploiting coarse structural patterns in the image features, even when fine-grained texture information was partially lost.

3.8.5 Gradient Boosting

Gradient Boosting is an ensemble technique that builds models sequentially, with each new model focusing on correcting the errors of the previous ones. Unlike Random Forest, which trains trees independently, Gradient Boosting optimises performance through iterative refinement.

This model was included to assess whether sequential learning could better capture subtle disease-related patterns within the feature space. Gradient Boosting is known for strong predictive performance but can be sensitive to noise and hyperparameter selection. Its inclusion enables comparison between different ensemble strategies and their interaction with denoised techniques.

3.8.6 Naive Bayes

Naive Bayes is a probabilistic classifier based on Baye's theorem, operating under the assumption that features are conditionally independent given the class label. Despite this strong assumption, Naive Bayes is computationally efficient and often performs surprisingly well on certain classification tasks.

In this study, Naive Bayes was included primarily as a baseline to evaluate the limitations of probabilistic models when applied to image data. Given the strong spatial and statistical dependencies between pixel values in images, Naive Bayes was expected to struggle, making it useful for highlighting the importance of feature relationships in plant disease classification.

3.9 Hyperparameter Optimisation

Hyperparameter Optimisation is the process of systematically identifying the most suitable configuration parameters for a machine learning model that are not learned directly from the training data. Unlike model parameters, which are estimated during

training, hyperparameters control the behaviour and capacity of the learning algorithm itself. Examples include the regularisation parameter and kernel settings in Support Vector Machines, the number of neighbours in K-Nearest Neighbours, and the number of estimators in ensemble models.

The use of hyperparameter optimization is essential because inappropriate hyperparameter choices can lead to underfitting or overfitting, resulting in suboptimal model performance regardless of the quality of the input data. In the context of this study, where high-dimensional image features were used, model sensitivity to hyperparameter settings was particularly pronounced. Consequently, relying on the default configurations would not provide a fair or reliable assessment of classifier capability.

In this project, hyperparameter optimization was primarily applied to the Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel using GridSearchCV. GridSearchCV performs an exhaustive search over a predefined set of hyperparameter combinations and evaluates each configuration using cross-validation. By splitting the training data into multiple folds, the method assessed how well each hyperparameter setting generalises to unseen data, thereby reducing the risk of selection bias. The configuration yielding the highest cross-validated performance is then then selected and retrained on a full training set before final evaluation on the test set.

This approach ensured that model performance improvements were attributed to meaningful parameter tuning rather than chance. Moreover, hyperparameter optimization played a critical role in achieving the highest overall classification accuracy. As such, hyperparameter optimization represents a key methodological component of the proposed AgriHealth AI framework.

4. Ethical Implications:

4.1 Ethical Considerations and Compliance:

This project was conducted in accordance with institutional ethical guidelines and focused exclusively on the use of secondary, publicly available data. No human participants were involved at any stage of the research. No personal, sensitive, or identifiable information was collected, processed, or stored. As a result, the ethical risks associated with this study are minimal.

The primary dataset used in this research is the Kashmiri Apple Plant Disease Dataset (D-KAP), obtained from Kaggle. The dataset consists solely of images of apple leaves representing healthy and diseased conditions and was originally introduced by Sharma et al. (2022). The dataset was accessed for academic research purposes only.

4.2   Dataset Source and Licensing:

The D-KAP dataset is distributed under the Database Contents License (DbCL) v1.0, which grants a worldwide, royalty-free, non-exclusive, perpetual, and irrevocable copyright license to use, share, adapt, and sublicense the database contents in any medium and for any purpose, including commercial use. The license applies specifically to the contents of the database and operates alongside the Open Database License (ODbL) for aspects related to database structure and database rights.

In accordance with the license terms, the dataset was used "as is", without any claim of ownership or warranty, and full acknowledgement has been provided to the original authors. No attempt was made to modify, redistribute, or monetise the dataset beyond the scope permitted by the license. The dataset was used strictly for educational and research purposes within the context of this dissertation.

4.3   Absence of Human Participant Data:

This study does not involve human participant data in any form. The dataset contains only images of plant leaves and does not include information relating to individuals, locations, or personally identifiable attributes. Consequently, issues such as informed consent, participant recruitment, or withdrawal do not apply. The research therefore does not pose risks related to privacy invasion, confidentiality breaches, or data misuse involving human subjects.

4.4   Data Storage, Access, and Security:

The dataset and all derived experimental outputs were stored securely within a personal computing environment, with backup maintained via OneDrive. Access to the data was restricted to the author of this study and academic supervisor for the duration of the project. No third-party access was permitted at any stage of the research.

As the dataset is publicly available and does not contain sensitive or personally identifiable information, no specific data deletion timeline was required. Nevertheless, standard good practice was followed to ensure responsible handling, secure storage, and controlled access to the data throughout the research lifecycle.

4.5   Bias and Limitations of Secondary Data:

An ethical consideration inherent in the use of secondary dataset is the potential presence of bias introduced during the original data collection and annotation process. The D-KAP dataset reflects the sampling decisions, imaging conditions, and labelling strategies employed by the original authors. These factors may influence class distribution, disease representation, and visual variability, which in turn can affect model performance and generalisability.

To mitigate this risk, the dataset source and associated documentation were carefully reviewed to ensure legal and ethical compliance with the stated licensing terms. In addition all findings derived from the dataset are interpreted within the context of these limitations, and no claims are made regarding universal applicability beyond the dataset's scope.

4.6   Responsible Use and Transparency:

This study maintains transparency by clearly documenting all preprocessing steps, denoising techniques, classification models, and evaluation strategies employed throughout the research. Proper citation of the dataset source and the original authors has been provided to ensure academic integrity and appropriate acknowledgment of intellectual contributions.

Furthermore, the capabilities of the proposed system are not overstated. Although the results demonstrate promising classification performance, the system is positioned as a decision-support tool rather than a definitive diagnostic solution. This distinction is important to reduce the risk of misuse or over-reliance on automated predictions in real-world agriculture contexts.

4.7   Ethical Impact of the Proposed System:

From a broader ethical perspective, the proposed AgriHealth AI system aims to support sustainable agricultural practices by enabling earlier detection of plant disease and potentially reducing crop losses. By prioritising lightweight and interpretable machine learning techniques, the system promotes accessibility and practical deployment, particularly in resource-constrained environments.

It is acknowledged, however, that automated plant disease detection systems should complement rather than replace expert agricultural knowledge. Any future deployment should ensure that users are informed of system limitations and encouraged to seek professional guidance where appropriate.

5. Results and Analysis:

5.1 Analysis Overview:

This represents the results obtained from implementing and evaluating a classical machine learning pipeline for apple leaf disease classification. The evaluation was conducted in two stages.

1. First, baseline performance was established using the original images without denoising.
2. Second, image denoising techniques were applied to assess their impact on classification performance.

All models were evaluated using a consistent stratified 80/20 train-test-split, feature standardisation, and identical evaluation metrics to ensure fair comparison. Performance was assessed quantitatively using accuracy, precision, recall, F1-score, and confusion matrices, supported by qualitative interpretation of misclassification patterns.

5.2 Data Analysis Methods:

5.2.1 Evaluation Metrics:

The performance of the proposed classification models was evaluated using a combination of quantitative metrics to provide a comprehensive and reliable assessment. Although classification accuracy was used as the primary metric for comparison, additional evaluation measures were employed to account for class imbalance and to better understand model behaviour at a per-class level.

The use of multiple evaluation metrics is particularly important in plant disease classification, where misclassification of certain disease categories may have different practical implications. Relying on a single metric can therefore provide a misleading representation of model performance.

5.2.2 Accuracy:

Accuracy measures the proportion of correctly classified samples relative to the total number of samples in the test set. It is defined as the ratio of correct predictions to all predictions made.

Accuracy was used as the primary metric in this study because it provides a simple and intuitive measure of overall model performance and allows for straightforward comparison across different classifiers and denoising techniques. However, accuracy

alone does not reveal how well individual disease classes are identified, particularly when class distributions are uneven. For this reason, accuracy was complemented with additional metrics.

5.2.3 Precision:

Precision measures the proportion of correctly predicted positive samples among all samples predicted as positive for a given class. In the context of plant disease classification, precision reflects how reliable the model's disease predictions are.

Mathematically, precision is defined as:

$$\textbf{Precision} = \frac{TP}{TP + FP}$$

where,
- TP (True Positives) represents the number of disease leaf samples correctly classified as diseased,
- FP (False Positives) represents the number of healthy leaf samples incorrectly classified as diseased

High precision indicates that when the model predicts a specific disease, it is likely to be correct. This metric is particularly important when false positives are undesirable, such as incorrectly identifying healthy leaves as diseased, which could lead to unnecessary intervention, increased cost, or inappropriate treatment decisions.

5.2.4 Recall:

Recall, also called sensitivity, measures the proportion of actual positive samples that are correctly identified by the model. It reflects the model's ability to detect all instances of a given disease.

Mathematically, Recall is calculated as,

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

where,
- TP (True Positives) represents the number of diseased leaf samples correctly identified,
- FN (False Negatives) represents the number of diseased leaf samples incorrectly classified as healthy

In agricultural applications, recall is especially important because failing to detect diseased plants can cause disease spread and increased crop loss. A model with high

recall ensures that most diseased leaves are correctly identified, even if this comes at the cost of some false positives.

5.2.5   F1-Score:

The F1-Score is the harmonic mean of precision and recall, providing a single metric that balances both measures. It is particularly useful when dealing with imbalanced datasets, as it penalises extreme trade-offs between precision and recall.

The F1-Score is defined as:

$$\textbf{F1-Score} = 2 \times \frac{Precision \; x \; Recall}{Precision + Recall}$$

By combining Precision and Recall into a single value, the F1-Score provides a more comprehensive measure of classification performance, particularly in multi-class disease detection problems.

In this study, the F1-Score was used to evaluate per-class performance and to provide a more nuanced assessment of classifier effectiveness than accuracy alone. Reporting both macro-averaged and weighted F1-Scores allowed for evaluation of overall model behaviour while accounting for differences in class support.

5.2.6 Confusion Matrix:

The confusion matrix provides a detailed breakdown of classification results by showing the number of correct and incorrect predictions for each class. Each row of the matrix represents the true class labels, while each column represents the predicted class labels.

Confusion matrices were used extensively in this study to analyse misclassification patterns between disease categories. This qualitative insight enabled identification of systematic errors, such as confusion between visually similar diseases or misclassification of healthy leaves as diseased. Such analysis complements numerical metrics and supports a deeper understanding of model limitations.

5.3   Baseline Classification Results (No Denoising):

Baseline results were obtained using resized RGB images (48x48), flattened and standardised, without any denoising. These results establish a benchmark against which the effectiveness of preprocessing techniques can be evaluated.

