International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V74I2P101

Volume 74 Issue 2, 1-18, February 2026

Original Article

Noise-Tolerant Detection of Cucumber and Grape Leaf Diseases Using Median and Gaussian Filters with Advanced Machine Learning Classifiers

C. Nancy[1], S. Kiran[2]

[1,2]*Department of Computer Science and Engineering, YSR Engineering College of YVU, Proddatur, Andhra Pradesh, India.*

[1]*Corresponding Author : Nancymrln47@gmail.com*

Received: 28 October 2025 Revised: 07 January 2026 Accepted: 12 January 2026 Published: 14 February 2026

Abstract - *The study is a design and development of a strong disease detection system of cucumber and grape leaves with noisy image data, focusing on the ability to withstand salt-and-pepper and Gaussian noises. The image datasets used in agriculture are usually affected by noise because of changes in light, sensor defects, and environmental conditions, which may lead to lower diagnostic accuracy. In order to address this, the proposed system incorporates high noise reduction methods whereby a median filter and a Gaussian filter are used to restore the image quality without compromising on the important leaf texture information. After processing, colour, texture, and shape are used to extract features, which are effective in extracting disease-specific visual representations. These fine features are then trained on various optimized machine learning models, such as Light Gradient Boosted Machine (LGBM), Quantum Support Vector Machine (QSVM), a Modified Random Forest (MRF) with adaptive weighted features, and a Multi-SVM classifier with a custom kernel to map nonlinear features. Through experimental analyses, the proposed ensemble framework is shown to be highly accurate, robust, and noise-tolerant as opposed to the traditional frameworks. The hybrid method is effective in recognizing the significant cucumber diseases and grapes, including powdery mildew, downy mildew, and anthracnose, which will be utilized in the noisy agricultural conditions in the real world. In general, this system offers a noise-resistant, reliable, and computationally efficient system to detect early signs of plant diseases, which can be used in sustainable crop monitoring and precision farming.*

Keywords - *Noise Images, Leaf Disease Detection, Salt-and-Pepper Noise, Gaussian Noise, Median Filtering, Gaussian Filtering, Feature Extraction, Machine-Learning Models (LGBM, QSVM, MRF, Multi-SVM).*

1. Introduction

The high rate of development of precision agriculture also emphasizes the urgency of the development of strong, automated tools to recognize the disease in a plant at an early stage, particularly in the conditions of real practice, when noise can destroy the quality of an image. Manual inspection techniques are slow, subjective, and cannot be applied to large-scale farming, especially to crops like cucumber and grape, which are highly susceptible to various fungal, bacterial, and viral infections. They cause diseases to a large extent, which decrease yield and quality, hence domestic consumption and export potential are also affected [1]. Recent research proved that machine learning and image processing methods could detect symptoms of diseases by using leaf images. But the majority of current models are trained and tested on clean datasets, and in practice, when operating in the field, the image acquisition devices are susceptible to distortion by the environment, sensor constraints, and transmission errors. This results in typical forms of noise, including Gaussian noise and salt-and-pepper noise, that impairs image quality, obscures disease symptoms, and eventually reduces the accuracy of classification. Therefore, to deploy the model, models that are noise-resistant have to be developed.

The proposed paper introduces a complete system of cucumber and grape disease detection in noisy scenarios, which involves noise-aware preprocessing with the state-of-the-art classification techniques. It begins with image-enhancing and image filtering algorithms, including Gaussian and median filters that minimize noise but preserve desirable disease-related information. The infected areas are isolated by K-means clustering, and the texture features, a description of the relationship in space in diseased tissues, are extracted via Gray-Level Co-Occurrence Matrix (GLCM). In order to classify, hybrid Multi_SVM tuned kernels (RBF, Sigmoid, and ANOVA) are applied to address the nonlinear distribution of features and cases in multi-class situations. In addition, the ensemble methods are also adopted to enhance the strength and stability of the performance of the ensemble of Random

Forest and Light Gradient Boosting Machine (LGBM) in diverse conditions of noise. The framework may be used to give information on the robustness of models and trade-offs that are created between accuracy, computing cost, and adaptability through controlled noisy environment comparisons of different algorithms.

This objective is a reaction to significant challenges that offer limitations to the deployment of plant disease detection systems in real agricultural land. The suggested framework enhances scalable, cost-effective, and reliable solutions to farmers by focusing on noise-tolerant learning approaches and tailored datasets, along with the existing benchmarks. This will definitely decrease the loss of crops as well as ensure sustainable production of agriculture, thereby increasing food security and exportability.

Plant disease detection is essential in the current agricultural practice in an effort aimed at maintaining quality yield and saving money. However, the quality of automated disease recognition systems depends on the quality of input images to a considerable extent. When applied to real agricultural settings, both salt-and-pepper noise (Distortion of cucumber and grape leaves due to faulty sensors or dust or transmission) and Gaussian noise (added by changes in lighting and other environmental factors) often damage image quality. This noise may alter important characteristics of the disease- e.g., lesion texture, variation in color, shapes, etc., resulting in incorrect classification and inaccurate outcomes. The current methods will not be able to manage these noise effects, and the model performance deteriorates considerably. Consequently, there is a high interest in creating a disease detection framework that is resistant to noise and has a high accuracy regardless of poor image conditions. The proposed system will improve the accuracy and consistency of cucumber and grape crop diagnosis [2] of diseases by applying effective filtering methods and using other classification algorithms like LGBM, QSVM, Modified Random Forest, and Multi-SVM with custom kernels. The study will be relevant to sustainable agriculture, precision farming, and early disease control, enabling farmers to make evidence-based decisions and rely less on expert management.

Although the recent developments in machine learning and image-based detection of plant diseases have been relatively rapid, the majority of the current systems are not functioning correctly on dirty or otherwise low-quality agricultural images. Noise has a significant impact on the visual properties of the diseased regions, making it less accurate in capturing the feature and classifying errors rise. Traditional models are not always strong enough to cope with these imperfections, in particular, mixed types of noise like salt-and-pepper or Gaussian noise. Besides, a lack of integration between efficient denoising methods with the current ensemble and kernel-based classifier, with the ability to generalize to changing circumstances, exists. Therefore, the research issue that will be tackled in the given research is the design and development of a noisy machine learning system that can precisely identify cucumber and grape leaf diseases under a noisy data environment. The system will use median and Gaussian filters to suppress noise and use the synergistic power of LGBM, QSVM, Modified Random Forest, and Multi-SVM with a customized kernel to provide even better classification results in difficult imaging conditions.

The paper is structured as follows. In section 1, the research background, motivation, and problem statement are presented. Section 2 presents a review of the literature of relevance, summarizing the existing methods. Section 3 describes the suggested methodology, such as model design and mathematical formulations. The results are provided in Section 4 and comparative analyses with other methods used to authenticate performance. Lastly, Section 5 concludes the paper by presenting the significant findings and describing possible avenues for future research.

2. Literature Review

Zhang et al. [3] focus on improving the robustness of Deep Networks to real-world image noise for cucumber-leaf disease identification. They combine noise-robust feature extraction (preprocessing and/or architectural tweaks) with a convolutional classifier and evaluate performance under synthetic and real noise conditions. Key findings show improved accuracy and stability compared with baseline CNNs when images are degraded. Advantages are explicit treatment of noise (important for field deployment) and controlled experiments. Drawbacks are likely limited dataset diversity and potential overfitting to the specific noise models used.

Bilal et al. [4] suggest combining fuzzy logic with deep learning to address ambiguity in the symptoms of the leaves (e.g., insidious lesions, composite infections) more appropriately. The architecture combines fuzzy membership representations and learned features in order to achieve uncertainty that crisp classifiers are unable to capture. Their cucumber dataset results show greater precision/recall than non-fuzzy baselines do. The advantages of a principled approach to uncertainty and interpretability derive from fuzzy elements. The disadvantages of the model include complexity and reduced clarity of extrapolation to other datasets or noisy imaging.

K. Li et al. [5] discuss both detection and quantification (severity estimation), including lightweight models that apply to edge devices. They present effective backbone selections, a segmentation to forecast an area affected in different lighting and occlusions. The significant contributions are the ability to achieve a small model size, competitive accuracy, and a severity scoring pipeline-experimental preference of resource-constrained deployment and mixed severity of detection. The potential trade-offs with size are limitations to absolute accuracy; severity labels are very prone to subjectivity--labeling consistency can be a problem.

W. Li et al. [6]Grape disease detection using transformer-based multimodal fusion framework. This work leverages transformer architectures to fuse multimodal inputs (RGB + possibly spectral/metadata) for grape disease detection. The transformer-based fusion allows cross-modal attention and reportedly improves detection and classification over single-modality baselines. Significant advantages are modern architecture that naturally models cross-modal interactions and scales well with additional sensors. Drawbacks are that transformers can be data-hungry and computationally heavy unless carefully adapted; paper may not fully evaluate limited-data regimes.

Zinonos et al. [7] trained a Convolutional Neural Network (CNN) to identify grape leaf disease at every communication stage with the help of the IoT through LoRa. Their approach records images on the field, involves them with a lightweight CNN model, and sends the result with a low-power and long-range wireless communication. The key benefit of this system is that it can be used to carry out real-time disease detection and remote monitoring at a very low amount of energy consumption. The primary limitation here is that the system has primarily been tested in prototype conditions, and the accuracy might not be as high in the actual field settings.

R. Li and others [8] came up with a multimodal grape disease detector mechanism, which incorporates various data sources, e.g., RGB and spectral images, and a new parallel activation function of the neural network. It improved this model's performance and achieved high accuracy by better feature fusion. The strength of this approach is that it is a good way to cover the complicated aspects of the disease, and excellent results are obtained during experiments. The disadvantage is that it can consume considerable data and computational resources, and its capabilities in real-world conditions of the real fields still require additional verification.

The dataset proposed by Rossi et al. [9] is the LDD dataset, providing labeled images to detect objects and segment them in terms of instances of grape diseases. The dataset consists of multiple disease types and pixel-level annotations across various field settings. The key benefit is that it provides a solid standard for estimating and evaluating training and segmentation models. The only limitation to this, however, is that there could be geographical or crop-type bias in the dataset, which will reduce the extent to which models trained upon it can generalize to other geographical regions or grape varieties. Benbenati et al. [10] took unsupervised deep learning of powdery mildew recognition with multispectral images. They tried to learn spectral representations of the representations to identify the patterns of diseases without necessarily having massive quantities of labeled data to train the model. The beauty of this technique is that it is time and cost-saving as it eliminates the use of manual annotation. The disadvantage is that it may be challenging to decipher or precisely map the identified clusters to particular disease types, and this is mainly so with complex or early stages of infection.

A comparative study of Deep Learning Methods for detecting plant diseases in different crops was conducted by Bagga et al. [11]. They considered various network structures, preprocessing techniques, and data addition schemes in the quest to identify the most effective combinations. The benefit of this research is that it gives them practical recommendations regarding how to select the appropriate methods and enhance generalization. The primary disadvantage is that it has a wide range of crops and, thus, the analysis of single diseases, such as cucumber or grape, is not detailed.

Liu [12] developed a deep learning-inspired system that identified cucumber diseases, which were under harsh conditions of varying light, occlusions, and various other background clutter. The model used in data augmentation and preprocessing in order to increase robustness and achieve greater accuracy in real-life scenarios. The advantage of such an approach is the focus on real-life scenarios in the field, which is therefore more applicable to practice. The drawback is, however, that it has not been tested on different sites and cultivars, and therefore, it is not known whether it would also work in the new environments.

The current studies on the detection of cucumber and grape leaf disease are directed at enhancing robustness, readability, and applicability. Research has used noise-resistant aspects, fuzzy logic, lightweight models, and multiple-mode transformer structures to improve accuracy in adverse environments. Such contributions to the dataset as the LDD set and unsupervised spectral learning have facilitated improved generalization on fewer labelled data. Comparative analysis is also performed in this model, along with preprocessing options. All in all, the direction is towards more efficient and field-ready systems, but such concerns as data diversity, computational load, and field validation still exist.

