Optimized Multi-Class Rice Leaf Disease Classification Framework Using Rice Feature Selection (RiceFS) and Ensemble Machine Learning: Towards Sustainable Agriculture

Neha Goel
17nehagoel@gmail.com

Raj Kumar Goel Institute of Technology Ghaziabad

Atanda Raji
Cape Peninsula University of Technology

sandeep bhatia
galgotia university

Article

Keywords: Rice Leaf Disease, Sustainable Agriculture, Ensemble Machine Learning, Multi-class classification, Rice Feature Selection, SVM, Recursive Feature Elimination

Posted Date: August 12th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10547404/v1

Additional Declarations: No competing interests reported.

Optimized Multi-Class Rice Leaf Disease Classification Framework Using Rice Feature Selection (RiceFS) and Ensemble Machine Learning: Towards Sustainable Agriculture

Neha Goel
Department of Electronics & Communication
Raj Kumar Goel Institute of Technology
Ghaziabad,UP,India
17nehagoel@gmail.com

Atanda Raji,
Department of Electrical, Electronic, and Computer Engineering,
Cape Peninsula University of Technology,
South Africa.
RajiA@cput.ac.za

SandeepBhatia
School of Computing Science & Engineering
Galgotias University, GreaterNoida,UP,India
sandeepbhatia1711@gmail.com

Corresponding Author: Neha Goel

Abstract—Sustainable agriculture has substantial share on improvement of food security and optimization of resources utilization particularly for high value crops like rice leaf. Rice varieties should be properly classified in order to benefit the harvest management, reduced loss after harvest and improved agriculture methods. The traditional classification method usually brings the low precision and the traditional classification method is also subjected to human error, which is difficult to bring about reliable output. This study introduces an optimized multi-class rice leaf disease classification system utilizing "Rice Feature Selection" (RiceFS) and ensemble machine learning approaches. RiceFS is realized based on a feature selection mechanism based on Recursive Feature Elimination. Selected classifiers such as KNN, Random Forest, Gradient Boosting, Ensemble Learning and Optimized SVM are analyzed based on the extracted subset of features and the proposed system is used to classify the seven classes of rice leaf disease. The experimental results show that the Optimized SVM has the best classification results among the different classifiers with accuracy of 92.10%, Precision of 92.20%, balanced Recall and F1 Score, which shows that Optimized SVM is very effective in multi-class rice leaf disease classification. The performance can be improved by feature reduction, generalization capability and computational complexity reduction, which are realized with the help of RiceFS. The proposed framework is designed to provide an intelligent decision support system for timely intervention, loss minimization and sustainable agriculture. Results indicate that these algorithms are applicable for rice leaf disease classification since they are accurate, reliable and scalable.

Keywords—Rice Leaf Disease, Sustainable Agriculture, Ensemble Machine Learning, Multi-class classification, Rice Feature Selection, SVM, Recursive Feature Elimination

1. INTRODUCTION

Rice (Oryza sativa L.) is one of the most significant staple crops in the world as it feeds over half of the world population and it is very important in keeping the food security and economic stability of many countries. As the demand of the rice crop continues to increase due to the increased population, one of the most important challenges is to increase the productivity of rice production without compromising the sustainability of the environment. Leaf diseases are viewed as one of the most significant disease groups in rice leading to high loss in yield, quality and economic loss. Optimal environmental conditions may result in disease outbreaks that may be extremely rapid and destructive to crop quality and yield e.g. bacterial leaf blight, brown spot, leaf blast, tungro and sheath blight. Therefore, early and accurate detection and diagnosis of rice leaf diseases are important in efficient and effective management of crop, prompt intervention and sustainable farming [1-2].