Model	Accuracy
SVM (GridSearchCV RBF kernel)	0.62
Random Forest	0.56
K-Nearest Neighbour (KNN)	0.55
Gradient Boosting	0.55
Logistic Regression	0.54
SVM (Linear kernel)	0.49
SVM (RBF kernel)	0.48
Naive Bayes	0.36

Table 5.1: Baseline Accuracy of Classification Models

Here's a visual representation of the comparison for all the classification models accuracy in descending order,
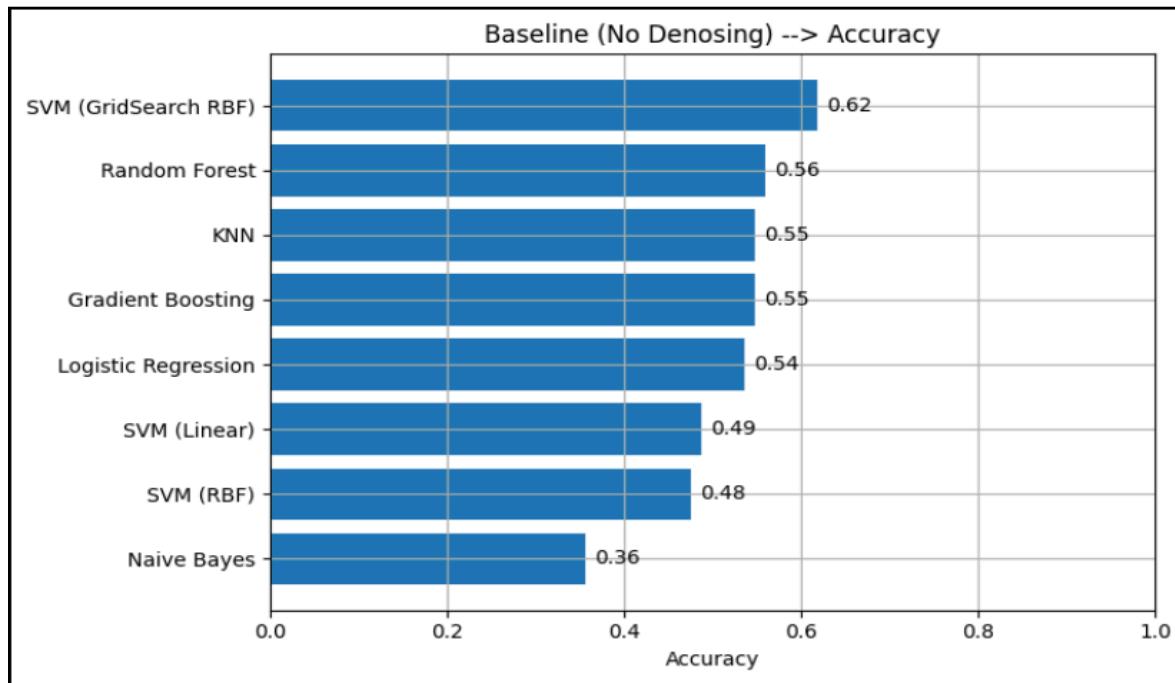
Figure 5.1: Baseline Accuracy of Classification Models

5.3.1 Support Vector Machine (SVM):

Expectation:
Given the non-linear nature of plant disease patterns, it was expected that SVM with an RBF kernel would outperform linear SVM.

Observed Results:
1. SVM with linear kernel achieved 0.481, confirming underfitting
2. Default RBF kernel SVM slightly underperformed at 0.476
3. Tuned RBF kernel SVM (GridSearchCV) achieved the highest baseline accuracy of 0.619

Analysis:
The poor performance of default RBF compared to the tuned version highlights the sensitivity of SVM to hyperparameter C and γ. Once optimised via GridSearchCV, the RBF kernel effectively captured non-linear disease patterns, making it the strongest baseline classifier.

SVM (Linear)				
	Precision	**Recall**	**F1-score**	**Support**
Apple Rot Leaves	0.64	0.50	0.56	28
Healthy Leaves	0.31	0.57	0.40	7
Leaf Blotch	0.47	0.53	0.50	15
Scab Leaves	0.47	0.44	0.45	34
Accuracy			0.49	84
Macro avg	0.47	0.51	0.48	84
Weighted avg	0.51	0.49	0.49	84

Table 5.2: Classification Report for SVM (Linear) with No Denoising
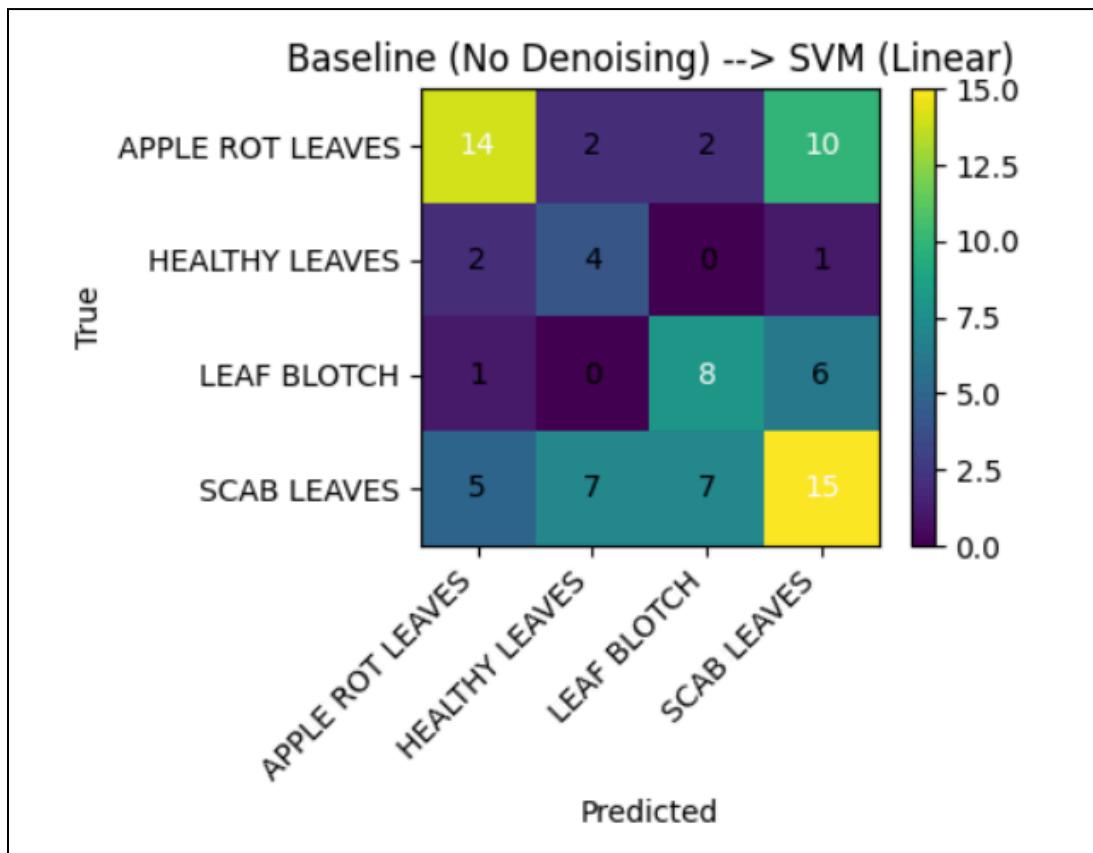
Figure 5.2: Confusion Matrix for SVM (Linear) with No Denoising

SVM (RBF)			
	Precision	**Recall**	**F1-score**	**Support**
Apple Rot Leaves	0.86	0.21	0.34	28
Healthy Leaves	0.00	0.00	0.00	7
Leaf Blotch	0.50	0.20	0.29	15
Scab Leaves	0.44	0.91	0.60	34
Accuracy			0.48	84
Macro avg	0.45	0.33	0.31	84
Weighted avg	0.55	0.48	0.41	84

Table 5.3: Classification Report for SVM (RBF) with No Denoising
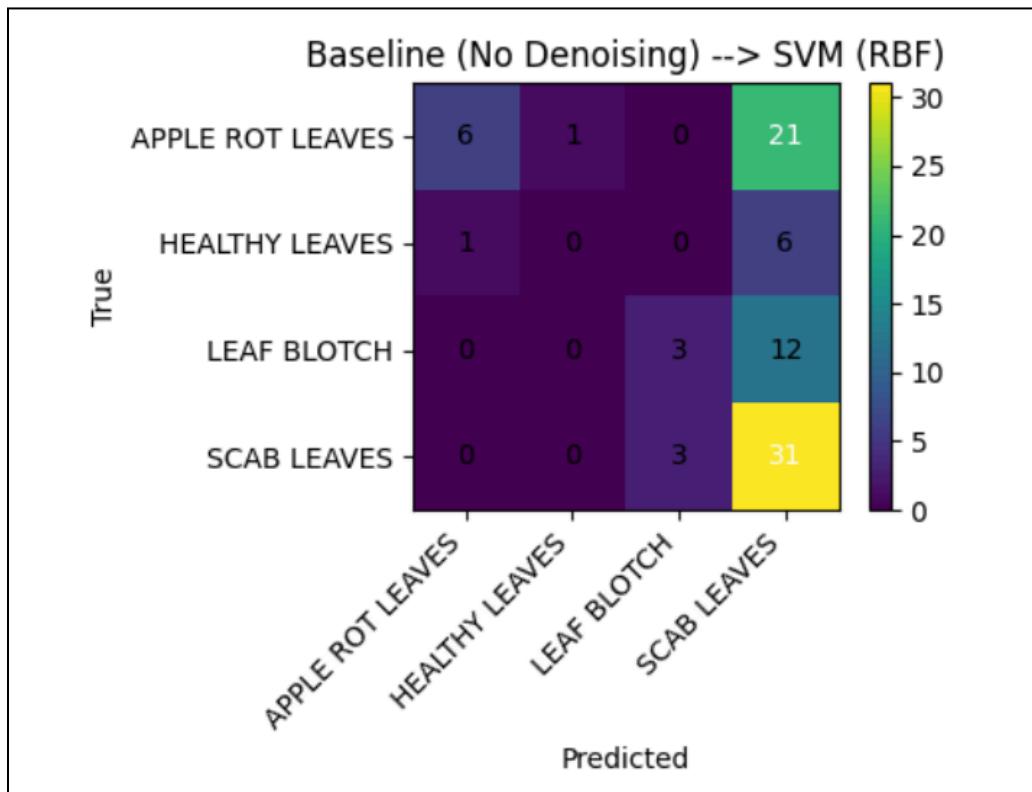
Figure 5.3: Confusion Matrix for SVM (RBF) with No Denoising

SVM (GridSearchCV RBF)				
	Precision	**Recall**	**F1-score**	**Support**
Apple Rot Leaves	0.73	0.57	0.64	28
Healthy Leaves	0.50	0.71	0.51	7
Leaf Blotch	0.60	0.80	0.69	15
Scab Leaves	0.59	0.56	0.58	34
Accuracy			0.62	84
Macro avg	0.61	0.66	0.62	84
Weighted avg	0.63	0.62	0.62	84