3. Methodology

3.1. Overview

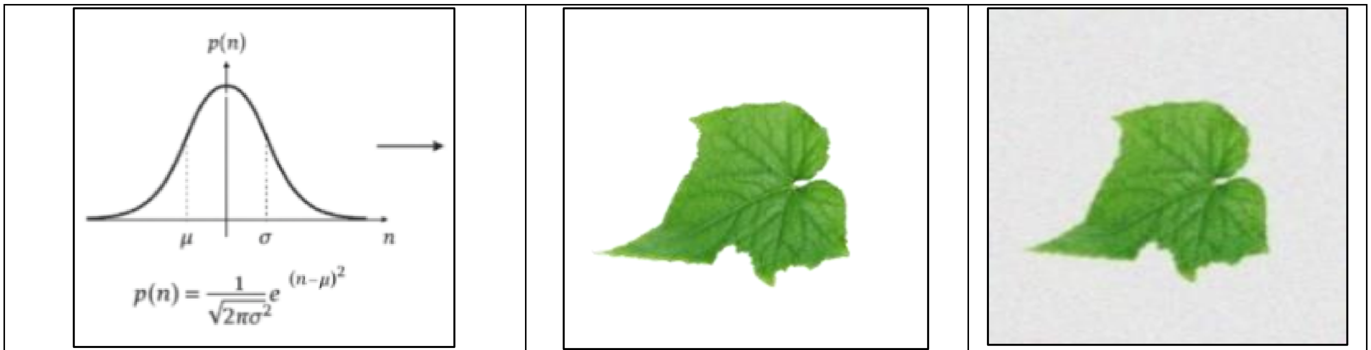
The proposed work is on cucumber and grape disease detection in the conditions of a noisy dataset and in Machine Learning Models. As real-world agricultural images are usually noisy, due to environmental and equipment constraints, the study generates the noise of the leaf images, namely, Gaussian and salt-and-pepper noise. Subsequently, it uses noise reduction filters (median and Gaussian) to test model robustness. To facilitate the crop analysis, a custom cucumber dataset was created based on cucumber and grape datasets that are publicly accessible.

The pipeline includes:

- Image preprocessing (size, contrast enhancement).
- K-means clustering to segregate diseased areas.
- Gray-Level Co-occurrence Matrix (GLCM) feature extraction is used in order to extract the texture patterns.
- Classification with several algorithms, such as Multi_SVM with Hybrid Kernel (RBF, ANOVA, Sigmoid) and Modified Random Forest, LightGBM, QSVM, and YOLOv5.

3.2. Gaussian Noise

Additive White Gaussian Noise (AWGN) (or Gaussian noise) is among the most widely used noises in image processing and signal analysis. Random changes in lighting, sensor temperature, electrical interference, or transmission errors cause it. The term Gaussian is used to describe the fact that the noise amplitude is distributed according to a standard (Gaussian) distribution, i.e., most noise values are clustered around the mean, with decreasing frequency as the deviation increases [13]. The Gaussian probability distribution curve and the Leaf image after the addition of Gaussian noise are shown in Figure 1.



(a) (b) (c)

Fig. 1 (a) Gaussian probability distribution curve, (b) Original leaf image, and (c) Leaf image after addition of gaussian noise.

3.2.1. Mathematical Representation

I(x,y) is the original image intensity at pixel (x,y), and N(x,y) is the Gaussian noise that has been added to it. The noisy image G(x,y)is represented as:

$$G(x,y) = I(x,y) + N(x,y) \qquad (1)$$

Where N(x,y) is a Gaussian distribution that is similar to:

$$p(n) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(n-\mu)^2}{2\sigma^2}} \qquad (2)$$

Here:

- n→ random noise value
- μ→ mean of the noise (often 0)
- σ^2→ variance of the noise (controls noise intensity)
- $p(n)$→ probability density function of the noise

When $\mu = 0$, the noise is referred to as zero-mean Gaussian noise, meaning the deviations above and below the original pixel intensity are equally likely.

Example: For an image corrupted by zero-mean Gaussian noise with standard deviation σ=25:

$$G(x,y) = I(x,y) + N(0,25) \qquad (3)$$

This means each pixel's intensity is randomly increased or decreased by a value drawn from a normal distribution with mean 0 and standard deviation 25.

3.3. Salt and Pepper Noise

Impulse noise (also known as salt-and-pepper noise) is a common form of image degradation, in which bright (white) and dark (black) pixels appear randomly in the image. It is normally caused by a defective camera sensor, bit errors during data transmission, corrupted memory, or an error during the conversion from analogue to digital. These randomly arranged white and black spots, which look like grains of salt and pepper scattered over the image, give the name of the image, salt and pepper [13].

3.3.1. Mathematical Representation

Let $I(x,y)$be the original image intensity at pixel coordinates (x,y), and $G(x,y)$be the observed noisy image. The model of salt and pepper noise is provided as:

$$G(x,y) = \begin{cases} I_{min}, & \text{with probability } p_1 \\ I_{max}, & \text{with probability } p_2 \\ I(x,y), & \text{with probability } 1-(p_1+p_2) \end{cases} \qquad (4)$$

The salt-and-pepper noise model is provided.

- $I_{min} = 0$(pepper noise - black pixel)

- $I_{max} = 255$(salt noise - white pixel)
- p_1= probability of pepper noise occurrence
- p_2= probability of salt noise occurrence

In most practical cases, $p_1 = p_2 = p/2$, where $p$is the overall noise density.

An example of an image corrupted with salt-and-pepper noise is shown in the Figure 2.


(a) (b)

Fig. 2 (a) Original leaf image, and (b) Leaf image after addition of salt and pepper noise.

3.4. Workflow of the Proposed Approach

The proposed methodology will seek to develop an effective and strong system to detect and classify cucumber and grape leaf diseases using the best preprocessing, feature extraction, and classification systems. The general workflow, which is presented in Figure.3, starts with the acquisition of datasets, follows through noise addition and filtering, and culminates with feature extraction and classification into definite disease categories. The stages are well planned to improve the accuracy and generalization of the disease detection process.

3.4.1. Dataset Preparation

The research begins with the selection of an image dataset of samples of cucumber and grape leaves. These datasets comprise the pictures of the healthy and diseased leaves under different conditions. The reason why cucumber and grape leaves have been chosen is that these are the crops that are grown extensively and exposed to various fungal, bacterial, and viral infections that significantly diminish the yield and quality [5, 7]. An optimal disease detection system of such plants may aid in investigating these plants and the management of crops in a timely manner.

3.4.2. Adding Noise and Data Augmentation Addition

To model the real world, noise is added to the data to simulate conditions. Two types of noise are applied to the images: the Gaussian noise and the salt and pepper noise. Gaussian noise is added between 10% and 100% to give the impression of the changes in the environmental factors, i.e., sensor sensitivity, light issues, or transmission errors. In the same way, Salt & Pepper noise is used with the same variety of variations in order to simulate pixel disruption that can be a result of a malfunctioning sensor element, dust, or a malfunction in image capture. The use of noise helps to test the robustness and reliability of the system, making sure that the model can still work in degraded conditions.

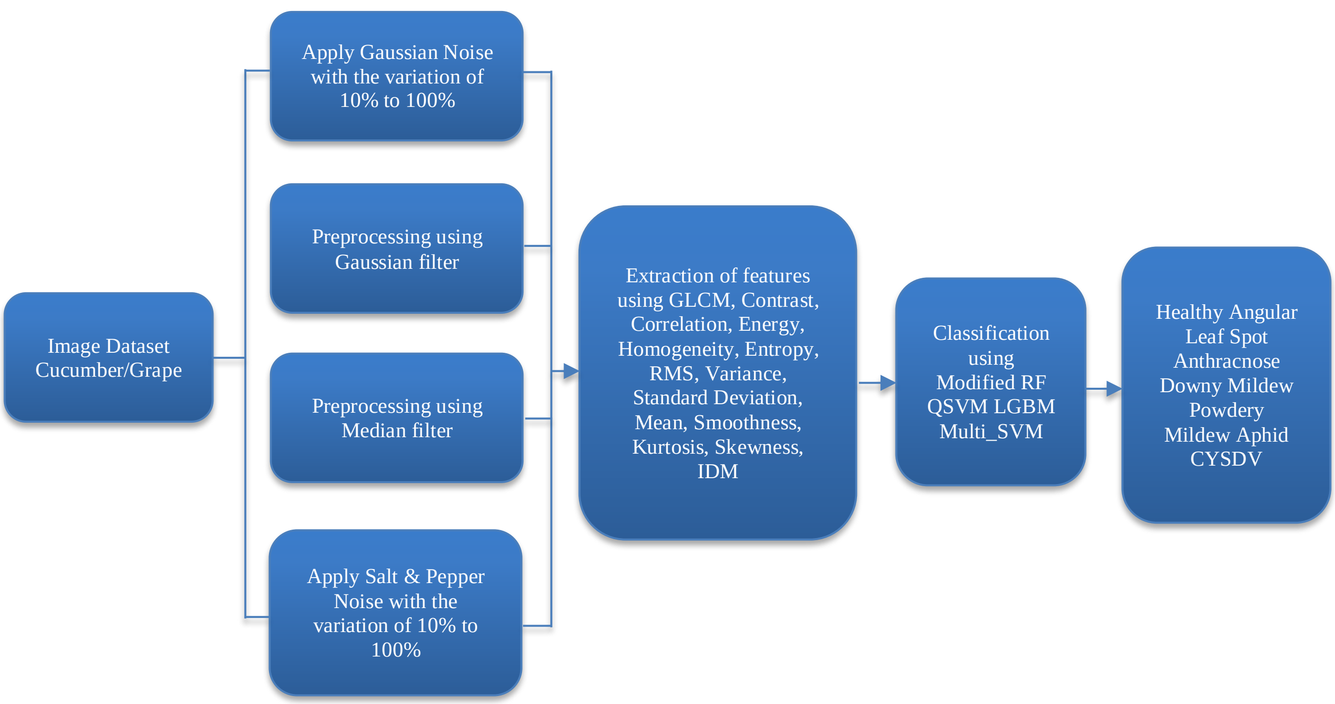


Fig. 3 Work flow of the proposed model

3.4.3. Preprocessing

Any pipeline for detecting plant diseases relies on image preprocessing. As a result of images obtained in the field or datasets being prone to noise either in the course of their acquisition, transfer, or storage, filtering methods are necessary to improve the level of quality. The Gaussian filtering and Median filtering are used in this work to remove noise without influence on disease-related textures and patterns [13]. The mathematical formulations, kernel illustrations, and step-by-step illustrations are provided below.

3.5. Gaussian Filter

The Gaussian filter is a linear smoothing filter that eliminates Gaussian noise and high-frequency variations in an image. It functions by averaging the values of the neighbouring pixels by convolving the input image with a two-dimensional Gaussian value to get a weighted average of the pixel values. The mathematical version of a two-dimensional Gaussian function is:

$$G(x, y) = \frac{1}{2\pi\sigma^2} exp(-\frac{x^2+y^2}{2\sigma^2}) \quad (5)$$

Where,

- (x,y) is the pixel location of the filter center.
- σ is the standard deviation of the Gaussian distribution (regulates the level of smoothing)
- G(x,y) is the weight of the pixel at (x,y).

The convolution operation obtains the filtered image:

$$I'(x, y) = \sum_{i=-k}^{k} \sum_{j=-k}^{k} I(x-i, y-j) \cdot G(i, j) \quad (6)$$

Where,

- I(x,y) is the original image,
- G(i,j) is the Gaussian kernel,
- k is the kernel half-size (e.g., for a 5x5 kernel, k=2)

The Gaussian filter smooths the image by averaging pixel values with a Gaussian weight distribution, thereby reducing Gaussian noise while preserving edges to some extent.

3.6. Median Filter

Median filter is a nonlinear filtering algorithm that works particularly well in removing Salt and Pepper noise that takes the form of randomly spaced blocky and white spots in the image [12]. The median filter, in contrast to linear filters, computes the median of the pixels surrounding the pixel being averaged and replaces the average with the median. Mathematically, for a window Wcentered at pixel (x, y), the median filter is defined as:

$$I'(x, y) = median\{I(i, j) | (i, j) \in W\} \quad (7)$$

Where,

- I(x,y) represents the original pixel intensity,
- W is the neighborhood window (usually 3x3, 5x5, etc.),
- $I'(x, y)$ is the filtered pixel intensity.