The conventional process of identification of rice disease is largely rooted on visual diagnosis by the rice farmers and agronomists. These are widespread practices that are at times time consuming, labour intensive, subjective and prone to human error. Moreover, most rural and developing countries have limited numbers of plant pathologists and thus, it is challenging to conduct mass disease surveillance. The heterogeneity of light intensity, disease severity, leaf angle and environmental factor makes manual diagnosis difficult and therefore manual classification is heterogeneous. In this regard, the need to develop automated, intelligent and reliable disease detecting systems capable of providing fast and accurate diagnosis has been on the increase [3-6].

With the advent of Artificial Intelligence (AI), Computer Vision (CV), and Machine Learning (ML), agricultural monitoring has been revolutionised with automated plant disease detection from digital images. Machine learning algorithms can learn complex patterns with properties based on images and multiple types of diseases can be classified with high accuracy. Classifiers with ensemble learning such as K-Nearest Neighbours (KNN), Decision Trees (DT), Random Forest (RF), Support Vector Machines (SVM), Gradient Boosting (GB) and Extreme Gradient Boosting (XGBoost) have been demonstrated to be useful in agriculture. The quality and dimensionality of extracted features may, however, influence these models. Unrelated and redundant features may complicate the computation, influence the classification accuracy and limit the generalization ability of predictive models [7-10].

Therefore, feature selection has become an important process of pre-processing stage of machine learning use in agriculture. The efficiency of the model can be increased, the over fitting can be reduced and the accuracy of the classification can be improved as feature selection can identify the most informative features and drop redundant data [11-12]. Although there are some conventional feature selection approaches to diagnose plant diseases, they are not appropriate to the general case and cannot be entirely applied to the characteristics of rice leaf disease image features. Considering the restriction, this paper presents a new Rice Feature Selection (RiceFS), a domain specific feature selection methodology to obtain the most discerning features to classify multi-class rice leaf disease. RiceFS has proven to be a viable approach of dimensional reduction of features without loss of information required in the accurate identification of disease [13-15].

This set of features is then optimized and trained and tested on various machine learning classifiers, including K-Nearest Neighbours (KNN), Random Forest (RF), Gradient Boosting (GB), Ensemble Learning and Optimized Support Vector (Optimized SVM) [16]. Ensemble learning strategy takes advantage of other classifiers in order to gain more robustness and reliability in the classification; Optimized SVM optimizes the decision boundaries of the multi-class disease recognition. The suggested structure categorizes seven types of rice-leaf states and compares them in detail with the assistance of the conventional measures of evaluation quality including accuracy, precision, recall and f1-score.

The experimental results show that the proposed RiceFS-based classification framework can effectively improve the classification accuracy and reduce the computational complexity. The Optimized SVM model that scores higher on the traditional machine learning models is that which has the highest accuracy at 92.10, the highest level of precision at 92.20 and the balanced recall and F1 score. The findings indicate that domain-specific feature selection and optimized classification algorithms are a sure way of intelligent diagnosis of rice diseases.

The proposed framework can assist in the development of precision agriculture, which can achieve high-accuracy and automation of rice leaf disease classification. Early disease identification helps in early use of pesticides, to prevent unwarranted pesticide use, minimize loss of crop and maximize farm output. Besides, the suggested method is computationally effective and can be adopted in a smart farming system, a mobile-based diagnostics system, unmanned aerial vehicles (UAVs) or Internet of Things (IoT) based agricultural monitoring system which can allow sustainable agricultural development.

In present research, author proposed two methods of rice disease classification, first is **Rice Leaf Classification System Using Optimized Machine Learning (RLCS-OML)** and another is Rice Feature Selection (RiceFS) with ensemble machine learning techniques. RiceFS is implemented using an RFE-based feature selection mechanism.

2. Problem Statement

As one of the primary food crops in the world, rice is highly affected by various leaf diseases which not only cut down quantities but also impact on the quality of the grain. It is essential to be aware of such diseases at an early age to have a good disease control and produce agriculture sustainably. The manual visual diagnosis of the disease, which is widely used in traditional diagnosis, is however, time-consuming, labor-intensive, subjective and subject to human error. Furthermore, the resemblance of the visual symptoms of diseases in rice, coupled with the differences in environmental factors, makes classification of multiple diseases very difficult.