Table 5.4: Classification Report for SVM (GridSearchCV RBF) with No Denoising
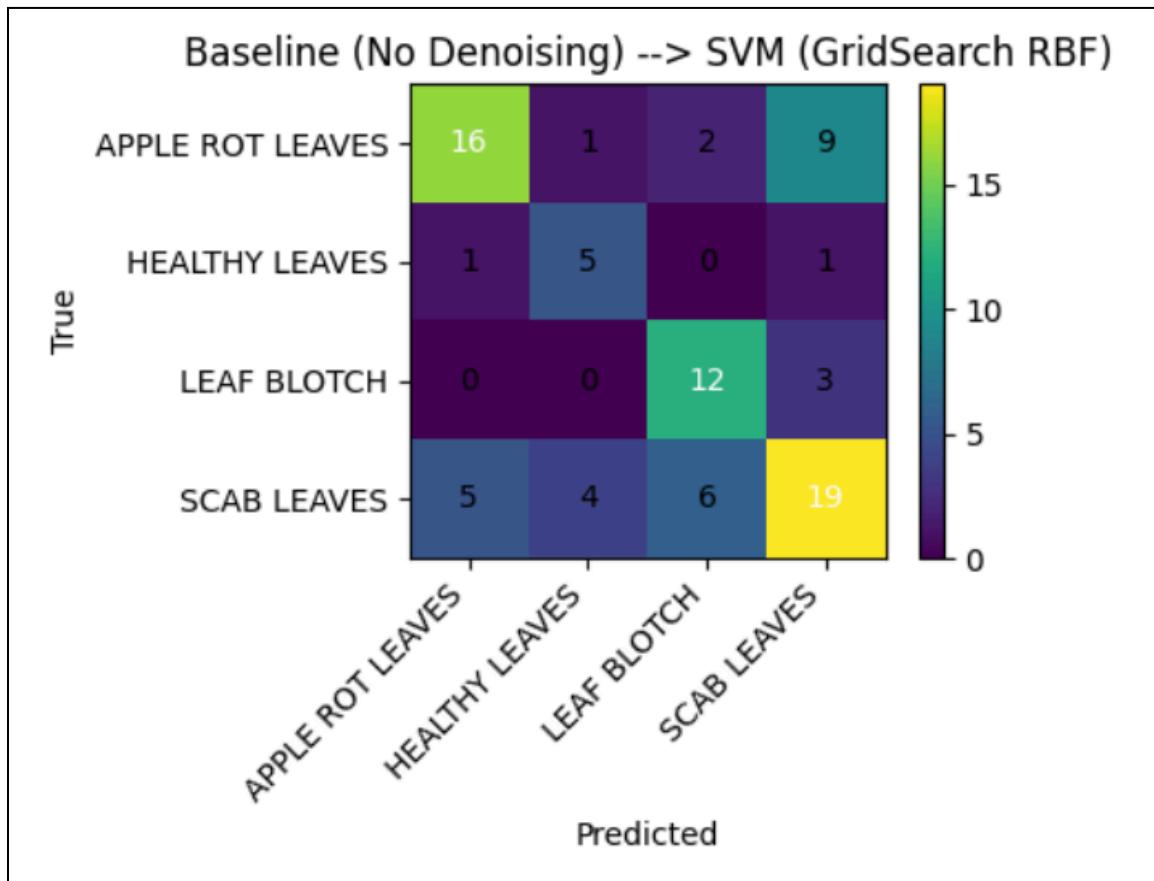
Figure 5.4: Confusion Matrix for SVM (GridSearchCV RBF) with No Denoising

5.3.2 Logistic Regression:

Expectation:
As a linear classifier, Logistic Regression was expected to perform similarly to linear SVM and struggle with complex visual patterns.

Observed Results:
Accuracy of 0.536

Analysis:
While Logistic Regression outperformed linear SVM, its linear decision boundary still limited it. It performed reasonably on Healthy Leaves but showed confusion between disease classes with similar lesion textures.

Logistic Regression				
	Precision	**Recall**	**F1-score**	**Support**
Apple Rot Leaves	0.74	0.50	0.60	28
Healthy Leaves	0.56	0.71	0.62	7
Leaf Blotch	0.44	0.53	0.48	15
Scab Leaves	0.47	0.53	0.50	34
Accuracy			0.54	84
Macro avg	0.55	0.57	0.55	84
Weighted avg	0.56	0.54	0.54	84

Table 5.5: Classification Report for Logistic Regression with No Denoising
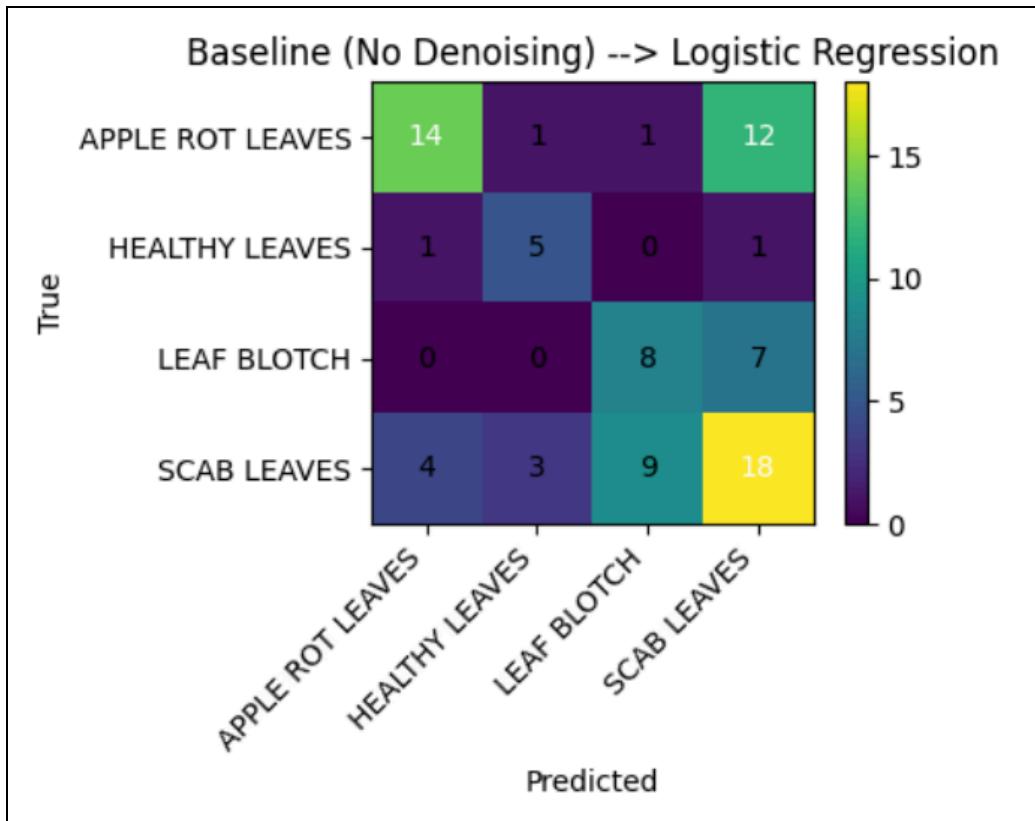
Figure 5.5: Confusion Matrix for Logistic Regression with No Denoising

5.3.3 K-Nearest Neighbours (KNN):

Expectation:
KNN was expected to suffer from high dimensionality and noise sensitivity.

Observed Results:
Accuracy of 0.548

Analysis:
KNN performed moderately but inconsistently. The confusion matrix shows frequent misclassification between Apple Rot and Scab Leaves, indicating that Euclidean distance in pixel space is not robust for subtle disease patterns.

K-Nearest Neighbour (KNN)			
	Precision	**Recall**	**F1-score**	**Support**
Apple Rot Leaves	0.76	0.46	0.58	28
Healthy Leaves	0.50	0.57	0.53	7
Leaf Blotch	0.43	0.60	0.50	15
Scab Leaves	0.53	0.59	0.56	34
Accuracy			0.55	84
Macro avg	0.55	0.56	0.54	84
Weighted avg	0.59	0.55	0.55	84

Table 5.6: Classification Report for KNN with No Denoising
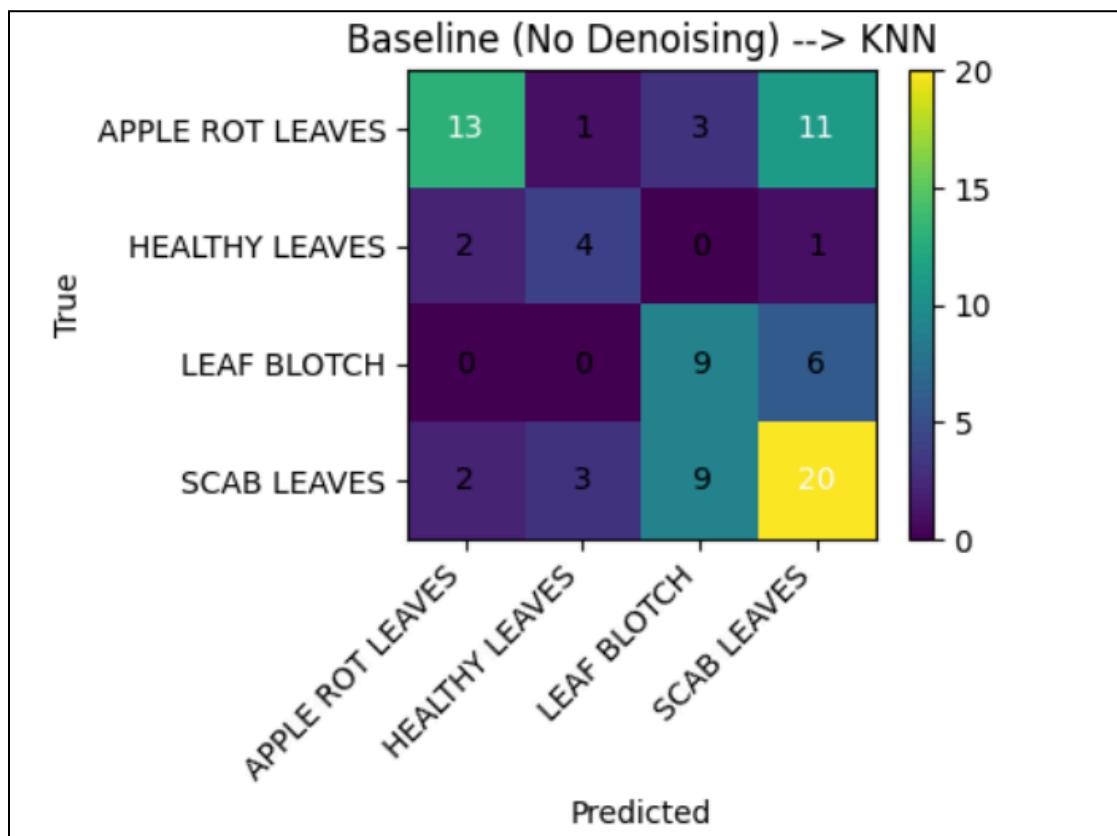
Figure 5.6: Confusion Matrix for KNN with No Denoising

5.3.4 Random Forest:

Expectation:
Random Forest was expected to handle non-linearity and noise better than linear models.

Observed Results:
Accuracy of 0.560

Analysis:
Random Forest achieved stable but not superior performance. While ensemble voting improved robustness, raw pixel features limited its ability to learn fine-grained texture differences.