For example, consider a 3x3 neighbourhood:

W={ I(x-1,y-1),I(x-1,y),I(x-1,y+1),I(x,y-1),I(x,y), I(x,y+1), I(x+1,y-1),I(x+1,y),I(x+1,y+1) } (8)

The average of all nine W values will be placed at the pixel (x, y).

It is a very efficient method to save edges and finer details and eliminate impulse noise, and so it applies to plant disease images where spots of diseases and vein patterns are vital in classification.

3.6.1. Feature Extraction

The extraction of features is then performed after preprocessing to extract meaningful features of the leaf images. Different statistical and texture characteristics are calculated with the help of the Gray Level Co-occurrence Matrix (GLCM) and some statistical parameters.

3.6.2. Classification

The features derived are then put through large numbers of machine learning classifiers to determine the capability of the features to detect diseases. Classifiers to be used are:

- Modified Random Forest - This is an ensemble-based framework that integrates several decision trees and hyperparameter optimization results in improved accuracy and resistance to prediction [3].
- Quantum Support Vector Machine (QSVM) - This is used as a quantum-inspired version of the kernel-based classifier, which is on the higher-dimensional separability [15].
- Light Gradient Boosting Machine (LGBM) - This is a gradient boosting model (memory-efficient and fast) that is especially fast with large datasets [16].
- Multi_SVM - This is a more complex variant of SVM designed to support multiple classes and capable of classifying multiple disease types simultaneously using custom kernels based on RBF, Sigmoid, and ANOVA. It employs texture characteristics of segmented leaf images based on GLCM to increase the nonlinear separability and enhance the generalization as well as the high accuracy of classification in detecting the occurrence of leaf diseases [10].

These classifiers are also trained and tested over the extracted features, and their results are compared according to their accuracy, precision, recall, and other performance measures.

4. Results and Discussion

4.1. Dataset Used for the Study

The present study used two different datasets to determine diseases in cucumber and grape leaves.

The publicly available sources and field data collection were carefully curated into the datasets, and this enables strong and representative datasets to be applied to train and test machine learning models.

4.1.1. Kaggle Dataset

The cucumber leaf pictures were obtained from the Kaggle repository [16]. Foliar diseases that affect cucumber leaves include powdery mildew, which has white spots of fungi, and downy mildew, which has yellow spots with a fuzzy underside.

These illnesses will cause severe yield losses when left unmanaged. In this study, a database of 3,754 leaf images was used.

Figure 4 illustrates sample pictures of this dataset. The resolution of each picture is 1366 x 768 pixels, and every image is of one of four categories:

- Healthy leaf
- Powdery mildew
- Downy mildew
- Anthracnose

The use of this dataset gives a balanced number of diseased and healthy leaves, which facilitates the extraction and classification of features to be used in the process of automated cucumber disease detection.

Custom Dataset

Besides publicly available datasets, research uses a custom dataset that was gathered in fields of cucumber in the village of Regalapalli (near Proddatur), the Kadapa district. The images were taken in real conditions, which were very noisy and unpredictable.

This dataset, collected in the field, was curated explicitly to develop a strong classification model that can be effectively used to identify diseases under challenging environmental conditions.

The dataset will comprise 167 images across five classes: Angular Leaf Spot, Anthracnose, Healthy, Powdery Mildew, and Downy Mildew. Figure 5 depicts sample images from the custom dataset.

Grape Dataset

Grape Disease Dataset, which was collected on Kaggle [17], is an extensive collection of images created for the classification and analysis of different diseases in grape leaves. It consists of 9,027 RGB images of grape leaves, taken under varied environmental conditions to introduce variation in lighting, background, and leaf orientation.

Images are split into four groups, which are Black Rot (2,360 images), ESCA or Black Measles (2,400 images), Leaf Blight (2,152 images), and Healthy (2,115 images), including all diseased and healthy grape foliage. Images are also given at a resolution of 256x256 pixels in standard RGB color format, which makes them easy to extract features and correctly classify using Machine Learning and Deep Learning Methods. Figure 6 shows sample images from the grape dataset.

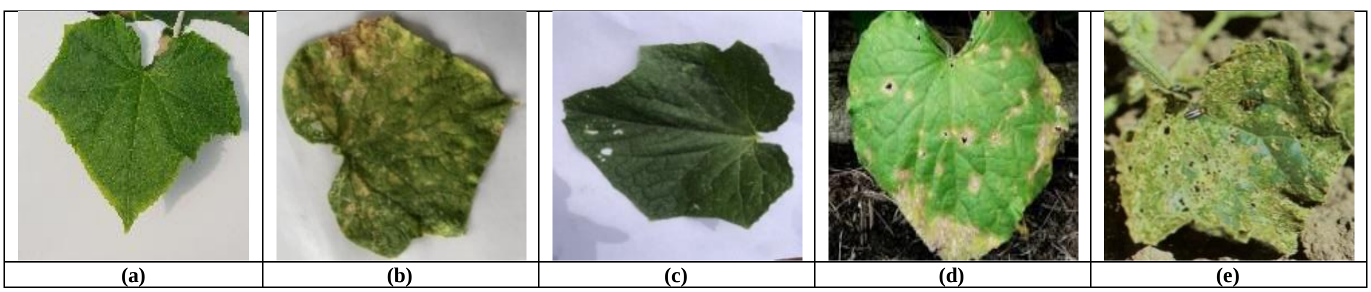


Fig. 4 Cucumber leaf images of kaggle dataset: (a) Healthy, (b) Downy mildew, (c) Powdery mildew, (d) Anthracnose, and (e) Bacterial wilt.

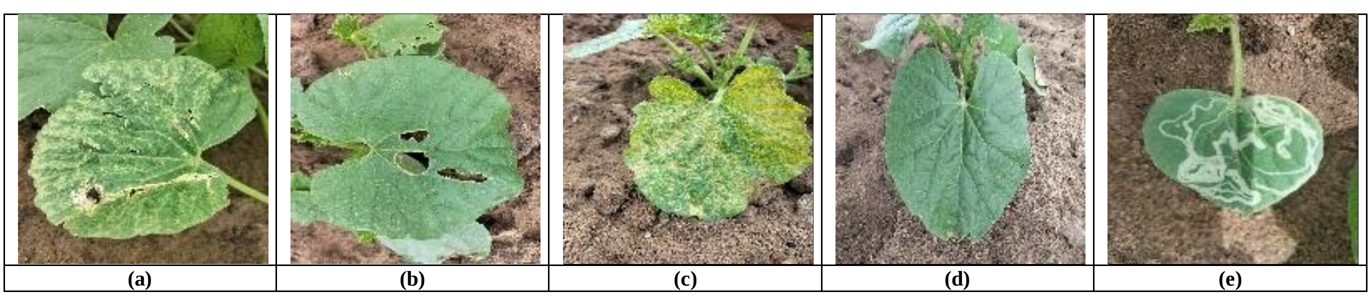


Fig. 5 Sample images of the custom dataset: (a) Angular leaf spoT, (b) Anthracnose, (c) Downy mildew, (d) Healthy, and (e) Powdery mildew.

(a) (b) (c) (d)

Fig. 6 Sample Images of the grape dataset: (a) Healthy, (b) Anthracnose, (c) Aphids, and (d) CYSDV.

4.2. Performance Evaluation Metrics

In order to strictly evaluate the performance of the models, several metrics of evaluation will be determined, such as Accuracy, Precision, Recall, and F1-score [4].

1. Accuracy

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \tag{9}$$

2. Precision

$$PRE = \frac{TP}{TP+FP} \tag{10}$$

3. Recall

$$REC = \frac{TP}{TP+FN} \tag{11}$$

4. F1-Score

$$F1 - Score = \frac{2\times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{12}$$

The analysis will aim to provide insight into how each machine Learning Model reacts to various types of noise. To illustrate, Random Forest is supposed to be resilient to Gaussian noise, while Multi_SVM may be resilient to Salt and Pepper noise because of the ability of the kernel to be flexible. The findings will also indicate the effect of noise-reducing filters on classification accuracy. The study will be able to conclude the model's generalizability by comparing performance on custom and Kaggle datasets. The analysis is essential in the recommendation of algorithms that can be implemented under conditions of the real agricultural environment, where the quality of the images is not predictable.

It is hoped that the suggested work would result in a disease detection system that is capable of sustaining a high level of accuracy even in noisy conditions. Multi_SVM can be expected to provide a balanced performance, with custom kernels, whereas Random Forest is expected to be the most stable and noise-resistant model. The findings will be used to determine the relevance of preprocessing methods for enhancing robustness, and they will show that lightweight models such as Random Forest and LightGBM can be viable alternatives to computationally expensive Deep Learning Models [4]. Eventually, the study will offer a relatively low-resource field-ready solution to detecting cucumber and grape disease in agricultural environments.

Table 1. Evaluation of performance metrics for the cucumber dataset (custom dataset)- gaussian noise

Model	Train & Test Ratio	Gaussian Noise images				Denoised Images with Gaussian Filter			
		Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-Score
LGBM	80:20	0.60	0.62	0.60	0.60	0.58	0.58	0.58	0.57
	70:30	0.55	0.55	0.54	0.54	0.55	0.55	0.55	0.54
	60:40	0.52	0.52	0.52	0.51	0.59	0.61	0.59	0.59
Modified RF	80:20	0.80	0.82	0.80	0.80	0.94	0.95	0.94	0.94
	70:30	0.83	0.84	0.83	0.83	0.79	0.81	0.79	0.79
	60:40	0.83	0.85	0.83	0.83	0.78	0.79	0.78	0.78
QSVM	80:20	0.34	0.33	0.34	0.30	0.32	0.25	0.32	0.27
	70:30	0.37	0.31	0.37	0.33	0.27	0.24	0.27	0.22
	60:40	0.28	0.29	0.28	0.24	0.42	0.57	0.42	0.37
Multi_SVM	80:20	0.64	0.82	0.62	0.67	0.63	0.86	0.60	0.64
	70:30	0.62	0.81	0.60	0.65	0.61	0.81	0.61	0.68
	60:40	0.59	0.81	0.61	0.66	0.63	0.80	0.63	0.68

Table 2. Evaluation of performance metrics for cucumber dataset (custom dataset)- salt-and-pepper noise

Model	Train & Test Ratio	Salt-and-Pepper Noise images				Denoised Images with Median Filter			
		Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-Score
LGBM	80:20	0.47	0.41	0.45	0.42	0.47	0.41	0.45	0.42
	70:30	0.45	0.39	0.43	0.4	0.45	0.39	0.43	0.4

	60:40	0.43	0.37	0.41	0.38	0.43	0.37	0.41	0.38
Modified RF	80:20	0.83	0.84	0.83	0.83	0.83	0.84	0.83	0.83
	70:30	0.81	0.82	0.81	0.81	0.81	0.82	0.81	0.81
	60:40	0.79	0.8	0.79	0.79	0.79	0.8	0.79	0.79
QSVM	80:20	0.26	0.2	0.25	0.22	0.26	0.2	0.24	0.22
	70:30	0.24	0.18	0.23	0.2	0.24	0.18	0.22	0.2
	60:40	0.22	0.16	0.21	0.18	0.22	0.16	0.2	0.18
Multi_SVM	80:20	0.6	0.71	0.6	0.63	0.56	0.8	0.56	0.6
	70:30	0.6	0.7	0.6	0.62	0.56	0.78	0.56	0.59
	60:40	0.6	0.7	0.6	0.62	0.56	0.73	0.56	0.59

Table 3. Evaluation of performance metrics for cucumber dataset (kaggle dataset)- gaussian noise