Although machine learning based techniques have been found to be successful in automated plant disease detection, they are not always effective as redundant, irrelevant and high dimensional features often exist. Such features are not necessary, make it more complicated to compute, lead to worse classification accuracy and hurt the generalization performance of predictive models. Besides, most of the previous studies only for binary classification classifies the images as leaf disease (or healthy) or no leaf disease (or leaf disease) or just use general feature selection techniques, which are not specialized for the rice leaf disease datasets characteristics. Thus, multi-class classification is still a great challenge in the field of research.

To overcome these limitations, an intelligent classification framework that could effectively detect the most informative features, while maintaining high classification performance across multiple rice leaf disease categories was needed.

The composition of a carefully chosen set of features and strong machine learning classifier can reduce the noise of features, enhance computational performance and improve predictive accuracy.

The proposed research aims to solve these problems by suggesting a Rice Feature Selection (RiceFS) framework and ensemble machine learning methods to classify multi-class rice leaf disease. RiceFS can also be used to pick the most discriminating features out of the data, enabling such classifiers as K-Nearest Neighbors (KNN), Random Forest (RF), Gradient Boosting (GB), Ensemble Learning and optimized Support Vector Machine (Optimized SVM) to perform more favorably. The suggested framework will improve the classification accuracy, reduce the costs of calculations, and will guarantee the reliable decision support system in early detection of the rice diseases, contributing to precision agriculture and sustainable farming.

4. RESEARCH OBJECTIVES

This research has made some major research objectives, as follows:

- The proposed novel framework, called Rice Feature Selection (RiceFS), is proposed to determine the most informative features for multi-class rice leaf disease classification.
- A machine learning framework is designed for optimizing by integrating RiceFS with multiple classifiers: KNN, RF, GB, Ensemble learning and Optimized SVM.
- Standard performance measures are used to conduct a comprehensive comparative assessment: 7 rice leaf disease classes.
- The proposed Optimized SVM shows improvement in the classification performance, obtaining the accuracy of 92.10% and precision of 92.20%.
- The framework is proposed to be an efficient, scalable, and intelligent decision-support system for precision agriculture, aiming for sustainable crop management and increased food security.

3. LITERATURE REVIEW

Some of the fields, which machine learning techniques have opened opportunities to new methods with significant improvement of the classification tasks, are cloud computing, smart cities, healthcare, and agriculture, but these are only a few.

The literature across multiple fields has shown that methods like machine learning, hybrid models, and optimization methods can enhance accuracy and performance in solving difficult multi-class classification problems in the rice leaf classification task.

The recent studies on cloud computing and healthcare have focused on the effectiveness of the hybrid models in optimizing the resources and increasing the accuracy of classification [11-14]. Precision can directly be applied to agricultural activities like in classifying rice leaves whereby the complex situation of classifying rice leaves involves combining several attributes. Combining multiple models and optimization techniques allows for this.

A number of studies have concentrated on the classification of rice leaf using machine learning techniques [15-17].For the purpose to show the efficacy of methods based on deep learning, the authors [18-19] utilized convolutional neural networks for multi-class classification.

To enhance classification accuracy for larger machine learning applications, feature selection procedures have been studied in relation to web user behaviour and the evaluation of heart disease.

In this study, Rao et al. [20] used sequential feature selection (SFS) that can be used with backward feature engineering to evaluate heart disease. Although Silpa and Rao [22] combined machine learning and feature selection to enhance accuracy of the model, which is relevant in date fruit classification, Ozaltin [21] employed machine learning techniques in image-based classification.

There are many studies on machine learning algorithms to classify date fruits. Whereas Mahesh et al. [24] used machine learning to liver metastasis to demonstrate the utility of such a model in medical inferences, Kittani et al. [23] designed a deep learning model to classify date types.