Random Forest				
	Precision	**Recall**	**F1-score**	**Support**
Apple Rot Leaves	0.72	0.46	0.57	28
Healthy Leaves	0.42	0.71	0.53	7
Leaf Blotch	0.50	0.67	0.57	15
Scab Leaves	0.56	0.56	0.56	34
Accuracy			0.56	84
Macro avg	0.55	0.60	0.56	84
Weighted avg	0.59	0.56	0.56	84

Table 5.7: Classification Report for Random Forest with No Denoising
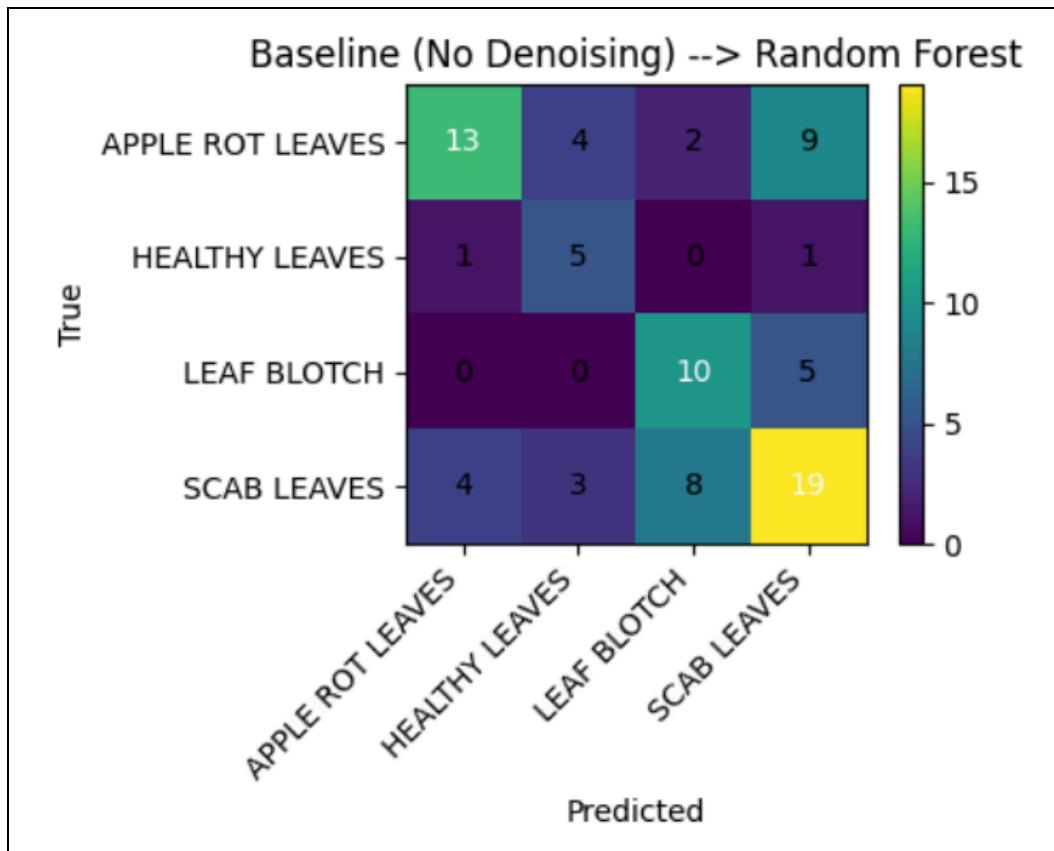
Figure 5.7: Confusion Matrix for Random Forest with No Denoising

5.3.5 Gradient Boosting:

Expectation:
Gradient Boosting can outperform Random Forest when strong weak learners exist.

Observed Results:
Accuracy of 0.548

Analysis:
The model struggled due to the lack of discrimination handcrafted features. Boosting amplified weak pixel-level patterns but did not significantly improve separability.

Gradient Boosting				
	Precision	Recall	F1-score	Support
Apple Rot Leaves	0.74	0.50	0.60	28
Healthy Leaves	0.50	0.71	0.59	7
Leaf Blotch	0.46	0.40	0.43	15
Scab Leaves	0.50	0.62	0.55	34
Accuracy			0.55	84
Macro avg	0.55	0.56	0.54	84
Weighted avg	0.57	0.55	0.55	84

Table 5.8: Classification Report for Gradient Boosting with No Denoising
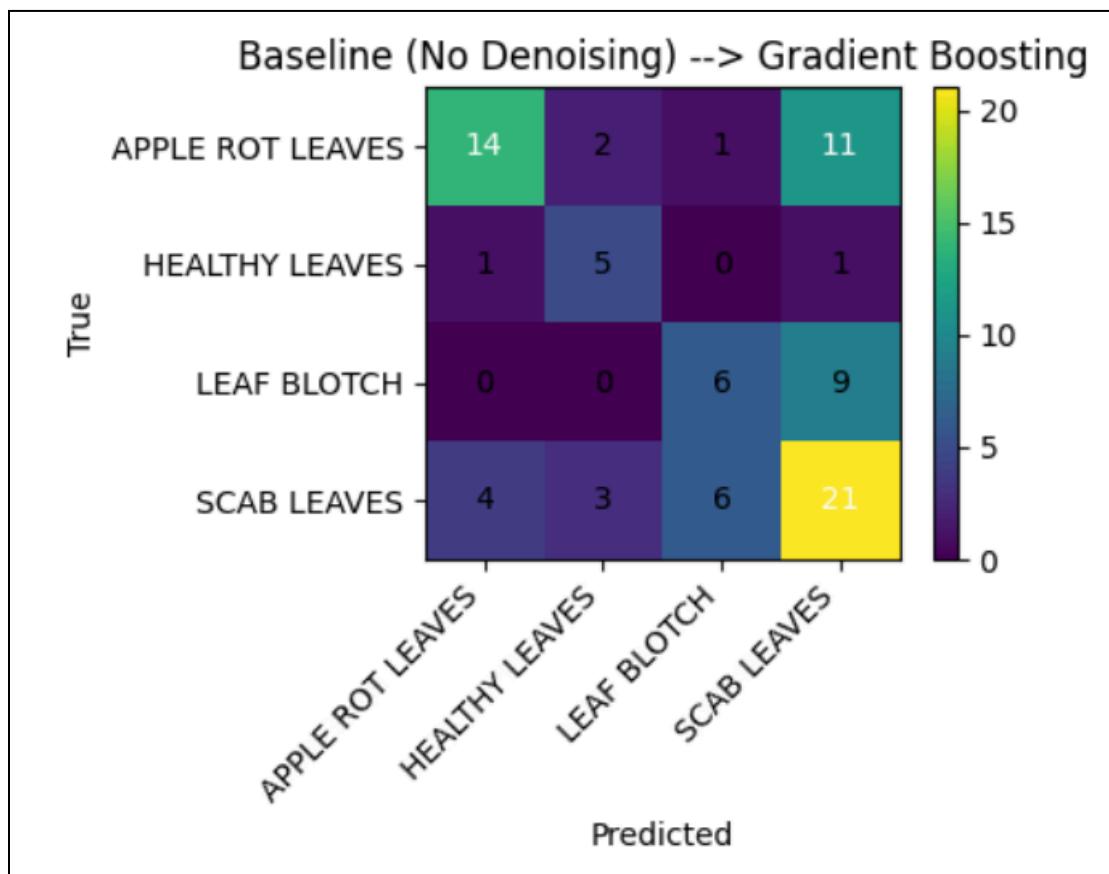

Figure 5.8: Confusion Matrix for Gradient Boosting with No Denoising

5.3.6 Naive Bayes:

Expectation:
Naive Bayes was expected to underperform due to its strong feature independence assumption.

Observed Results:
Lowest accuracy of 0.357

Analysis:
Pixel intensities are highly correlated spatially, violating the Naive Bayes assumption. Confusion matrices show near random predictions for some classes, confirming its unsuitability for this task.

Naive Bayes				
	Precision	**Recall**	**F1-score**	**Support**
Apple Rot Leaves	0.58	0.25	0.35	28
Healthy Leaves	0.29	1.00	0.45	7
Leaf Blotch	0.31	0.87	0.46	15
Scab Leaves	0.50	0.09	0.15	34
Accuracy			0.36	84
Macro avg	0.42	0.55	0.35	84
Weighted avg	0.48	0.36	0.30	84

Table 5.9: Classification Report for Naive Bayes with No Denoising
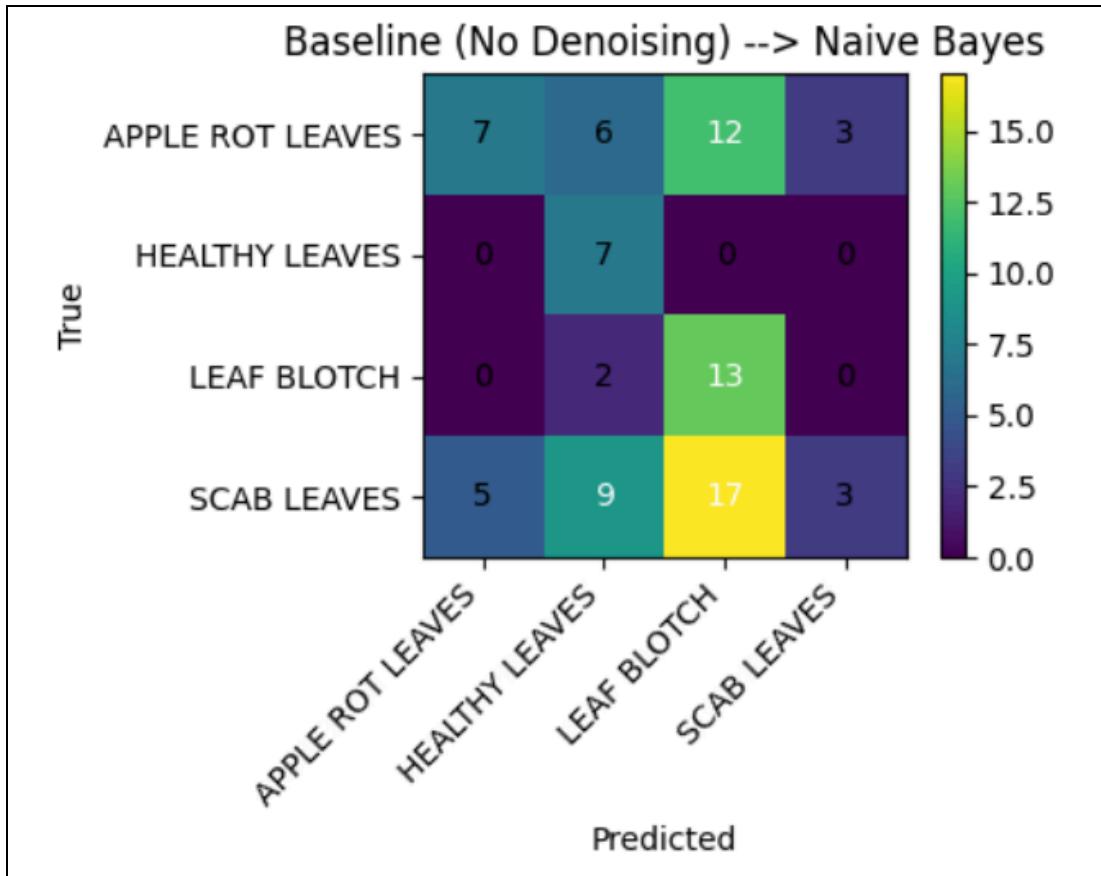
Figure 5.9: Confusion Matrix for Naive Bayes with No Denoising

5.4 Results with Gaussian Denoising:

Gaussian denoising was the first preprocessing technique evaluated due to its simplicity and computational efficiency. Gaussian blur reduces high-frequency noise by smoothing local pixel variations while largely reserving overall image structure.