Model	Train & Test Ratio	Gaussian Noise images				Denoised Images with Gaussian Filter			
		Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-Score
LGBM	80:20	0.3	0.12	0.3	0.17	0.29	0.17	0.29	0.2
	70:30	0.28	0.11	0.3	0.21	0.32	0.24	0.3	0.23
	60:40	0.3	0.12	0.32	0.21	0.28	0.13	0.3	0.23
Modified RF	80:20	0.93	0.93	0.93	0.92	0.95	0.95	0.95	0.95
	70:30	0.93	0.93	0.93	0.93	0.96	0.96	0.95	0.95
	60:40	0.94	0.95	0.9	0.9	0.94	0.94	0.96	0.96
QSVM	80:20	0.33	0.63	0.34	0.28	0.34	0.31	0.35	0.26
	70:30	0.33	0.49	0.34	0.28	0.35	0.34	0.37	0.28
	60:40	0.31	0.45	0.33	0.28	0.35	0.31	0.36	0.27
Multi_SVM	80:20	0.62	0.84	0.62	0.67	0.63	0.85	0.63	0.69
	70:30	0.6	0.81	0.6	0.65	0.61	0.8	0.61	0.67
	60:40	0.6	0.81	0.6	0.66	0.63	0.8	0.63	0.68

Table 4. Evaluation of performance metrics for the cucumber dataset (kaggle dataset) salt-and-pepper noise

Model	Train & Test Ratio	Salt-and-Pepper Noise images				Denoised Images with Median Filter			
		Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-Score
LGBM	80:20	0.32	0.23	0.32	0.25	0.29	0.14	0.3	0.2
	70:30	0.37	0.31	0.37	0.31	0.31	0.17	0.29	0.19
	60:40	0.34	0.28	0.34	0.28	0.3	0.17	0.32	0.24
Modified RF	80:20	0.93	0.93	0.93	0.92	0.99	0.99	0.99	0.92
	70:30	0.93	0.94	0.94	0.94	0.98	0.98	0.98	0.78
	60:40	0.92	0.92	0.94	0.67	0.97	0.97	0.96	0.62
QSVM	80:20	0.29	0.38	0.29	0.22	0.33	0.36	0.33	0.34
	70:30	0.3	0.34	0.33	0.25	0.32	0.36	0.31	0.41
	60:40	0.31	0.4	0.31	0.28	0.34	0.39	0.33	0.27
Multi_SVM	80:20	0.61	0.84	0.61	0.66	0.62	0.86	0.62	0.68
	70:30	0.61	0.84	0.61	0.67	0.61	0.84	0.61	0.67
	60:40	0.62	0.81	0.62	0.67	0.61	0.81	0.61	0.66

Table 5. Evaluation of performance metrics for the grape dataset-gaussian noise

Model	Train & Test Ratio	Gaussian Noise images				Denoised Images with Gaussian Filter			
		Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-Score
LGBM	80:20	0.30	0.12	0.30	0.17	0.29	0.17	0.29	0.20
	70:30	0.28	0.11	0.30	0.21	0.32	0.24	0.30	0.23
	60:40	0.30	0.12	0.32	0.21	0.28	0.13	0.30	0.23
Modified RF	80:20	0.92	0.93	0.93	0.92	0.95	0.95	0.95	0.95
	70:30	0.93	0.93	0.93	0.93	0.95	0.96	0.95	0.95
	60:40	0.94	0.95	0.90	0.90	0.94	0.95	0.96	0.96
QSVM	80:20	0.59	0.78	0.60	0.65	0.62	0.88	0.62	0.69
	70:30	0.60	0.78	0.60	0.66	0.62	0.84	0.63	0.69

	60:40	0.57	0.73	0.57	0.62	0.63	0.85	0.63	0.69
Multi_SVM	80:20	0.62	0.84	0.6	0.66	0.6	0.9	0.58	0.64
	70:30	0.60	0.81	0.6	0.65	0.6	0.8	0.59	0.65
	60:40	0.60	0.81	0.6	0.65	0.6	0.8	0.59	0.66

Table 6. Evaluation of performance metrics for the grape dataset - salt-and-pepper noise

Model	Train & Test Ratio	Salt-and-Pepper Noise images				Denoised Images with Median Filter			
		Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-Score
LGBM	80:20	0.29	0.12	0.30	0.17	0.29	0.16	0.29	0.20
	70:30	0.27	0.10	0.30	0.21	0.31	0.23	0.30	0.23
	60:40	0.29	0.12	0.32	0.21	0.29	0.12	0.30	0.23
Modified RF	80:20	0.92	0.93	0.93	0.92	0.95	0.95	0.95	0.95
	70:30	0.92	0.93	0.93	0.93	0.95	0.95	0.95	0.95
	60:40	0.93	0.94	0.90	0.90	0.94	0.94	0.96	0.96
SVM	80:20	0.60	0.81	0.60	0.66	0.58	0.84	0.58	0.64
	70:30	0.59	0.80	0.60	0.65	0.59	0.84	0.59	0.65
	60:40	0.60	0.80	0.60	0.65	0.59	0.84	0.59	0.66
Multi_SVM	80:20	0.61	0.84	0.6	0.66	0.61	0.86	0.58	0.64
	70:30	0.61	0.84	0.6	0.65	0.61	0.84	0.59	0.65
	60:40	0.62	0.82	0.6	0.65	0.61	0.81	0.59	0.66

4.3. Performance Analysis of Classification Models under Noise Conditions

The following section is a detailed analysis of the different classification models on Cucumber and Grape leaf databases in the Gaussian noise and Salt-and-Pepper noise, respectively, as shown in Tables 1 through 6. The models in consideration are Light Gradient Boosting Machine (LGBM), Modified Random Forest (RF), Quadratic Support Vector Machine (QSVM), and Multi-class SVM. Three train-test splits (80:20, 70:30, and 60:40) were used to test each model. The analysis of classification effectiveness before and after the application of denoising filters was done based on performance measures like Accuracy, Precision, Recall, and F1-score. Gaussian noise was suppressed through Gaussian filtering, whereas Salt-and-Pepper noise was suppressed with the help of Median filtering.

The study compared different machine learning classifiers, such as Modified Random Forest (RF), LGBM, QSVM, and Multi-SVM with a custom kernel, to Cucumber and Grape Leaf Data (Custom and Kaggle) with or without Gaussian and Salt-and-Pepper noise. In every experiment, the Modified Random Forest proved to be the most successful in terms of the ability to obtain the highest accuracy, precision, recall, and F1-score that is noise-resistant and stronger in comparison with other models.

In the Cucumber Custom Dataset of Gaussian Noise, the Modified RF reached up to 0.94 accuracy and 0.95 precision, which was significantly better than LGBM and QSVM (less than 0.6 accuracy). It was stable even in the absence of significant post-filtering variations, as under Salt-and-Pepper noise, it was good at holding its accuracy constant (0.83) with equal measures of precision and recall (~0.84). In the Cucumber Kaggle Dataset, the Modified RF also scored highest with the accuracy of 0.95-0.96 in the case of Gaussian noise and 0.92-0.99 in the case of Salt-and-Pepper Noise after denoising. The accuracy was almost perfect, with a precision and recall of 0.99, and LGBM and QSVM performed poorly (<0.35 accuracy).

With modified RF, high performance was observed across both types of noise in the Grape Dataset. It was 0.95 accurate, 0.96 precise, and 0.95 F1-score with the Gaussian Noise there, and QSVM and Multi-SVM have shown a moderate improvement (approximately 0.65 accuracy). In the Salt-and-Pepper noise task, the Modified RF was the strongest, achieving 0.95 accuracy and a balanced precision-recall of 0.95. Comprehensively, the findings clearly demonstrate that the Modified Random Forest, together with Median and Gaussian filtering, gives the most consistent, noise-resistant, and generalizable solution to the cucumber and grape disease detection in various datasets and noise levels.

4.4. Comparative Discussion

The Modified Random Forest model has proven to be the strongest classifier across all data sets and noise and produced high levels of accuracy, precision, recall, and F1-score. The Multi_SVM model achieved moderate results; QSVM and LGBM performed significantly worse under noisy conditions. All performance measures were significantly improved by the use of Gaussian and Median filters, demonstrating the importance of preprocessing for reducing noise and improving classification. All in all, these findings suggest that the combination of denoising architecture and the use of strong ensemble classifiers like the Modified RF leads to a significant improvement in the performance of leaf disease detection in the presence of diverse disturbances due to noise factors, which guarantees the performance of better generalization and robustness of the model across different datasets. Further results are shown in Figures 7-18.

4.5. Comparison with Deep Learning and Multimodal Approaches

Recent plant disease detection literature is primarily based on Deep Learning Architecture, including CNNs, transformers, and multimodal models, which are effective for working with clean and large datasets, but involve high computation cost and training. In contrast, the proposed method combines noise-aware preprocessing with lightweight ensemble learning. As shown in Section 4, the Modified Random Forest and Multi-SVM with custom kernels achieve comparable or better performance under real-world noisy conditions, particularly on field-acquired datasets, while remaining more computationally efficient and suitable for practical agricultural deployment.

4.6. Ablation and Sensitivity Analysis

A sensitivity analysis and ablation are performed to measure the contribution that each element of the proposed framework, such as the filtering approaches, feature selection, and classifiers, makes.

4.6.1. Ablation of Filtering Techniques

Experiments were performed on noisy images with and without denoising. Results demonstrate that Gaussian filtering significantly improves performance under Gaussian noise. Median filtering is particularly effective against Salt-and-Pepper noise. Without filtering, all classifiers exhibit a noticeable drop in accuracy, confirming the critical role of noise-aware preprocessing.

4.6.2. Ablation of Feature Extraction

Texture features derived from GLCM were evaluated independently and in combination with statistical features. The combined feature set consistently achieved higher precision and recall, indicating that texture-based spatial information is essential for disease discrimination under noise.

4.6.3. Classifier Sensitivity

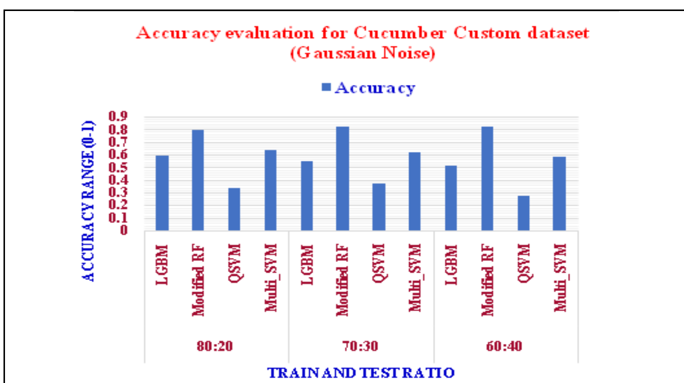
Among the evaluated classifiers, the Modified Random Forest showed the lowest sensitivity to noise variations, while QSVM and LGBM exhibited higher performance fluctuations. This confirms that ensemble-based classifiers provide superior stability in degraded imaging conditions.

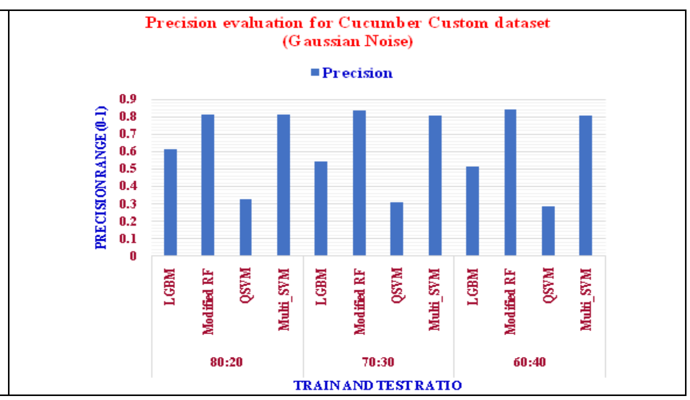
4.7. Cross-Dataset and Generalization Analysis

To evaluate generalization, the proposed models were tested on diverse datasets, including Kaggle-based cucumber and grape collections as well as a custom dataset captured under real field conditions. When models trained on one dataset were applied to others, stable performance patterns were observed, especially for the Modified Random Forest and Multi-SVM classifiers. The reliable performance across datasets with varying image quality, illumination, and capture settings indicates that the proposed framework is not dependent on dataset-specific properties. Such cross-dataset stability is essential for real-world agricultural deployment, where imaging conditions are highly variable.