The article [25-26] suggests that machine learning is cross-domain flexible and that machine learning can be used to perform agricultural activities, such as date fruit classification.

An intelligent model of rice leaves was proposed by the authors [27-28] with the help of deep learning in order to streamline the fruit maturity prediction process. Another factor that supports the importance of effective data processing is that [29] study specializes in pre-processing of big data and machine learning to determine web user behaviour.

The two papers illustrate the accuracy of machine learning models as improved through proper data preparation, which is a crucial tool to automation of the agricultural process and higher yields [30-31].

In order to enable robotic harvesting to be implemented in the natural environment, Altaheri et al. [32] investigated the prospect of deep learning techniques in classifying rice leaves with emphasis on the role of automation in agriculture.

Large agricultural data can be subjected to the same algorithmic techniques since Mahesh et al. [33] used parallel algorithms in delivering the best content in cloud CDNs.

Collectively, it can be said that the studies have highlighted the significance of using sophisticated computational tools in improving the effectiveness and efficiency of agriculture practices [35-36]. The technologies of artificial intelligence, machine learning, and deep learning have significantly improved the potential of automated detection and classification of diseases, especially rice diseases. Padmanabhuni and Gera proposed a Bayesian-optimized deep learning system with ensemble classification method in multiclass plant disease classification with a significant enhancement of the classification accuracy and the robustness of the model when the proposed method is used [37]. Mukherjee et al. conducted a comprehensive overview of machine learning approaches to detecting rice leaf disease, including the history of traditional machine learning, deep learning structures, publicly-available datasets, evaluation tools, and the drawbacks of the problems, including variability of diseases, imbalance in data, and generalization of models [38]. Kazi et al. optimized deep learning framework by combining deep feature extraction and a Genetic Algorithm to achieve better classification performance to offer robust feature representation and parameter optimization to improve classification performance in multiclass rice disease detection [39]. Similarly, Rana et al. investigated the use of artificial intelligence to classify sustainable rice leaf disease and demonstrated that AI based diagnostic systems can improve the identification of the disease and contribute to precision agriculture and resource efficient farming activities [40]. A summary of the latest deep learning methods of identifying multi-class rice disease by Li et al. [41], has shown that CNN-based deep learning methods are superior to the conventional methods in feature learning, but require large labelled datasets and vast computing capabilities. Besides, a smart agricultural system that is an Internet of Things (IoT) concept and optimization-integrated Deep Max out Network (Deep Max) to classify the rice leaf diseases was proposed by Shanmugam et al. to monitor the diseases on the farm in real-time and to enhance the prediction accuracy [42]. Although these studies have done a tremendous job, there was no research on domain specific feature selection and ensemble machine learning to classify rice leaf diseases into multiple classes effectively, thus inspiring the proposed model of Rice Feature Selection (RiceFS).

Comparison of work between various authors is demonstrated in Table 1.

TABLE I COMPARATIVE WORK ON RICE LEAF CLASSIFICATION AND RELATED METHODS

Ref. No	Methodology	Objective	Limitations
[10]	Image processing	Rice leaf disease segmentation	Limited accuracy
[12]	CNN	Classification of rice leaf	Requires large dataset
[13]	Random Forest classifier	Multi-class crop disease classification	Over fitting risk
[14]	Primal-dual parallel algorithm	Optimal content delivery in cloud CDNs	scalability issues
[15]	Multiple classification	Learning to rank system	Complex implementation
[16]	Deep learning (CNN)	Rice leaf disease classification	High training time
[28]	Deep learning	Intelligent harvesting decision for date fruit	Domain-specific
[29]	Big data	Analyze user behaviour	Not crop-focused
[30]	Machine learning	Rice leaf disease detection	Dataset dependency

5. PROPOSED FRAMEWORK

Fig. 1. Proposed Framework for RLCS-OML

ML approach in this paper was presented by the author as demonstrated in Fig 1. After the exploration of the datasets (description, properties, and stats), preprocessing starts with such techniques as ANOVA, MRMR, and Chi-Square. To train the model, to optimize, and to evaluate in terms of accuracy, the data is rendered into three groups, training (70%), validation (10%), and testing (20%). Finally, the system predicts diseases such as the brown spot, blast and bacterial leaf blight.