Expectation:
Prior to experimentation, it was expected that Gaussian denoising would:
1. Reduce irrelevant pixel-level noise
2. Improve generalisation for classifiers sensitive to noise
3. Provide moderate but consistent accuracy improvements across models

Here's a visual representation of the comparison for Gaussian Denoising across all the classification models accuracy in descending order,
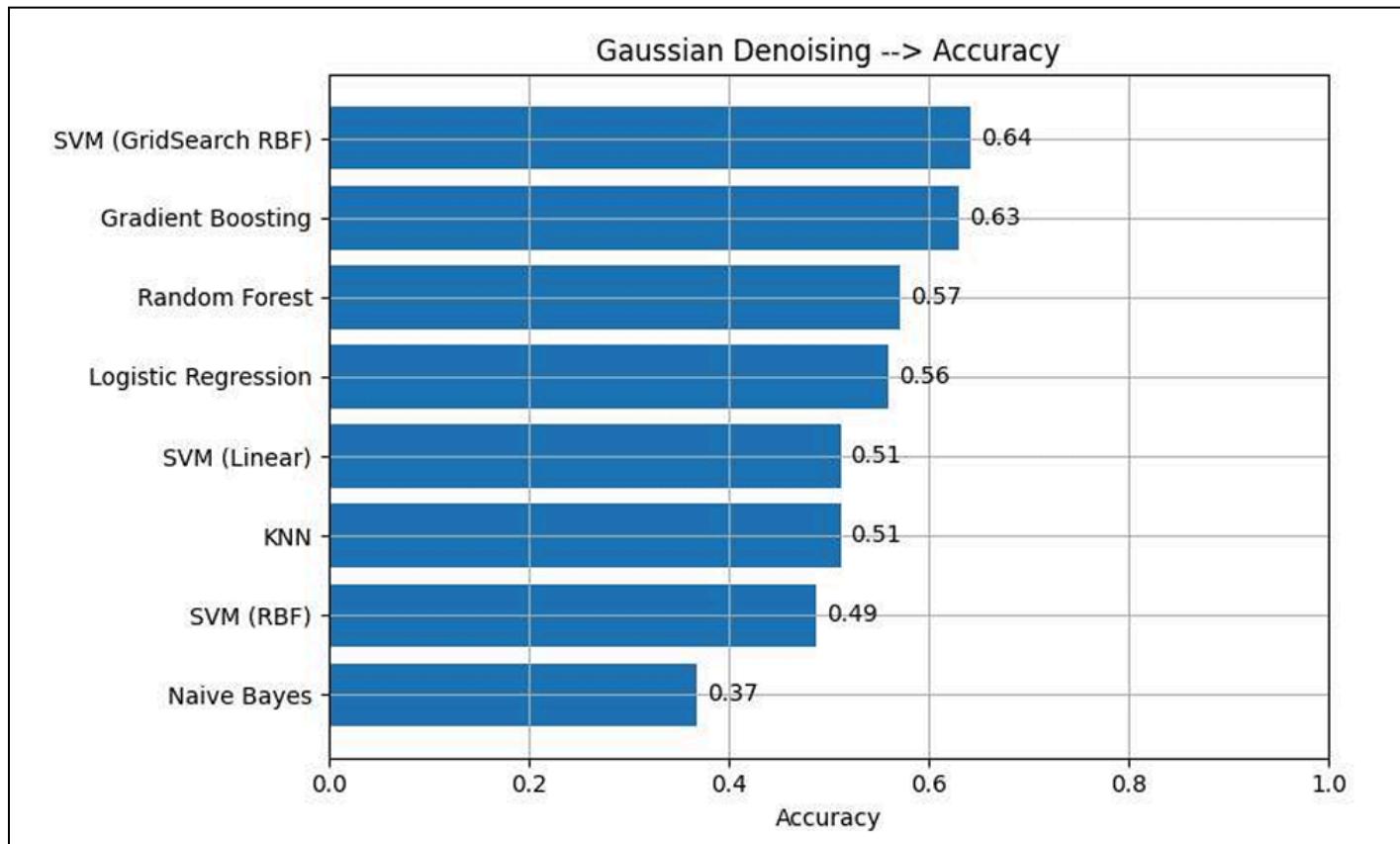

Figure 5.10: Gaussian Denoising Accuracy Across Classifiers

5.4.1 Analysis of Gaussian Denoising Results:

What was expected vs what was observed:
While it was expected that Gaussian denoising would offer general improvements, the results showed that its effectiveness was highly model-dependent. The tuned SVM achieved the strongest performance with RBF kernel, which reached an accuracy of 0.64, outperforming all other classifiers under Gaussian denoising.

Why Gaussian and Tuned SVM (RBF) worked best:
1. Gaussian smoothing reduced high-frequency noise that can distort margin-based classifiers.
2. The RBF kernel allowed the SVM to model non-linear disease patterns effectively.

3. Hyperparameter Optimisation via GridSearchCV identified an appropriate balance between margin width and classification error, enabling better generalisation.
4. Moderate smoothing preserved disease-related textures (spots, lesions) while stabilising the feature distribution.

This confirms that simple denoising combined with careful model tuning can be more effective than complex preprocessing alone.

Why other classifiers did not outperform the tuned SVM:
1. ***Random Forest and Gradient Boosting:*** While both ensemble models benefited from reduced noise, they were limited by the use of raw pixel features, which lack explicit texture abstraction. Gaussian smoothing improved stability but did not provide sufficient discriminative enhancement to surpass the tuned SVM.

2. ***Logistic Regression and Linear SVM:*** These linear models remained constrained by their inability to model complex non-linear decision boundaries, even after noise reduction.

3. ***KNN:*** Distance-based classification remained sensitive to high dimensionality, and smoothing did not significantly improve neighbourhood structure.

4. ***Naive Bayes:*** The independence assumption continued to be violated, and Gaussian denoising did not address this fundamental limitation.

5.5 Results with Median Denoising:

Median denoising replaces each pixel with the median value of its neighbourhood, making it effective for impulse noise removal while preserving edges.

Expectation:
Prior to experimentation, Median Blur was expected to:
1. Preserve sharp lesion boundaries
2. Outperform Gaussian denoising for texture-based disease patterns

Here's a visual representation of the comparison for Gaussian Denoising across all the classification models accuracy in descending order,
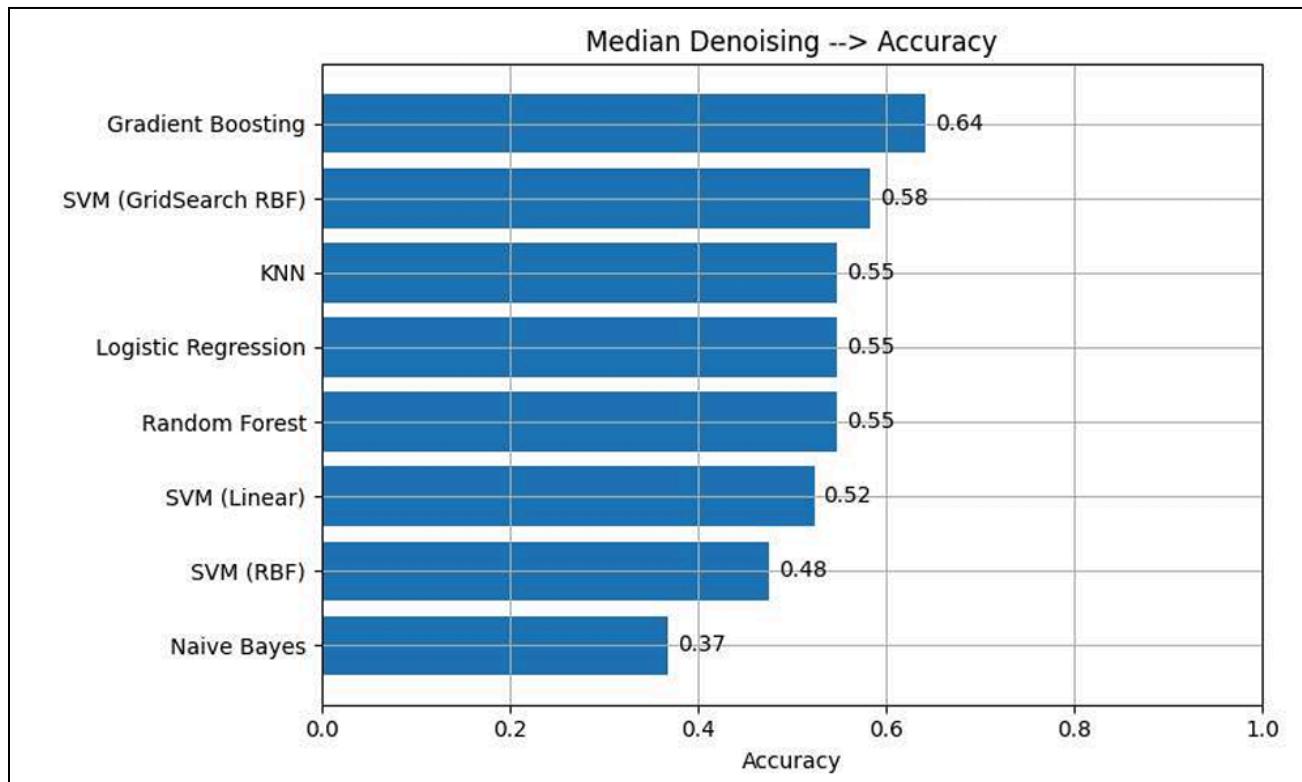

Figure 5.11: Median Denoising Accuracy Across Classifiers

5.5.1 Analysis of Median Denoising Results:

Median denoising unexpectedly favoured Gradient Boosting, which achieved an accuracy of 0.64, matching the highest score observed across all experiments. This indicates that median filtering created a representation well-suited to boosting's sequential error-correction mechanism.

However, median filtering disrupted continuous texture patterns, which are critical for disease such as apple scab and leaf blotch. These disruptions negatively impacted SVM performance, particularly when compared to Gaussian smoothing. This result highlights that edge preservation alone is insufficient when disease discrimination relies on distributed texture rather than sharp boundaries.

The clustering of multiple models around similar accuracies suggests that median filtering reduces both noise and discriminative signal, resulting in a flattened feature space.

5.6 Results with Bilateral Denoising:

Bilateral filtering performs edge-preserving smoothing by combining spatial proximity and pixel-intensity similarity.

Expectation:
Prior to experimentation, Bilateral Blur was expected to:
1. Preserve disease boundaries,
2. Maintain structural integrity of lesions,
3. Outperform Gaussian and Median denoising by balancing smoothing and edge preservation

Observation of Results:
The classification results obtained after Bilateral denoising are summarised in the table below:

Here's a visual representation of the comparison for Gaussian Denoising across all the classification models accuracy in descending order,
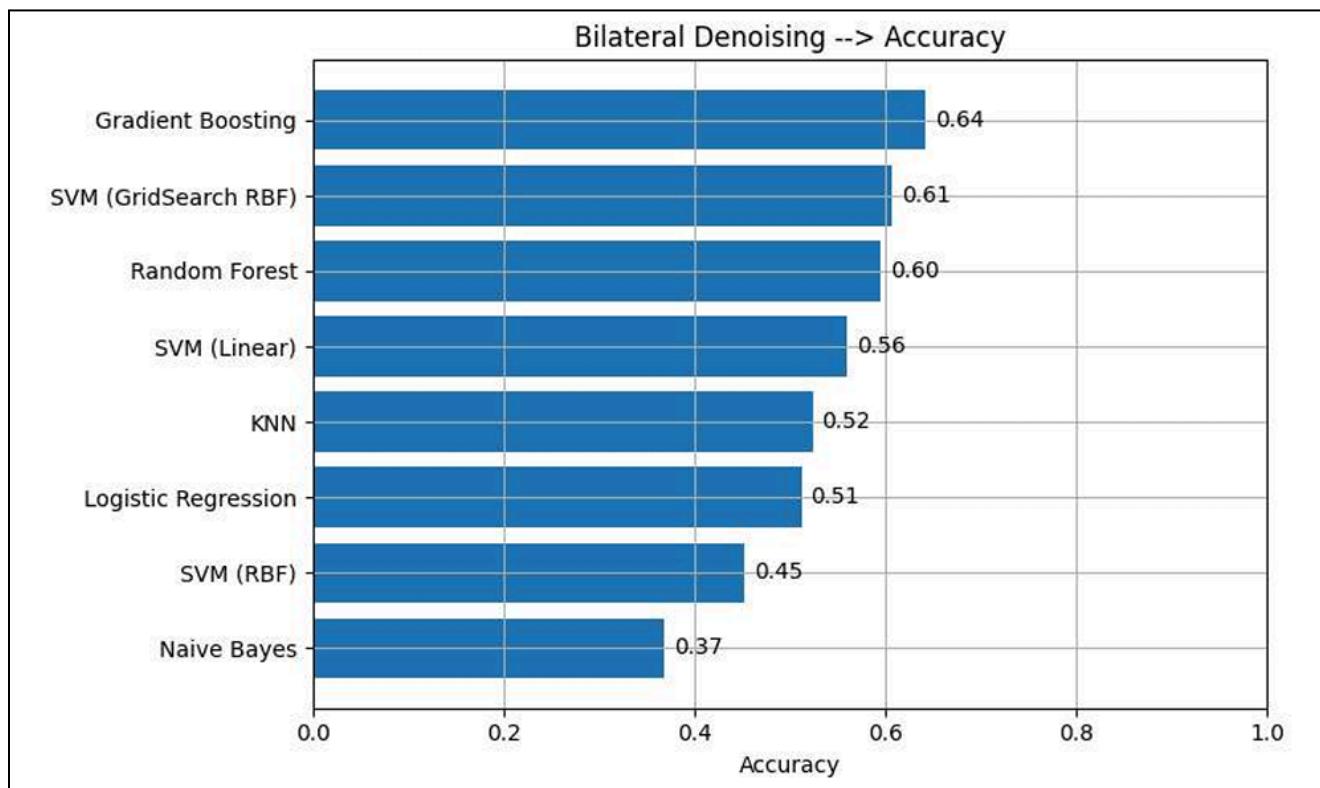