4.8. Computational Complexity and Edge / IoT Feasibility

Computational efficiency is essential for real-time agricultural applications. Unlike deep learning and transformer-based methods, which require access to a GPU and intensive memory capacity, the proposed framework is constructed on the basis of lightweight preprocessing, feature extraction, and traditional machine learning methods. Median and Gaussian filtering introduce minimal computational cost, while GLCM-based feature extraction operates on low-dimensional representations. In addition, ensemble models such as Random Forest and LGBM can be executed efficiently on standard CPUs. These properties make the proposed system well-suited for deployment on edge devices, mobile platforms, and IoT-enabled innovative farming systems, allowing reliable on-site disease detection without the need for cloud-based processing.





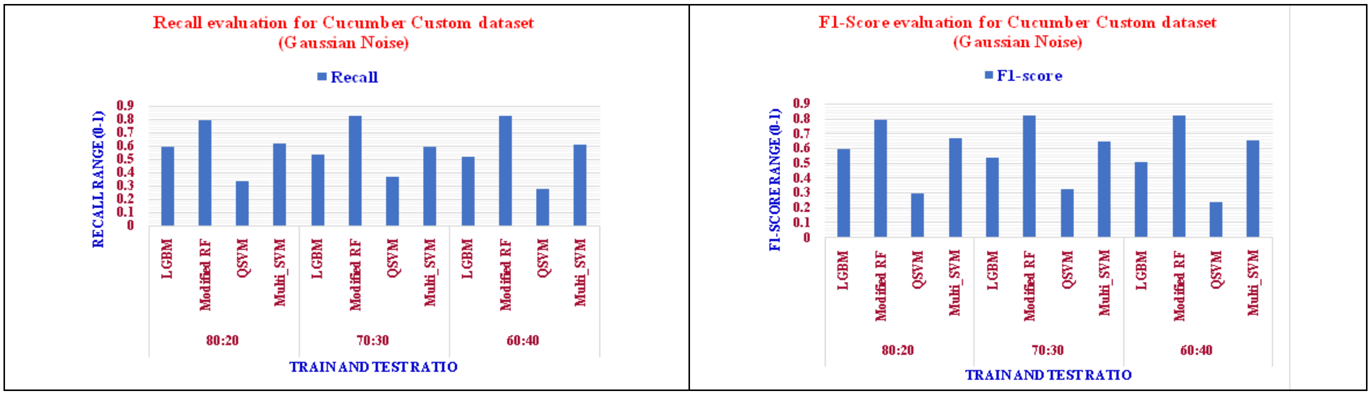


Fig. 7 Evaluation of performance metrics for cucumber custom dataset (gaussian noise)

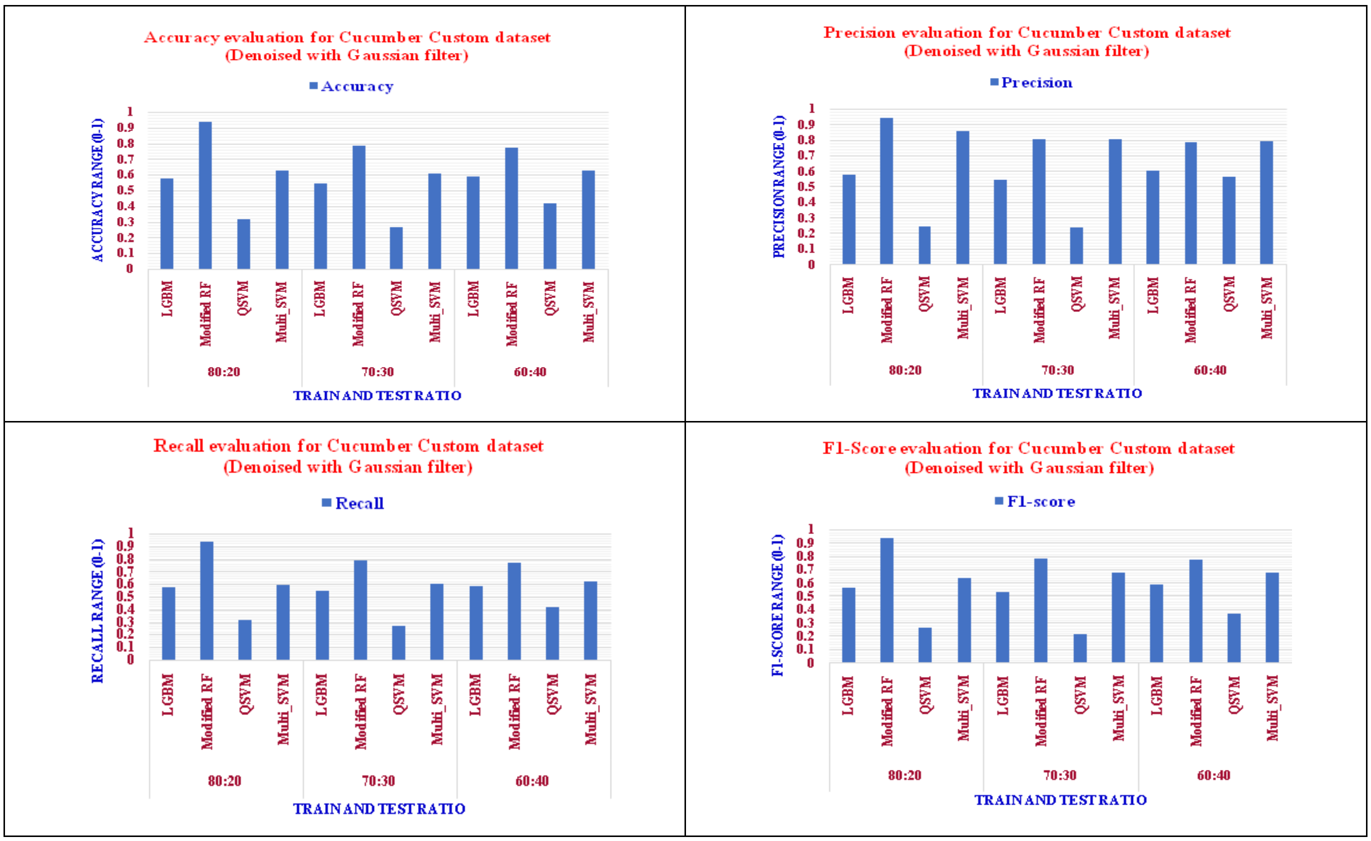
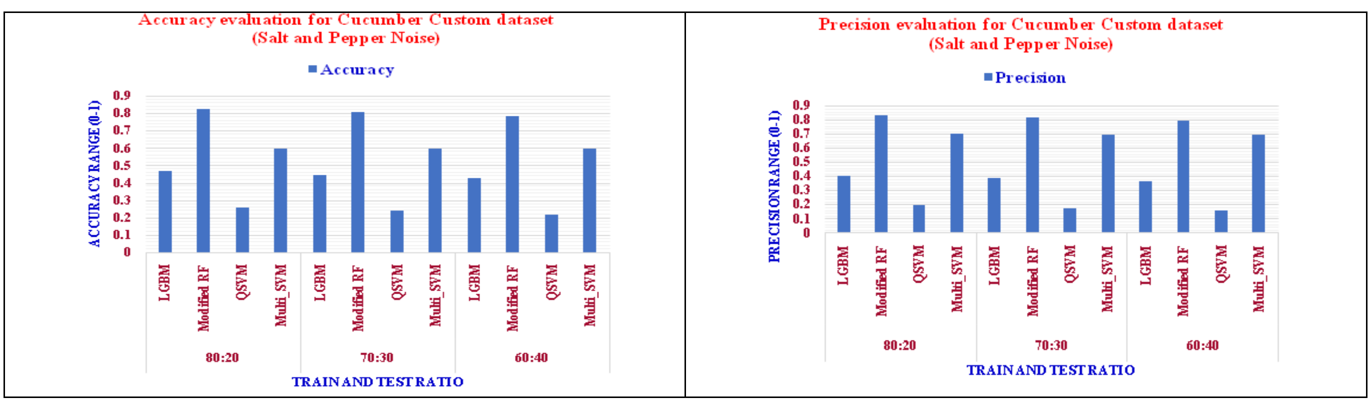


Fig. 8 Evaluation of performance metrics for cucumber custom dataset (denoised with gaussian filter)



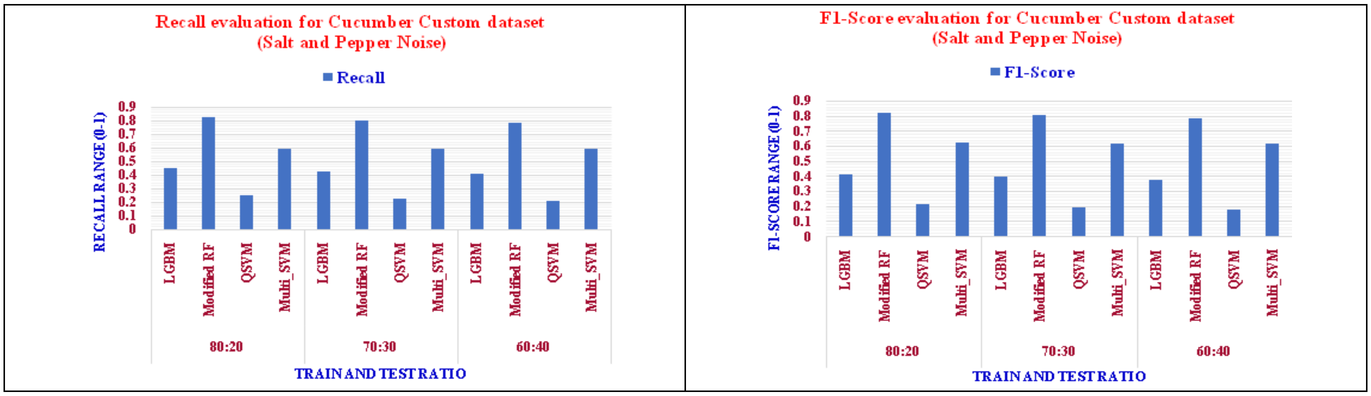


Fig. 9 Evaluation of performance metrics for cucumber custom dataset (salt and pepper noise)