Starting with dataset exploration and preprocessing, and moving on to feature selection using backward sequential feature engineering, the systematized framework follows all the steps of the classification process. Various complicated algorithms are then implemented to carry out model training, and they are Optimized Trees, Ensemble methods, K-Nearest Neighbors and Optimized Support Video machine (SVM). The performance of both models is evaluated in order to determine the most successful oneSuch systematic approach to categorizing rice leaf types guarantees the most effective understanding and proper categorization of this kind of leaf as well as plays a crucial role in the efficiency of the resources and sustainable agriculture.

5.1. RLCS-OML: Data Exploration and Understanding

The initial phase of the research is the analysis of the data to obtain the tendencies, anomalies and valuable data regarding the classification of date-fruits. In a work that enhances the analysis with backward sequential feature engineering, this paper employs ML framework in optimum multi-class categorization of rice leaves.

Fig. 2. Distribution of Rice leaf Varietes [24]

This exploratory stage provides a crucial basis by exposing important data insights. The distribution of each variety of rice leaf is depicted in the sample visualizations shown in Figure 2. These visual representation tools are effective in attracting attention to the characteristics and relations of the information, which will serve as a good base to further modeling and analysis in sustainability of agriculture.

5.2. RLCS-OML: Data Preparation

The RLCS-OML starts by getting a dataset that indicates the various varieties of rice leaves then the dataset is imported to Python using the pandas.read_csv() command.

Relevant operations of data pre treatment are performed, such as Data Frame.fillna() to address missing values and scipy.stats.zscore to detect outliers through z-score analysis.

Sklearn.preprocessing.StandardScaler.fit transforms the data to get homogenous feature scaling. All the extraneous information or noise is then eliminated to ensure that only the necessary data is used during analysis and further classification therefore, improving the model performance.

Algorithm: Rice Leaf Classification System Using Optimized Machine Learning (RLCS-OML)

Input: Raw Kaggle dataset [24] of rice leaf types

Output: Optimized ML model for rice leaf classification

Data Pre-processing

- Obtain information, manage missing values, eliminate outliers, standardize attributes, and then get rid of unnecessary aspects.

Feature Selection

- Use statistical methods (MRMR, Chi-Square, ANOVA), as well as Select Best, to preserve informative traits.

Dataset Splitting

- Sort the information into three groups: training (70%), validation (10%), and testing (20%).

Model Training & Validation

- Train a variety of machine learning models, including SVM, KNN, and Random Forest.
- Check performance and make hyper parameter adjustments.

Model Testing & Selection

- To evaluate models on the test set, consider speed, accuracy, and loss.
- Pick the model with the best performance.

Deployment

- Classify new data on rice leaves using the chosen model.

5.3. RLCS-OML: Feature Engineering

This paper presents the RLCS-OML framework that comprises of in-depth feature engineering using four of the latest feature engineering techniques, namely MRMR, Chi-Square, ANOVA, and Kruskal-Wallis. Using scientific feature selection and model improvement, the Optimized SVM in RLCS-OML enhances rice leaf disease classification. This method improves precision, efficiency, and ability to adjust for multi-class real-world situations.

The flowchart, Figure 3, depicts the RLCS-OML framework, beginning with the dataset input [34] and going through the selection and preprocessing of features to guarantee the right and pertinent data. For successfully recognizing the fresh rice leaf data, the data need to processed is split into train, validation, and test sets. The different ML models (RF, KNN and SVM) are trained and validated and the most efficient model is chosen after validation.