Figure 5.12: Bilateral Denoising Accuracy Across Classifiers

5.6.1 Analysis of Bilateral Denoising Results:

Bilateral filtering produced non-uniform smoothing, altering pixel distribution inconsistently across images. This negatively affected margin-based classifiers such as SVM and did not lead to improvements over Gaussian denoising. Ensemble models were more robust to these distortions but still failed to surpass the best Gaussian denoised results.

Gradient Boosting again emerged as the strongest classifier, demonstrating robustness to heterogeneous feature transformations. Random forest also benefited moderately due to its tolerance to feature perturbations.

Despite preserving edges, bilateral denoising did not outperform Gaussian smoothing overall, indicating that texture continuity is more important than sharp boundaries for apple leaf disease classification in low-resolution images.

5.7 Results with Non-Local Means (NLMeans) Denoising:

NLMeans denoising reduces noise by averaging similar image patches across the image.

Expectation:
Prior to experimentation, Non-Local Means (NLMeans) was expected to:
1. Preserve fine texture patterns,
2. Outperform simpler denoising methods due to its patch-based nature,
3. Improve classification of texture-driven disease patterns.

Here's a visual representation of the comparison for Gaussian Denoising across all the classification models accuracy in descending order,
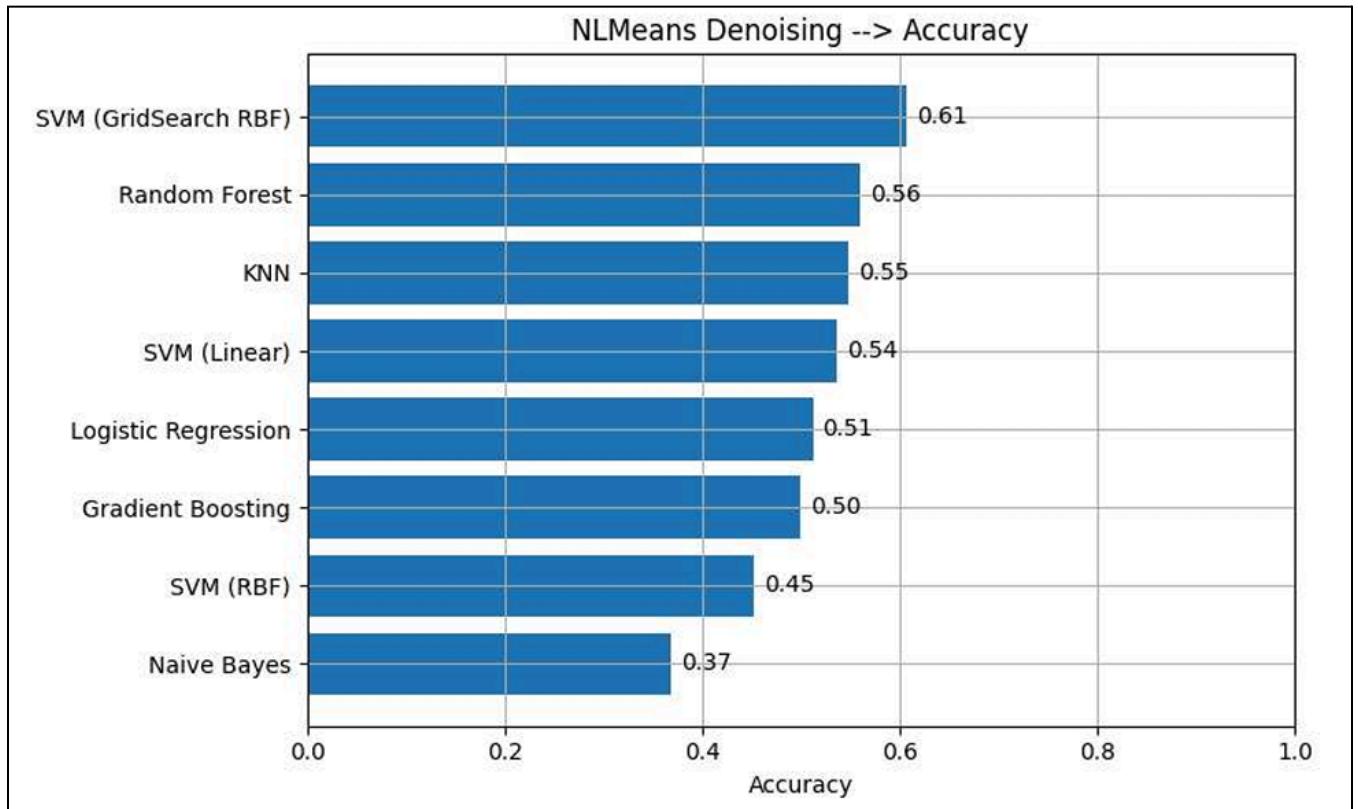
Figure 5.13: NLMeans Denoising Accuracy Across Classifiers

5.7.1 Analysis of Non-Local Means (NLMeans) Denoising Results:

Contrary to expectations, NLMeans consistently reduced classification performance. The primary reason lies in the low image resolution (48x48). At this scale, patch similarity becomes unreliable, and disease textures are easily smoothed out alongside noise.

Although the tuned SVM (RBF) remained the strongest model under NLMeans, its accuracy was significantly lower than under Gaussian denoising. Additionally, the computational cost of NLMeans makes it unsuitable for real-time agriculture applications.

5.8 Results with Kernel PCA Reconstruction:

Kernel PCA was evaluated as a reconstruction-based denoising technique, aiming to remove noise by projecting features into a lower-dimensional manifold and reconstructing them.

Expectation:
Prior to experimentation, Kernel PCA was expected to:
1. Reduce high-dimensional noise
2. Improve generalisation by filtering irrelevant variance,
3. Benefit classifiers prone to overfitting
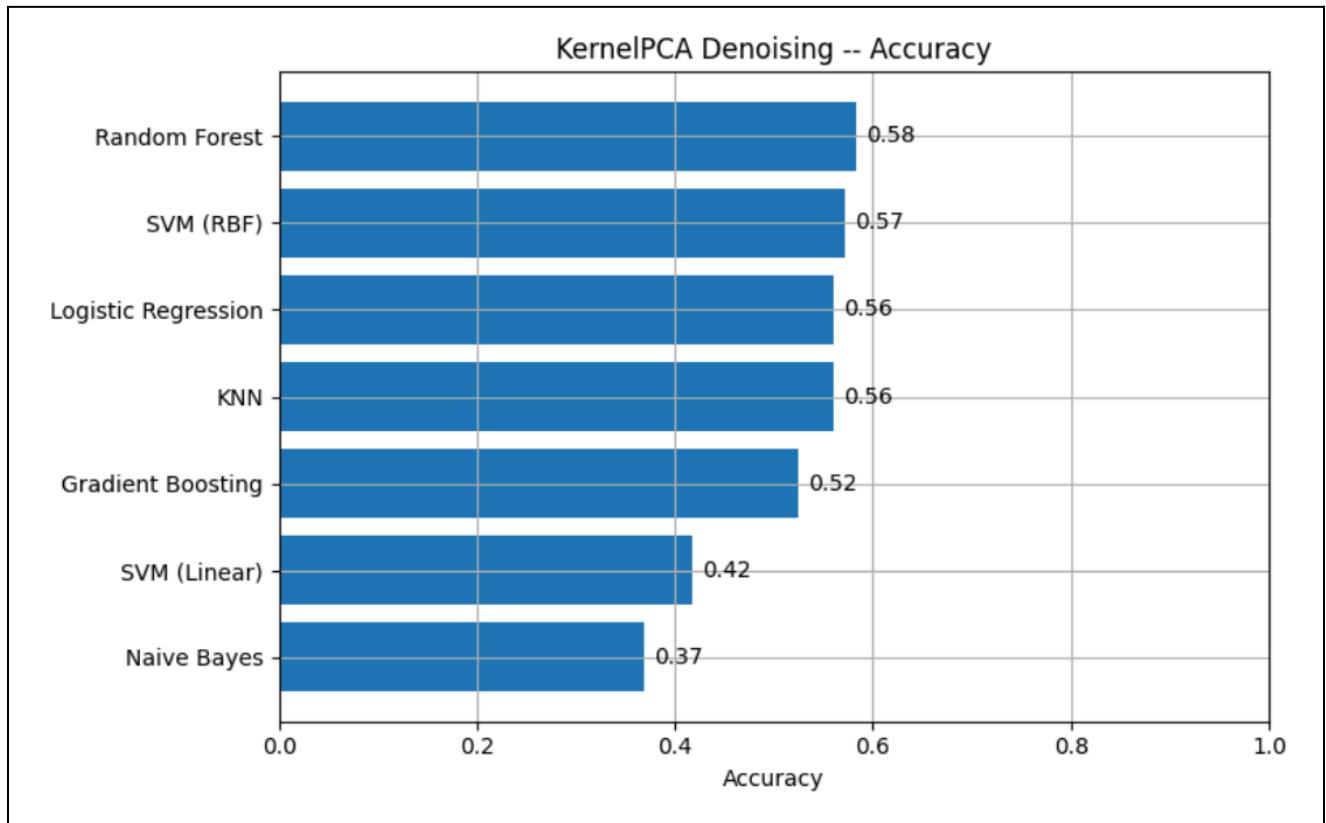
Figure 5.14: Kernel PCA Reconstruction Denoising Accuracy Across Classifiers

5.8.1 Analysis of Kernel PCA Reconstruction Results:

Kernel PCA reconstruction removed both noise and disease-related variance, leading to reduced separability between classes. Disease patterns do not lie on a simple low-dimensional manifold, making reconstruction-based denoising ineffective in this context.

Random Forest achieved the best result under Kernel PCA due to its ability to exploit coarse patterns that remained after reconstruction. However, overall performance was inferior to simpler denoising techniques, reinforcing the conclusion that aggressive feature compression can harm classification accuracy.