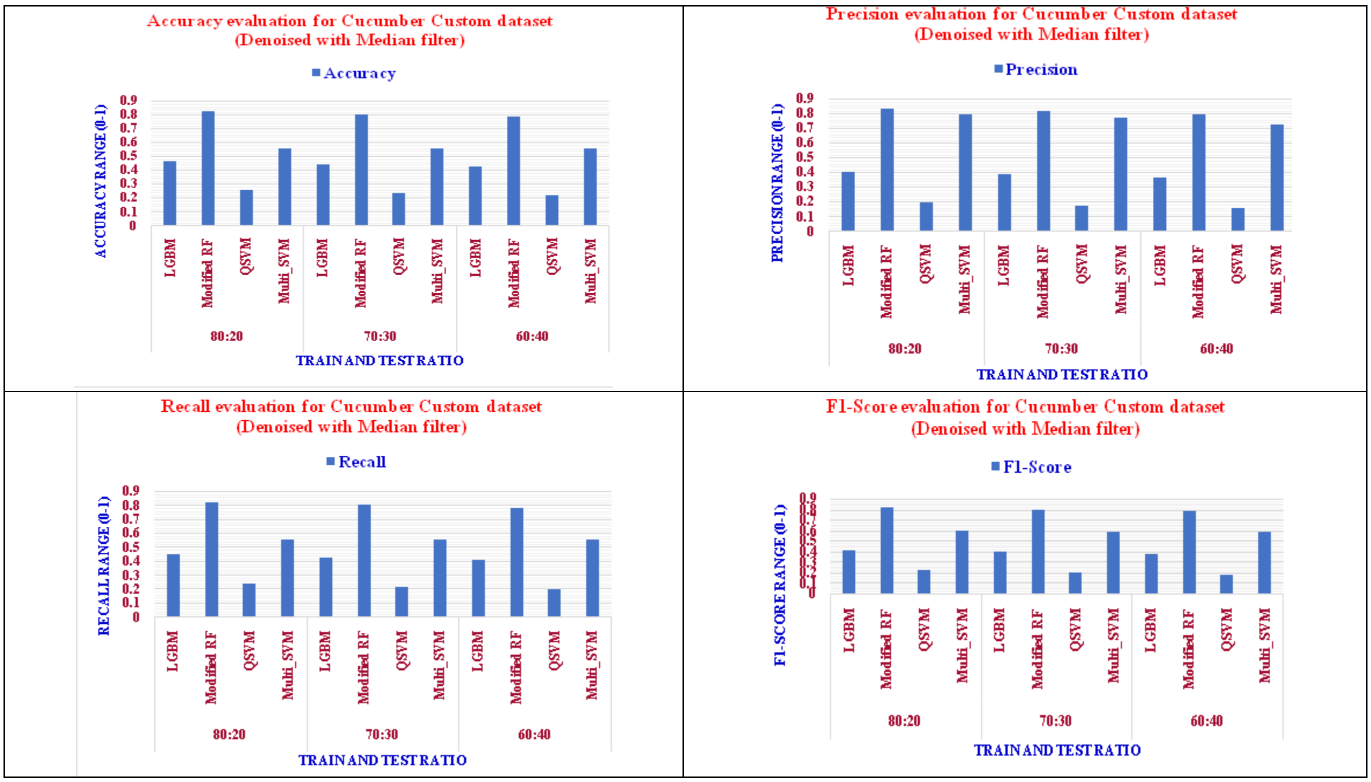
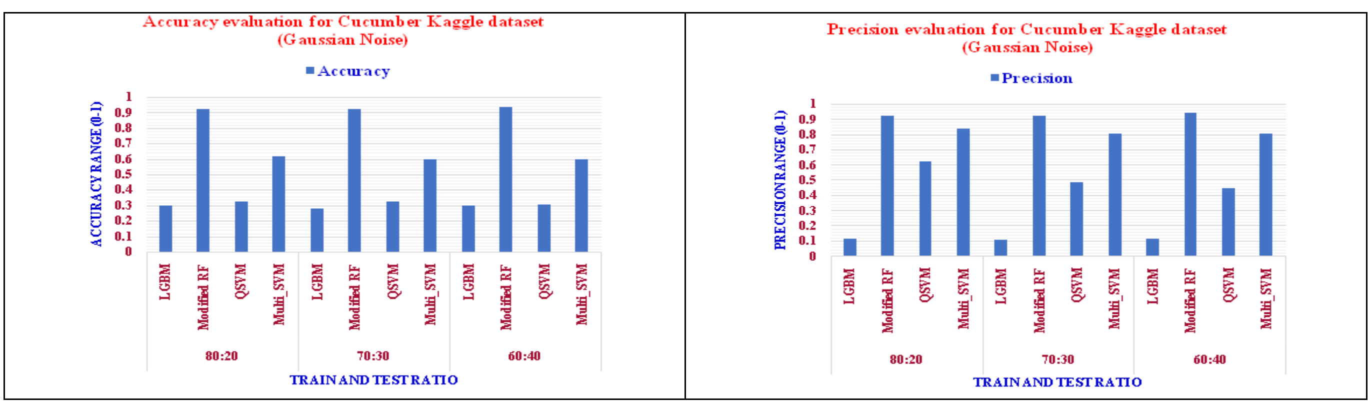


Fig. 10 Evaluation of performance metrics for cucumber custom dataset (denoised with median filter)



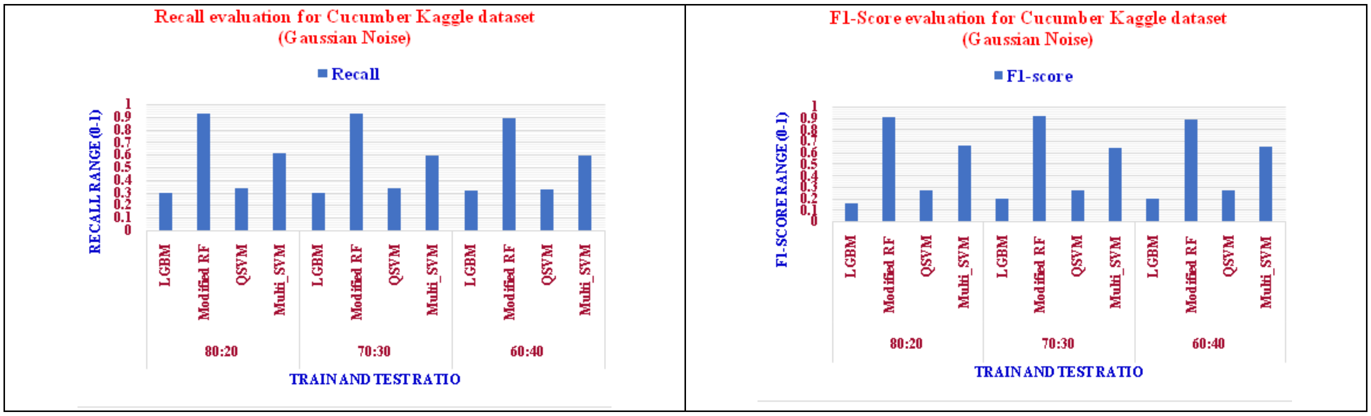


Fig. 11 Evaluation of performance metrics for cucumber kaggle dataset (gaussian noise)

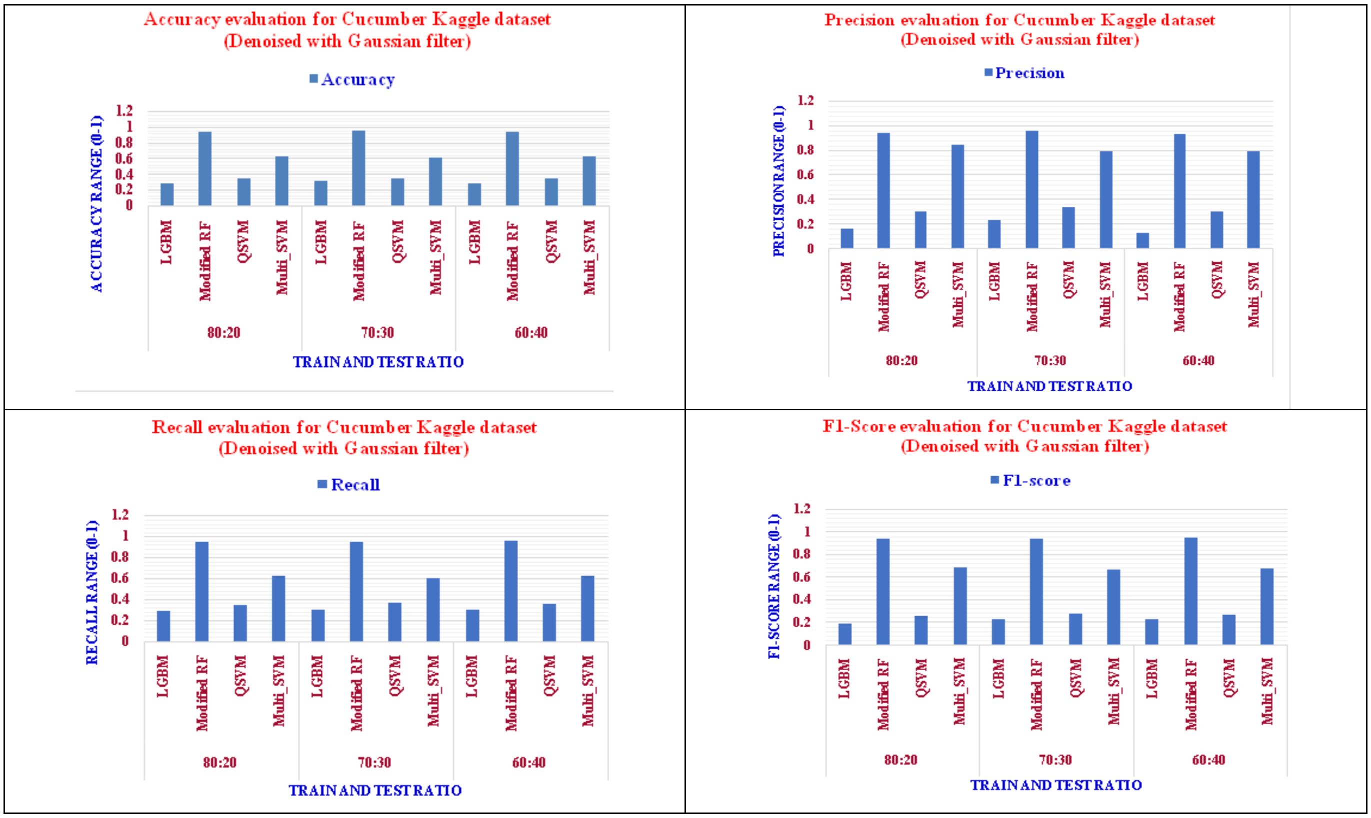
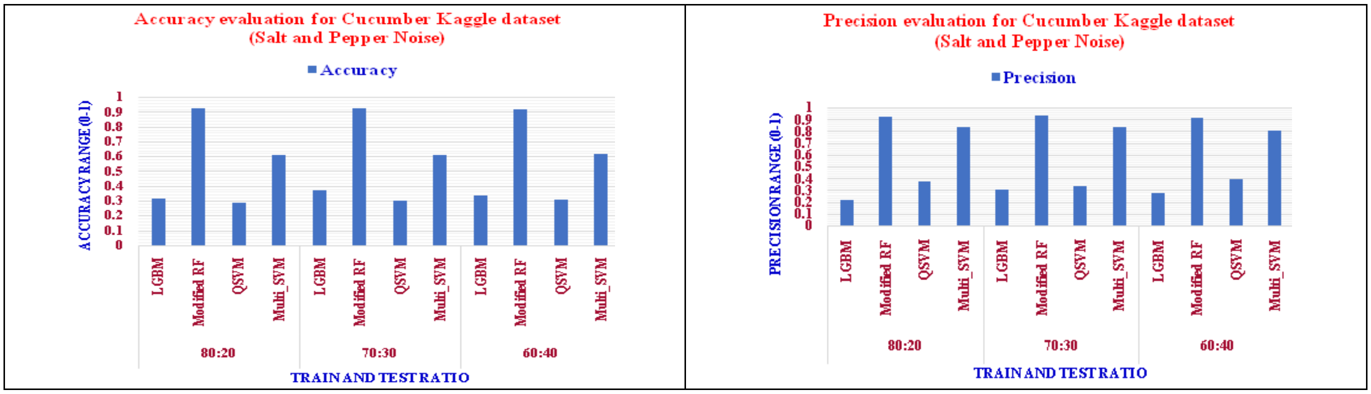


Fig. 12 Evaluation of performance metrics for cucumber kaggle dataset (denoised with gaussian filter)



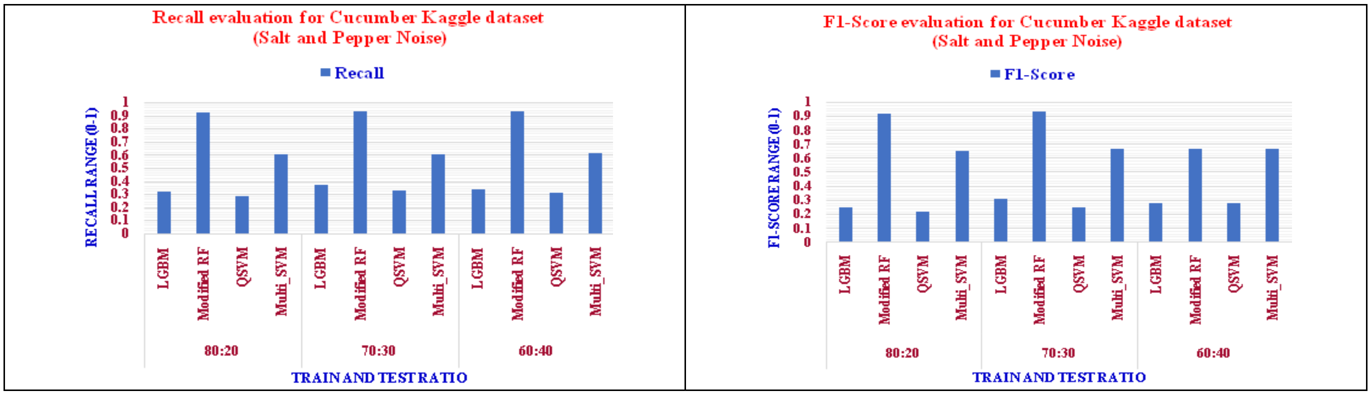


Fig. 13 Evaluation of performance metrics for cucumber kaggle dataset (salt and pepper noise)