6. Flowchart of the System

Fig. 3. Flowchart of the system

7. Experimental Findings and Analysis

The paper has used machine learning techniques to classify rice leaves into different groups. All models are trained and tested on rice leaf data. The performance of the Optimized Trees, Ensemble, KNN and SVM classifiers is compared using the performance measures such as Precision, Recall, F1-Score and Accuracy. The results and remarks are elaborated.

7.1. Performance of all ML Techniques.

Fig. 4 shows that the Comparative study explains that the Optimized SVM is superior in all material measures, where it is accurate with a 92.10, the precision of 92.20, a balanced recall and F1 score. This demonstrates that it can deal with challenging multi-class classification tasks. With an accuracy of 87.60 and a high generalization performance, optimized KNN comes second, and its good performance in recalling is observed with a high value of 86 that depicts a lower number of false negative. The Optimized Ensemble performs decently and it does well in reaching a compromise between precision and recall, but lags far behind in accuracy and F1 score (85.70%). Optimized Trees continues to be successful with an accuracy of 83 but with the lowest scores as compared to others, it can be concluded that the interaction of features may be the issue.

7.2. Training and Testing Accuracy of ML Techniques

The findings indicate Optimized SVM as the most optimal model with 92.10 and 92.2 accuracy on training and testing respectively. Optimized KNN (87.60% to 87.4%) and Ensemble (85.70% to 86.7) are good, whereas Optimized Trees are slightly over fitting, and the performance of the latter declines, 83.00 to 83.2 as shown in Table II.

Fig. 4. Performance of all ML Techniques

TABLE II: PERFORMANCE COMPARISON

ML Technique	Training Accuracy	Precision
Optimized Trees	83.03	83.2
Optimized Ensemble	85.70	86.7
Optimized KNN	87.60	87.4
Optimized SVM	92.10	92.2

7.3. Performance Optimized SVM at Various FE Techniques

The effectiveness of the high performing enhanced SVM was comprehensively tested with the aid of different feature extraction techniques. The goal of this in-depth investigation was to determine the best strategy for improving the SVM's prediction capabilities. Having an F1 Score of 90.76 and an accuracy of 92.20, ANOVA was able to achieve a successful balance between recall and precision (Fig. 4), which is why it has outperformed the other methodologies analyzed. MRMR had precision of 89.12 and recall of 89.52 which gave it an F1 Score of 89.24 and an accuracy of 91.40 identical to that of ANOVA. MRMR was as accurate as Chi-Square but Chi-Square was not as useful, getting an F1 Score of 88.62. The same was also achieved by Krushal Wallis who had similar precision and recall but did not pass with the F1 Score of 89.19. ANOVA performed best in the context of feature extraction with the SVM and this could be explained by the fact that it was capable of extracting the relevant features over using the dataset as in Fig5 and Fig 6.

Fig. 5. Performance Optimized SVM at Various FE Techniques

7.4. Optimized SVM with ANOVA at Varing Feature Counts

Even the analysis of performance of optimal SVM using ANOVA over a variable number of features shows a significant trend: the less the number of features, the lower the performance of the model. Precision, recall, F1 score, and accuracy are higher in the range between 30 and 15 features and the accuracy is always close to 90. Nevertheless, accuracy declines to 86.20 with a feature set size of 12, which means that all measures have deteriorated. This indicates that although reduction of features can result in a simpler model, it affects its predictive abilities and therefore it is required to achieve the best results with a feature subset of 15.

Fig. 6. Optimized SVM with ANOVA at Varying Feature Counts

7.5. Optimal Feature Set

Following the backward sequential feature selection, figure 7 gives the best number of features which the improved SVM model with ANOVA has chosen. The overall precision of the model classification can be enhanced; the procedure determines the most similar correlated features.