5.9 Grand Comparative Analysis of Classifiers and Denoising Techniques:

After evaluating each denoising technique independently, a grand comparative analysis was conducted to assess overall trend across all classifiers and preprocessing methods. The objective of this comparison is to identify the most effective combinations, understand model-preprocessing interactions, and determine whether denoising techniques consistently improves classification performance.

5.9.1 Consolidated Accuracy Comparison:

	Original	Gaussian	Median	Bilateral	NLMeans	Kernel PCA
Gradient Boosting	0.55	0.63	0.64	0.64	0.50	0.52
KNN	0.55	0.51	0.55	0.52	0.55	0.56
Logistic Regression	0.53	0.56	0.55	0.51	0.51	0.56
Naive Bayes	0.36	0.37	0.37	0.37	0.37	0.37
Random Forest	0.56	0.57	0.55	0.59	0.56	0.58
SVM(GridSearchCV)	0.62	0.64	0.58	0.61	0.61	NaN
SVM (Linear)	0.49	0.51	0.52	0.56	0.53	0.42
SVM (RBF)	0.48	0.49	0.48	0.45	0.45	0.57

Table 5.10: Accuracy of all Classifiers Across
All Denoising Techniques

Here's a visual representation of the accuracy of all classifiers across all denoising techniques,
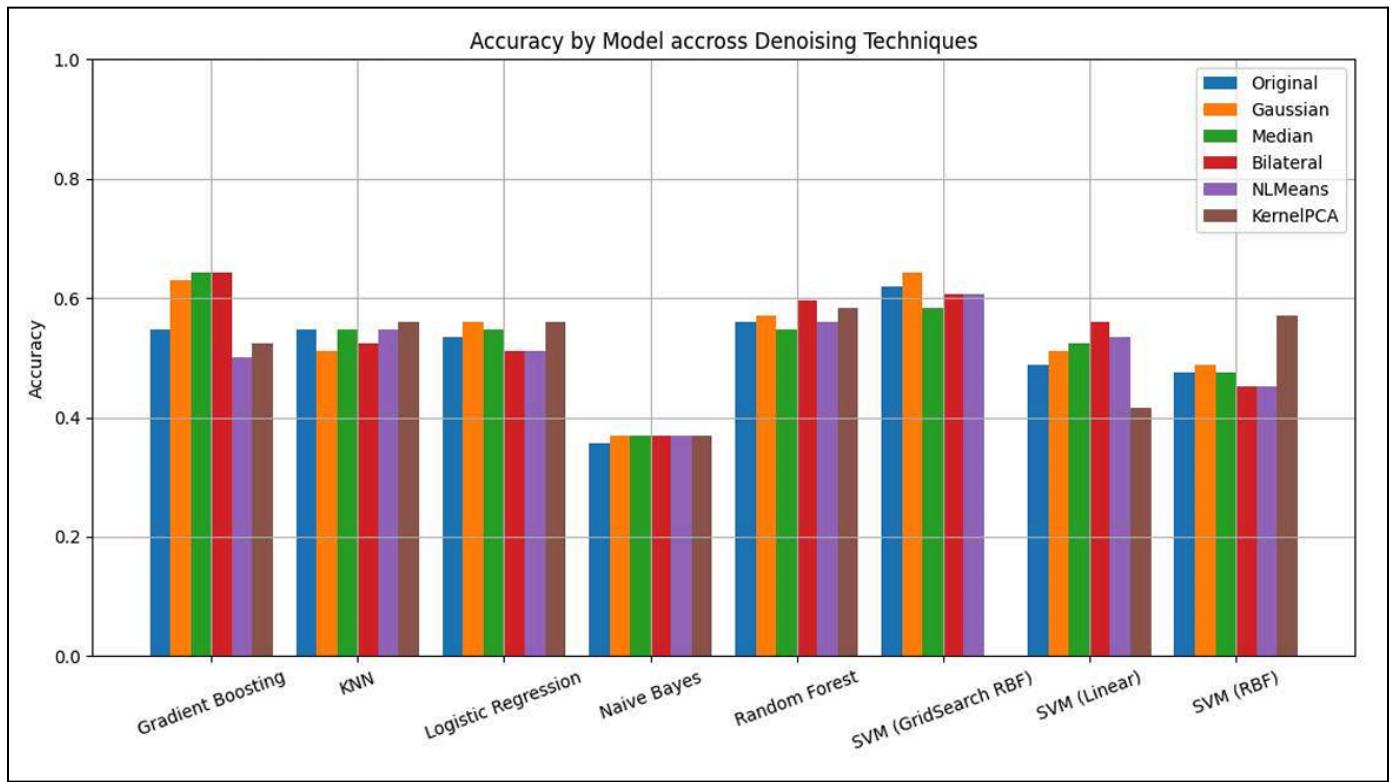

Figure 5.15 : Accuracy of all Classifiers Across All Denoising Techniques

5.10 Findings from the Grand Comparative Analysis:

The grand comparative analysis of classifiers and denoising techniques reveals several important findings regarding the interaction between preprocessing and classical machine learning models for apple disease classification. These findings provide insight into why certain pipelines consistently outperform others and clarify the conditions under which denoising improves or degrades classification performance.

A key finding from the grand comparison is that no single classifier dominates across all preprocessing strategies. While the tuned Support Vector Machine(SVM) with the RBF kernel and Gradient Boosting models frequently achieved the highest accuracies, their relative performance varied depending on the denoising technique applied.
1. The tuned SVM with RBF kernel achieved the best performance under no denoising, Gaussian denoising, and NLMeans denoising
2. Gradient Boosting achieved the best performance under Median and Bilateral denoising
3. Random Forest emerged as the strongest under Kernel PCA Reconstruction.

This variation demonstrates that classifier performance is highly sensitive to the statistical properties of the feature space produced by preprocessing. Consequently, classifier selection cannot be decoupled from preprocessing design, particularly when working with raw pixel features.

Another significant finding is that simple denoising techniques consistently outperformed more complex methods. Gaussian, Median, and Bilateral denoising all achieved peak accuracies of approximately 0.64, whereas complex methods such as NLMeans, and Kernel PCA failed to surpass these results. This outcome contradicts the intuitive expectation that more advanced denoising techniques should yield superior performance. Instead, the experiments indicate that:
1. Simple smoothing methods effectively reduce high-frequency noise while preserving discriminative disease patterns.
2. Complex denoising techniques tend to remove subtle texture variation that are essential for distinguishing visually similar disease classes.

These findings highlight the importance of matching preprocessing complexity to dataset characteristics, particularly image resolution and texture density.

5.10.1 The Most Stable Preprocessing Method:

Among all denoising techniques evaluated, Gaussian denoising emerged as the most stable and reliable across classifiers. It consistently improved or maintained performance relative to the baseline and produced the highest overall accuracy when combined with tuned SVM.
Gaussian denoising demonstrated:
1. Minimal performance degradation across models,
2. Low computational overhead,
3. Consistent improvement in feature stability.

The stability makes Gaussian denoising particularly suitable for real-time agricultural monitoring systems, where predictable behaviour and efficiency are critical.

5.10.2 Classifier-Denoiser Compatibility and Effects:

The grand comparison revealed clear compatibility effects between classifiers and denoising techniques:
1. Gradient Boosting benefited most from Median and Bilateral denoising, which preserved coarse structural information while removing impulsive noise.
2. SVM (RBF) benefited most from Gaussian denoising, which preserved continuous texture distributions necessary for margin-based optimisation.
3. Random Forest was comparatively more resilient under Kernel PCA Reconstruction, suggesting that tree-based models can exploit coarse features even when fine-grained information is lost.

These interactions confirm that preprocessing alters the geometry of the feature space in ways that favour different learning mechanisms.

5.10.3 Limitations of Advanced Denoising Techniques:

Despite their theoretical advantages, NLMeans and Kernel PCA consistently underperformed in the context of this study. Two primary factors explain this behaviour:

1. **Low Image Resolution (48x48):** At this resolution, patch-based similarity (NLMeans) and manifold reconstruction (Kernel PCA) become unreliable. Disease-specific textures are easily smoothed out along with noise.
2. **Loss of Discriminative Variance:** Both techniques remove variance that may be visually subtle but semantically important for disease differentiation, leading to reduced class separability.

This finding reinforces an important practical lesson: advanced preprocessing does not guarantee improved classification performance, particularly when applied indiscriminately.

5.11 Selecting the Best Technique and Recreating the Denoised Dataset:

5.11.1 Rationale for Selecting a Best Preprocessing Technique:

Following the grand comparative evaluation of classifiers and denoising techniques, a single preprocessing strategy was selected to construct the final classification pipeline. The purpose of this stage was to transition from exploratory benchmarking to the development of a consolidated and optimised model configuration that could represent the final outcome of the AgriHealth AI system.

Based on the comparative results, Gaussian denoising was selected as the most suitable preprocessing technique. This decision was guided by two main factors:

1. Gaussian denoising consistently achieved one of the highest classification accuracies during benchmarking
2. It is computationally lightweight, making it appropriate for real-time agricultural monitoring applications.

The selection of Gaussian denoising therefore reflects both empirical performance and practical deployment consideration.

5.11.2 Recreating the Best Denoised Dataset:

After selecting Gaussian denoising as the optimal preprocessing method, the entire dataset was reprocessed using this technique to create a clean and consistent feature representation for final model training and evaluation. All image samples were denoised using Gaussian smoothing and transformed into a unified feature matrix, ensuring that every instance underwent identical preprocessing. This step eliminated the possibility of mixing feature representations from earlier comparative experiments and formalised the preprocessing pipeline.

The denoised dataset was then partitioned into training and testing subsets using an 80/20 split, with stratification applied to preserve the original class distribution across both subsets. Maintaining stratification was particularly important given the class imbalance present in the dataset, as it ensured fair evaluation across disease categories. A fixed random state was used to ensure reproducibility and consistency with earlier experiments.

Prior to classification, feature standardisation was applied using z-score normalisation. This step ensured that all pixel-based features contributed equally during model training and prevented features with larger numerical ranges from disproportionately influencing the learning process. Feature scaling was especially critical for kernel-based methods such as Support Vector Machine (SVM), which are highly sensitive to feature magnitude when computing distances in the transformed kernel space.

Together, these steps resulted in a fully reconstructed dataset that reflected the selected best preprocessing technique and provided a stable foundation for final model optimisation.

5.11.3 Final Optimised Model on the Selected Technique:

Using the recreated Gaussian-denoised dataset, a tuned Support Vector Machine (SVM) with a Radius Basis Function (RBF) kernel was trained and evaluated. Hyperparameter Optimisation had previously identified suitable values for the regularisation parameter and kernel width, enabling the model to effectively balance bias and variance under the transformed feature distribution.

Evaluation on the test set yielded a final classification accuracy of 0.69, representing the strongest performance achieved throughout the study. This improvement is notable when compared to earlier benchmarking results, where the best accuracy under Gaussian denoising was approximately 0.64. The observed gain indicates that committing to a single best preprocessing strategy and rebuilding the dataset accordingly can lead to improved generalisation performance.

The increase in accuracy can be attributed to three key factors,
1. Focused optimisation reduced variability introduced by repeated transformations across different denoisers
2. Rebuilding the dataset ensured full consistency between preprocessing and model training
3. Hyperparameter tuning became more effective once applied to stable and well-defined feature space. Collectively, these refinements resulted in a more mature and reliable classification pipeline.