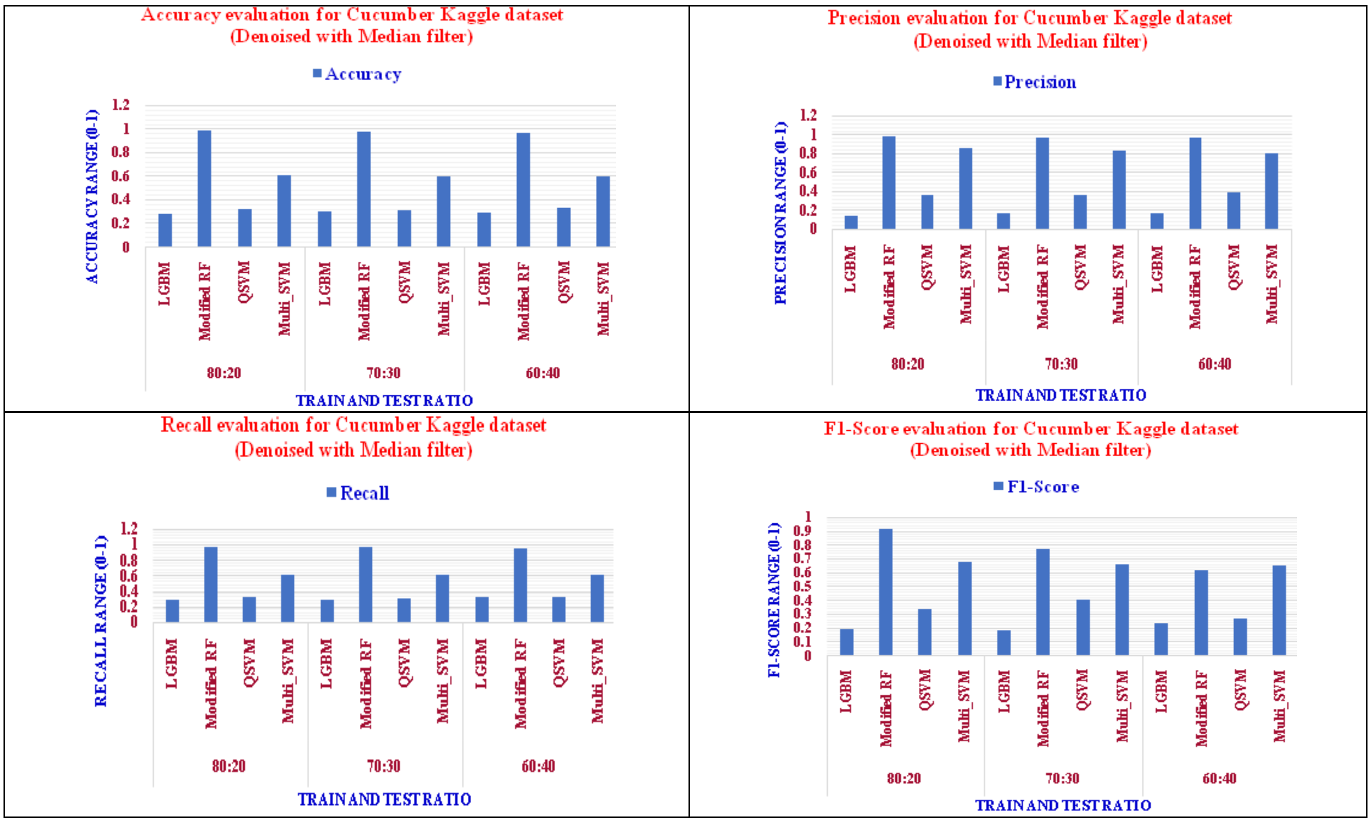
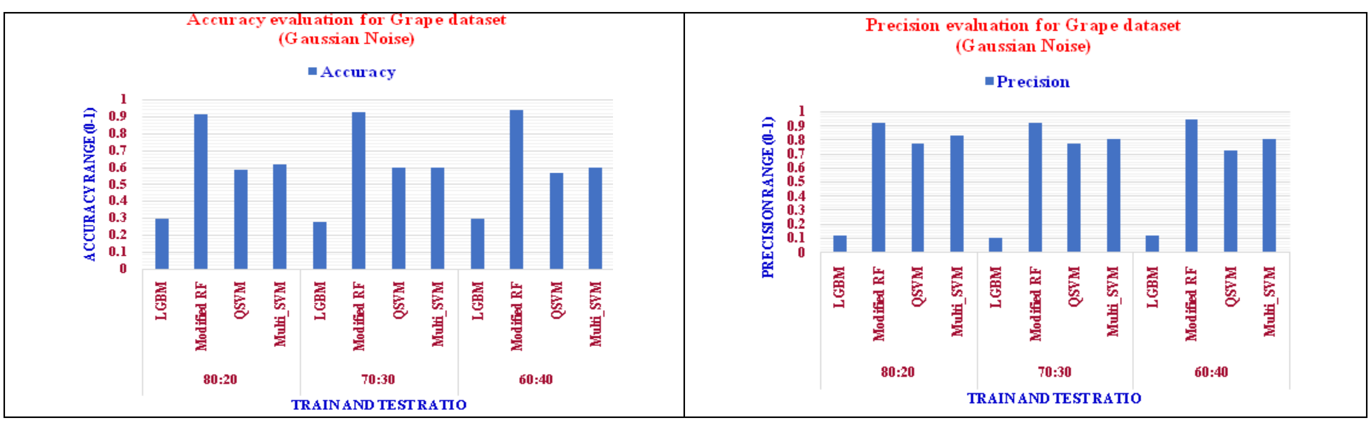


Fig. 14 Evaluation of performance metrics for cucumber kaggle dataset (denoised with median filter)



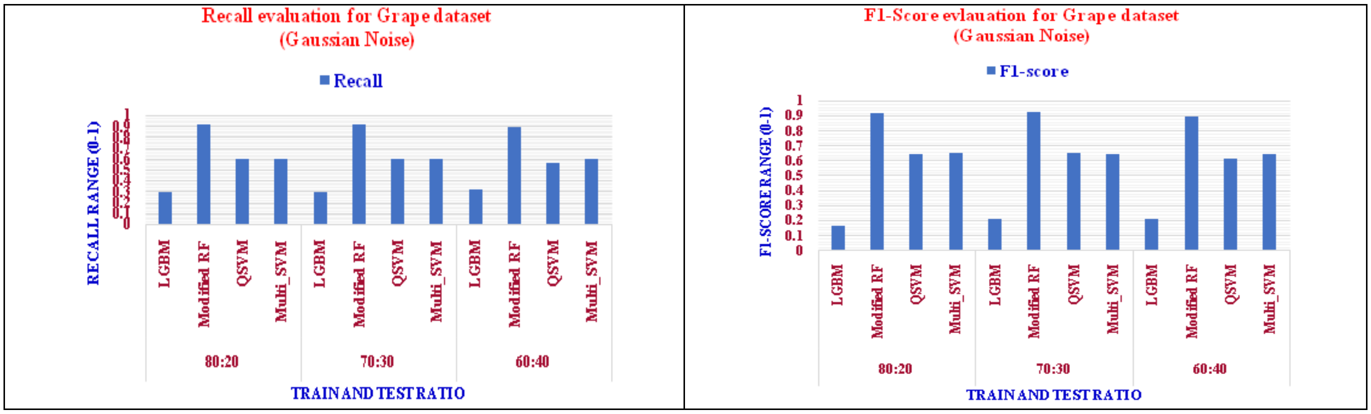


Fig. 15 Evaluation of performance metrics for the grape dataset (gaussian noise)

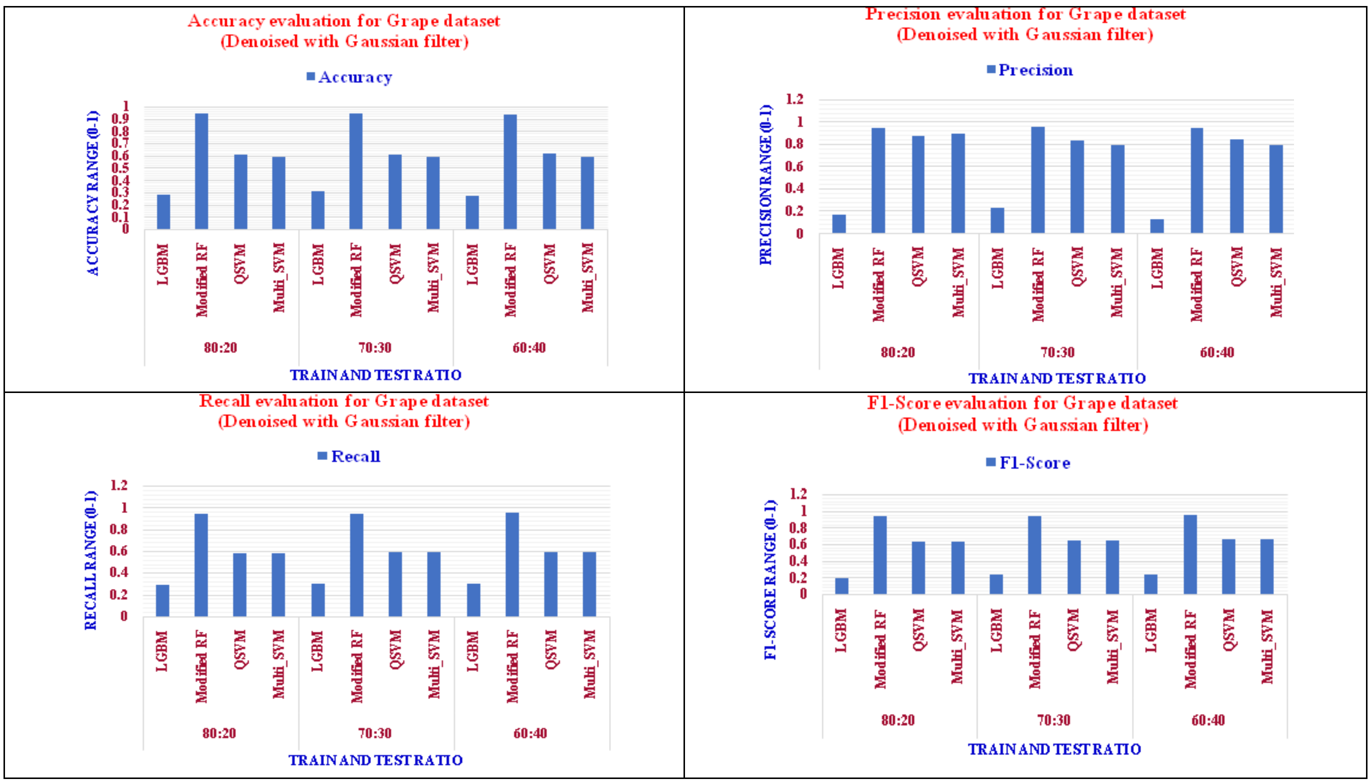
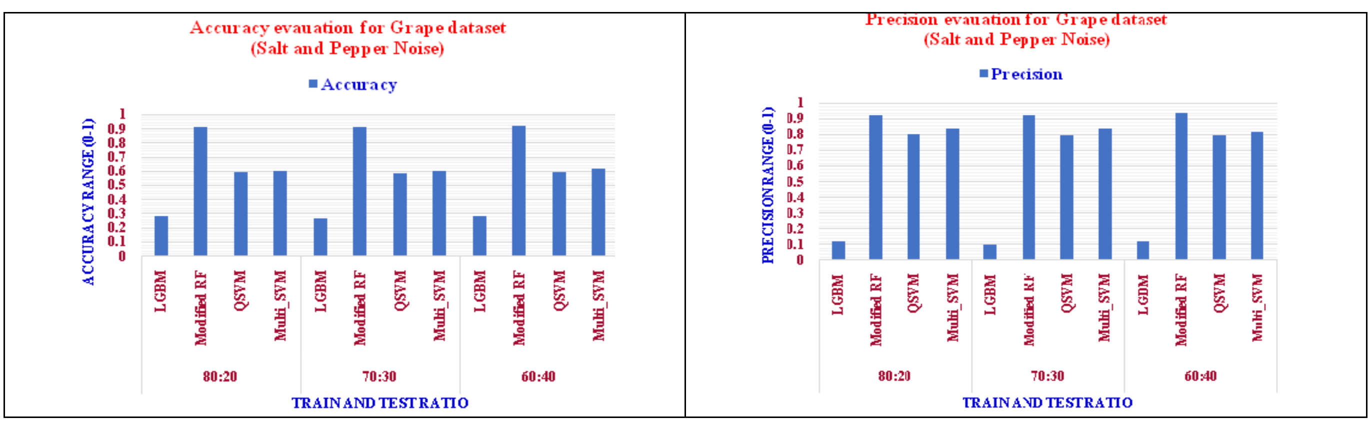