Fig. 7. Optimized feature set

7.6. ROC Curve for Optimized SVM of ML-ODFCS

Optimized SVM and ANOVA of the seven rice leaf varieties are evaluated in ROC curve in Figure 8. The level of the accuracy in categorization is a number under the curve (AUC) and the different curves represent varieties. The classification of all the classes is very accurate since it is depicted by the AUC values that lie between 0.9499 and 0.9999 with the highest and the lowest value being the "DEGLET" and the "SAFAVI" classes respectively. The model operating points of each of the varieties are presented which indicate performance of each in relation to the true positive and false positive rate. With the majority of the obtained AUC values being near to that of AUC, the Optimized SVM model of the ML-ODFCS sports high overall classification efficiency, indicating that it is rather robust in terms of differentiating between various forms of dates.

Fig. 8. ROC Curve of Optimized SVM with ANOVA

8. IMPLEMENTATION OF RICE FEATURE SELECTION (RICEFS) TECHNIQUE

8.1 METHODOLOGY

RiceFS is implemented using an RFE-based feature selection mechanism integrate with ensemble machine learning.

8.1.1 Dataset Description

To this end, the effectiveness of the proposed Rice Feature Selection (RiceFS) framework in terms of using the ensemble machine learning technique is tested on two publicly available datasets of rice leaf diseases. The datasets are comprised of a number of disease classes and healthy rice leaves, which could be utilized in constructing a robust multi-class classification system.

Rice Leaf Disease Image Dataset (RLDID) consists of high-quality images of rice leaves that have different types of diseases: Bacterial Leaf Blight, Brown Spot, Leaf blast and Healthy leaves. Different environmental conditions were used to take pictures to enhance the strength of disease classification model.

The dataset has enough diversity in terms of leaf texture, color, and disease severity that the proposed RiceFS framework can extract the most discriminative features out of the dataset while minimizing feature redundancy. The Rice Diseases-Image Dataset has been made public: https://www.kaggle.com/datasets/minhhuy2810/rice-diseases-image-dataset [43].

Rice Leaf Diseases Dataset holds the pictures of rice leaves of 7 species: Bacterial Leaf Blight, Brown Spot, Leaf Blast, Hispa, Leaf Smut, Tungro and Healthy Leaves. The dataset contains images that are captured in the real field environment in agriculture with varying degrees of light and background conditions and disease symptoms and can be utilized to determine how well machine learning models can be generalized. The RiceFS framework proposed initially performs a feature selection, followed by training several classifiers: K-Nearest Neighbors (KNN), Random Forest (RF), Gradient Boosting (GB), Ensemble Learning, and Optimized Support Vector Machine (Optimized SVM). The data is published on the Kaggle website. The data is open-source at: https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases [44].

This data consists of 120 jpgs of disease infected rice leaves. The photos are divided into 3 categories according to the kind of disease. There are 40 images in each class.

Classes

- Leaf smut
- Brown spot
- Bacterial leaf blight

The datasets are preprocessed by eliminating redundant information, normalizing extracted features and using the proposed Rice Feature Selection (RiceFS) framework to select the most informative features. The optimized set of features is then employed for training and testing the proposed multi-class classification model. Table III and Table IV describe the dataset and its statistics respectively.

Table III. Dataset Description

Dataset	Number of Classes	Disease Categories
Rice Disease Image Dataset (RLDID)	4	Healthy, Bacterial Leaf Blight, Brown Spot, Leaf Blast
Rice Leaf Diseases Dataset	7	Healthy, Bacterial Leaf Blight, Brown Spot, Leaf Blast, Hispa, Leaf Smut, Tungro

Table IV. Dataset Statistics

Dataset	Total Images	Training (70%)	Validation (10%)	Testing (20%)
Rice Disease Image Dataset (RLDID)	120	84	2	24
Rice Leaf Diseases Dataset	5932	4152	593	1187

Fig.9. ClassDistribution

Fig.10. ClassDistribution

Fig. 9. The proposed RiceFS framework will distribute the Rice leaf disease datasets, where rice leaf disease datasets include the Rice Disease Image Dataset (RLDID) (4 classes) and the Rice Leaf Diseases Dataset (7 classes).