5.11.4 Quantitative Results of the Final Model:

Tuned SVM (GridSearchCV + Gaussian)				
	Precision	**Recall**	**F1-score**	**Support**
Apple Rot Leaves	0.71	0.57	0.63	21
Healthy Leaves	0.50	0.44	0.47	9
Leaf Blotch	0.81	0.77	0.79	22
Scab Leaves	0.66	0.78	0.71	32
Accuracy			0.69	84
Macro avg	0.67	0.64	0.65	84
Weighted avg	0.69	0.69	0.69	84

Table 5.11: Classification Report for Tuned SVM (GridSearchCV + Gaussian)
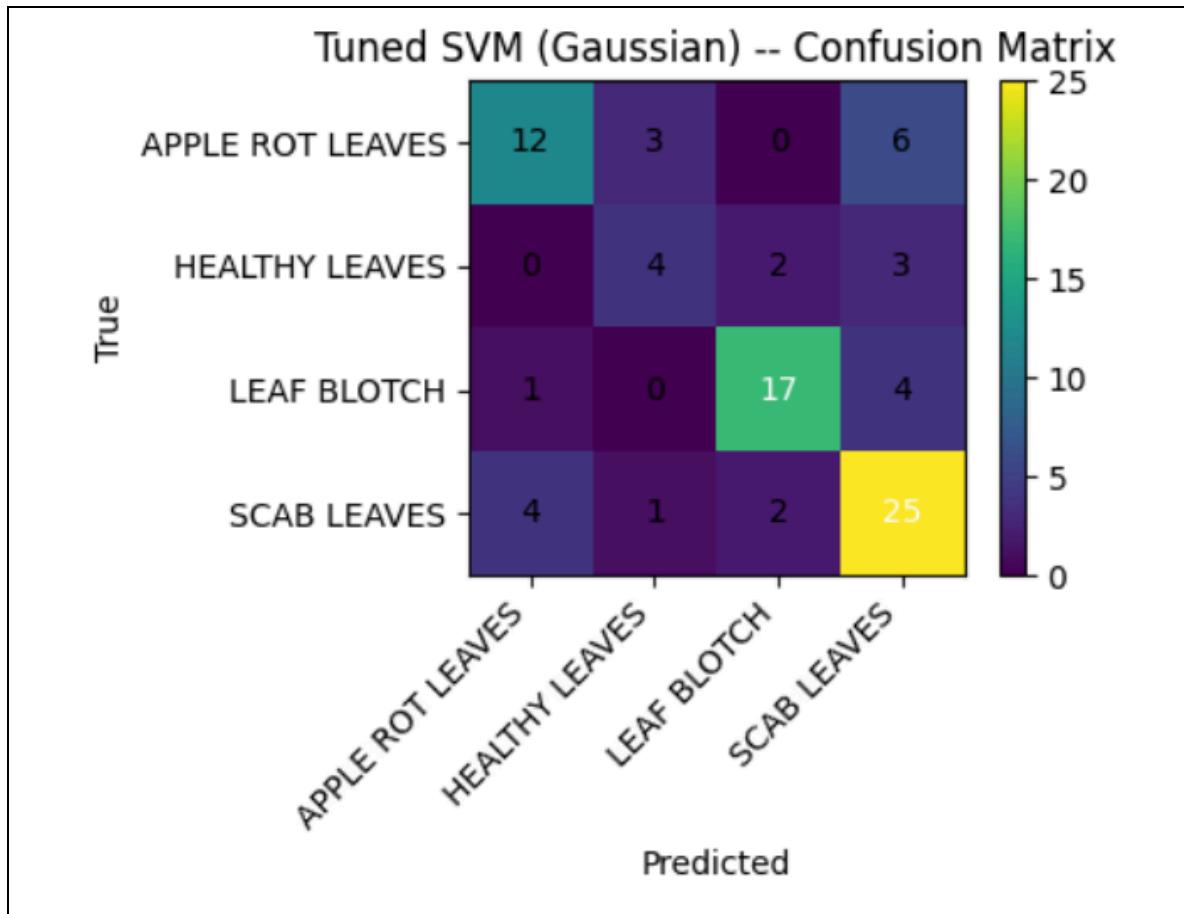
Figure 5.16 : Confusion Matrix for Tuned SVM (GridSearchCV + Gaussian)

6. Conclusion and Future Work:

6.1 Conclusion:

This study set out to investigate the effectiveness of classical machine learning techniques for automated apple leaf disease classification within the context of a real-time plant health monitoring system, that is AgriHealth AI. The primary objective was to evaluate how different image denoising techniques influence classification performance when using pixel-based feature representation and to identify an optimal combination of preprocessing and classification methods suitable for practical deployment.

To achieve this, a comprehensive experimental pipeline was developed using the "Kashmiri Apple Plant Disease Dataset", comprising four classes,

1. Healthy Leaves
2. Scab Leaves
3. Leaf Blotch
4. Apple Rot Leaves

All images were resized, converted into RGB feature representations, flattened into numerical vectors, and standardised prior to classification. Multiple classical machine learning models were systematically evaluated, including Support Vector Machines (SVM) with both Linear and RBF kernels, Logistic Regression, K-Nearest Neighbours, Random Forest, Gradient Boosting, and Naive Bayes.

A key contribution of this work lies in its extensive analysis of denoising techniques. Gaussian denoising, Median denoising, Bilateral denoising, Non-Local Means, and Kernel Principal Component Analysis (Kernel PCA) were applied and assessed in combination with each classifier. The results demonstrated that denoising has a model-dependent and data-dependent impact, rather than providing universal performance gains.

Among all evaluated configurations, Gaussian denoising combined with a tuned Support Vector Machine using an RBF kernel produced the strongest and most stable performance. After selecting Gaussian denoising as the optimal preprocessing technique and recreating the denoised dataset, the final optimised model achieved an accuracy of approximately 0.69, representing the highest performance obtained throughout the study. This result confirms that moderate noise reduction, when paired with appropriate model training, can significantly improve classification reliability without introducing excessive computational overhead.

The study also revealed that more complex denoising techniques, such as NLMeans and Kernel PCA, did not enhance performance and in some cases degraded accuracy. This outcome highlights an important practical insight: advanced preprocessing methods may remove subtle texture information that is critical for distinguishing visually similar plant diseases, particularly when working with low-resolution images. Ensemble models such as Gradient Boosting showed robustness under certain denoising conditions, while Naive Bayes consistently underperformed due to its strong independence assumptions, confirming its unsuitability for raw image-based classification.

Overall, this research demonstrates that carefully designed classical machine learning pipelines remain viable for plant disease classification tasks, especially when computational efficiency and interpretability are prioritised. The findings support the feasibility of integrating such models into real-time agricultural decision-support systems.

6.2 Key Contributions:

The key contributions of this study can be summarised as follows:
1. ***Systematic Evaluation of Classical Classifiers:*** This study provides a detailed comparative assessment of multiple classical machine learning algorithms applied to apple leaf disease classification, highlighting their relative strengths and limitations when using raw pixel-based features.
2. ***Comprehensive Analysis of Denoising Techniques:*** Five denoising and preprocessing methods were evaluated under consistent experimental conditions, offering empirical evidence of how different noise reduction strategies interact with various classifiers.
3. ***Identification of an Optimal Pipeline:*** Gaussian denoising combined with tuned SVM with RBF kernel was identified as the most effective and practical pipeline, achieving the highest observed accuracy while maintaining computational efficiency
4. ***Insight into Model Preprocessing Interactions:*** The results demonstrate that classifier performance is strongly influenced by preprocessing choices, reinforcing the importance of jointly optimising both components rather than treating them independently.
5. ***Foundation for Real-time Agricultural Application:*** By prioritising lightweight preprocessing and classical models, the study establishes a solid foundation for deploying plant disease classification systems in resource-constrained, real-time agricultural environments.

6.3 Limitation of the Study:

While the results are promising, several limitations should be acknowledged. First, the dataset size and image resolution constrained the ability to exploit fine-grained texture information, which may have limited the effectiveness of advanced denoising techniques. Second, the final model selection was based on comparative experimental results rather than a fully nested validation framework, which may introduce a degree of selection bias. Finally, the use of flattened pixel features restricts the model's capacity to capture spatial relationships inherent in leaf disease patterns.

Recognising these limitations is essential for contextualising the findings and guiding future improvements.

6.4 Recommendations for Future Work:

Several directions for future research emerge from this study:

1. Incorporation of Feature Engineering Techniques: Future work could explore texture-based descriptors such as Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP), or colour histograms to capture spatial and structural information more effectively than raw pixels.

2. Adoption of Deep Learning Approaches: Convolutional Neural Network (CNNs) could be investigated to automatically learn hierarchical features from leaf images, potentially improving classification accuracy and robustness, particularly with larger datasets.

3. Dataset Expansion and Augmentation: Expanding the dataset through additional samples or data augmentation techniques could help mitigate class imbalance and improve generalisation.

4. Cross-Dataset Generalisation Studies: Evaluating the trained models on apple disease datasets from different regions acquisition conditions would provide insight into the robustness and transferability of the approach.

5. Real-time System Integration: Future development of AgriHealth AI could focus on integrating the classification pipeline into a web-based or mobile application, enabling farmers to capture images and receive real-time disease diagnostics in the field.

6. Advanced Evaluation Strategies: Employing nested cross-validation or external validation sets would provide more rigorous estimates of generalisation performance and further strengthen experimental reliability.

6.5 Final Remarks:

In conclusion, this research demonstrates that effective apple leaf disease classification can be achieved using classical machine learning techniques when combined with carefully selected preprocessing strategies. The findings emphasise the importance of simplicity, interpretability, and systematic experimentation in applied machine learning projects. By balancing performance with computational efficiency, the proposed AgriHealth AI pipeline provides a meaningful step toward practical, real-time plant health monitoring systems and establishes a strong foundation for future advancements in precision agriculture.

7. References

[1] Elangovan, K. and Nalini, S. (2017) *Plant disease classification using image segmentation and SVM techniques*. International Journal of Computer Science and Mobile Computing, 6(5), pp. 20–29.

[2] Barbedo, J.G.A. (2018) 'Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification', *Computers and Electronics in Agriculture*, 153, pp. 46–53.

[3] Too, E.C., Yujian, L., Njuki, S. and Yingchun, L. (2019) 'A comparative study of fine-tuning deep learning models for plant disease identification', *Computers and Electronics in Agriculture*, 161, pp. 272–279.

[4] Mohanty, S.P., Hughes, D.P. and Salathé, M. (2016) 'Using deep learning for image-based plant disease detection', *Frontiers in Plant Science*, 7, Article 1419.

[5] Ferentinos, K.P. (2018) 'Deep learning models for plant disease detection and diagnosis', *Computers and Electronics in Agriculture*, 145, pp. 311–318.

[6] Barbedo, J.G.A. (2019) 'Plant disease identification from individual lesions and spots using deep learning', *Biosystems Engineering*, 180, pp. 96–107.

[7] Sharma, H., Padha, D. and Bashir, N. (2022) 'D-KAP: A deep learning-based Kashmiri apple plant disease prediction framework', *Proceedings of the 7th International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 576–581.

[8] Singh, A. *et al.* (2020) 'Apple leaf disease detection using machine learning', *Journal of Plant Diseases and Protection*, 127, pp. 701–712.

[9] Khan, M.A. *et al.* (2023) 'Image preprocessing techniques for plant disease classification', *Sensors*, 23(10), Article 4769.

[10] Li, Y. *et al.* (2023) 'Plant disease detection and classification by deep learning: A review', *Artificial Intelligence in Agriculture*, 7, pp. 19–34.