Fig. 16 Evaluation of performance metrics for the grape dataset (denoised with gaussian filter)



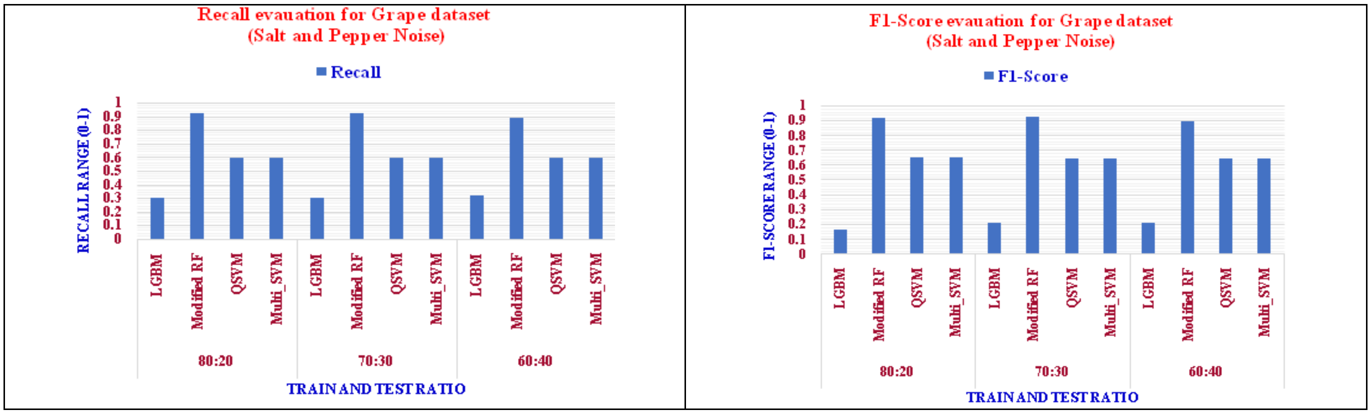


Fig. 17 Evaluation of performance metrics for the grape dataset (salt and pepper noise)

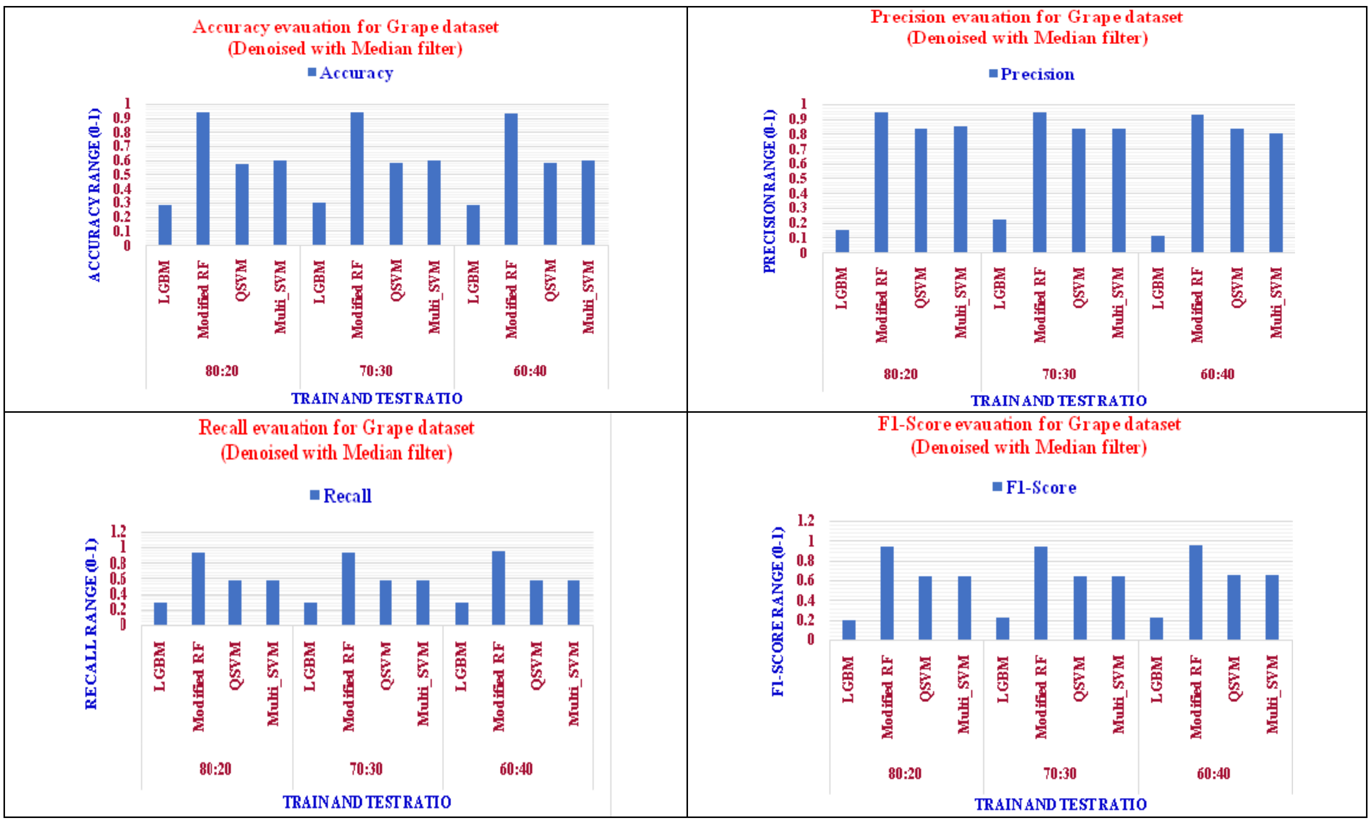


Fig. 18 Evaluation of performance metrics for grape dataset (denoised with median filter)

5. Conclusion

The proposed study on the design and development of a cucumber and grape disease detector, with noisy data, reveals a highly efficient, noise-tolerant methodology for accurate disease detection using Machine Learning Methods. The framework can mitigate the effects of salt-and-pepper and Gaussian noise, achieving superior image quality and feature localization through the combination of the median and Gaussian filters. Among the tested classifiers, the Modified Random Forest was the most accurate, precise, and recall, and F1-score on both custom and Kaggle data, and the rationale behind the high results is that it is robust and performs better when under the noisy setting, generalization. The findings indicate that effective denoising combined with ensemble-based learning contributes to a significant improvement in the model performance and stability in agricultural imaging activities.

5.1. Future Work

The implementation of Deep Learning-based Hybrid Models, Adaptive Filtering, and Automated Noise classification to correct in real-time can be used to extend this work in the future. Moreover, the system, as a part of the IoT-based smart farming and mobile application, could allow diagnosis of the disease on the field, and hyperspectral imaging and transfer learning methods could enhance the early diagnosis of the disease. In general, the research provides a solid foundation for the development of intelligent, scalable, and noise-resistant plant disease detection systems that enable accurate agriculture and sustainable crop management.

5.2. Ethical and Societal Implications

The proposed system supports sustainable agriculture by enabling early disease detection, reducing pesticide misuse, limiting crop loss, and improving food security.

Its computational efficiency also enhances access to precision farming technologies for small-scale farmers. Potential dataset bias due to regional and environmental variations is mitigated by using both public and real-field data, with future work aimed at broader geographic coverage. Overall, the framework promotes responsible and sustainable AI deployment.

Acknowledgements

I want to extend my deepest gratitude to YSR Engineering College, YVU, Proddatur, and my research guide, Dr. S. Kiran, who have been supportive and helpful all along this research project. Their knowledge, support, and resource availability played a key role in the achievement of this research, and I am genuinely grateful for the opportunities that they provided to make this research experience a success.

References

[1] Sharaf Alzoubi et al., "An Advanced Approach for Fig Leaf Disease Detection and Classification: Leveraging Image Processing and Enhanced Support Vector Machine Methodology," *Open Life Sciences*, vol. 18, no. 1, pp. 1-7, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Pa. Andal, and M. Thangaraj, "Comprehensive Review of Methods for Leaf Disease Identification," *Discover Artificial Intelligence*, vol. 5, no. 1, pp. 1-29, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[3] Jingyao Zhang et al., "Identification of Cucumber Leaf Diseases Using Deep Learning and Small Sample Size for Agricultural Internet of Things," *International Journal of Distributed Sensor Networks*, vol. 17, no. 4, pp. 1-13, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Anas Bilal et al., "Fuzzy Deep Learning Architecture for Cucumber Plant Disease Detection and Classification," *Journal of Big Data*, vol. 12, no. 1, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[5] Kaiyu Li et al., "A Severity Estimation Method for Lightweight Cucumber Leaf Disease based on DM-BiSeNet," *Information Processing in Agriculture*, vol. 12, no. 1, pp. 68-79, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[6] Weixia Li et al., "Grape Disease Detection Using Transformer-Based Integration of Vision and Environmental Sensing," *Agronomy*, vol. 15, no. 4, pp. 1-10, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[7] Zinon Zinonos et al., "Grape Leaf Diseases Identification System Using Convolutional Neural Networks and LoRa Technology," *IEEE Access*, vol. 10, pp. 122-133, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Ruiheng Li et al., "High-Performance Grape Disease Detection Method Using Multimodal Data and Parallel Activation Functions," *Plants*, vol. 13, no. 19, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Leonardo Rossi et al., "LDD: A Grape Diseases Dataset Detection and Instance Segmentation," *International Conference on Image Analysis and Processing*, Lecce, Italy, pp. 383-393, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Alessandro Benfenati et al., "Unsupervised Deep Learning Techniques for Powdery Mildew Recognition based on Multispectral Imaging," *arXiv Preprint*, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Manju Bagga, and Sonali Goyal, "Image-Based Detection and Classification of Plant Diseases Using Deep Learning: State-of-the-Art Review," *Urban Agriculture & Regional Food Systems*, vol. 9, no. 1, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Jun Liu, "Deep Learning Method for Cucumber Disease Detection in Complex Environments for New Agricultural Productivity," *BMC Plant Biology*, vol. 25, no. 1, pp. 1-18, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Thandiwe Nyawose et al., "A Review on the Detection of Plant Disease using Machine Learning and Deep Learning Approaches," *Journal of Imaging*, vol. 11, no. 10, pp. 1-37, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[14] Abdul Rehman et al., "Cucumber Leaf Disease Classification using Local Tri-Directional Patterns and Haralick Features," *2021 International Conference on Artificial Intelligence (ICAI)*, Islamabad, Pakistan, pp. 258-263, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Chaerur Rozikin et al., "Classification of the Severity of Downy Mildew Disease Using LGBM," *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, Jakarta Selatan, Indonesia, pp. 364-368, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[16] Kaushi Gihan, Cucumber Leaf Disease Dataset, Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/kaushigihanml/cucumber-leaf-disease-dataset

[17] Rajarshi Mandal, Grapevine Disease Dataset (Original), Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/rm1000/grape-disease-dataset-original