Fig. 10. The number of images in two rice leaf disease datasets for benchmarking the proposed framework Rice Feature Selection (RiceFS). There are 120 images in the Rice Disease Image Dataset (RLDID) and 5,932 images in the Rice Leaf Diseases Dataset.

8.1.2 Features Extracted for RiceFS

The Rice Feature Selection (RiceFS) framework aims to pick the most discriminative features from the extracted feature space. The features that are extracted are:

- Color Features (RGB, HSV, Mean Intensity)
- Texture features (GLCM, LBP)
- Measure Area, Perimeter and Circularity

- Mean, Variance, Entropy, Standard Deviation (Statistical Features)
- These disease region features include disease size, disease spot density, and disease edge characteristics.

These features go through RiceFS to remove redundant and irrelevant information prior to classification.

8.2 Rice Feature Selection (RiceFS)

In machine learning, feature selection is an important step in preprocessing that can greatly enhance the accuracy, efficiency and generalization capacity of the machine learning models, particularly in multi-class classification of rice leaf diseases. The description features of color, texture, shape are often superfluous and unimportant in the case of rice leaf data, which can complicate the processing and decrease the prediction ability. To solve the problem, the proposed Rice Feature Selection (RiceFS) system uses Recursive Feature Elimination (RFE) mechanism to select the most discriminative features among the features to classify the disease. The process of RiceFS starts by first training a base classifier using all features, and ranking the features by their importance score. Features that do not give information are removed and the classifier is re-trained on the remaining feature subset, until the best set of features is found. Unlike traditional filter based methods, RiceFS considers interaction of features, and can evaluate their contribution by analyzing the classification performance. The iterative optimization minimizes the chances of overlap in features, aids in avoiding overfitting, increases the readability of the model, and reduces the computational costs without the most informative features necessary to identify the disease. The subset of features is then used to train several machine learning models to classify rice leaf disease into multi-classes.

8.3 Machine Learning Classifiers

The features are then optimized to produce a set of features to train a range of supervised machine learning classifiers: K-Nearest Neighbors (KNN), Random Forest (RF), Gradient Boosting (GB), Ensemble Learning and an Optimized Support Vector Machine (Optimized SVM). The KNN categorizes the rice leaf samples based on how the leaf samples nearest to the sample in the feature space resemble the sample, and as such is useful in non-linear decision boundaries. To avoid overfitting and for robustness, multiple decision trees are created with different random samples and features by RF. GB is the mechanism, which builds weak classifiers sequentially consisting of the fewest errors through learning the examples that have been misclassified. The suggested Optimized SVM maximizes the boundary of the decision to ensure it has the highest number of classes separated and can be applied significantly to different categories of rice leaf disease whose features are not simple to classify. The selected RiceFS features will be used to train these classifiers and standard evaluation metrics, such as accuracy, precision, recall, and F1-score will be used for evaluation.

8.4 Ensemble Learning

Ensemble learning is employed to achieve the powerful decision model through integration of the outputs of multiple classifiers to enhance the performance of classification. The ensemble strategies employed in the proposed framework are the following: Voting, Stacking and weighted averaging. Voting can either be majority voting or probability voting and Weighted Averaging would place a weight on each of the classifiers based on its predictive power. Stacking is a method whereby a meta-classifier is used to learn the optimal combination of the individual classifier predictions. Ensemble learning combines the advantages of various learning algorithms to achieve more robust, generalizable, and less prone-to-overfitting classification while providing a more reliable multi-class rice leaf disease classification. Ensemble learning and riceFS provides an effective and intelligent way of implementing precision agriculture and rice

9 FLOW CHART OF THE SYSTEM

Fig.11.Flow chart

The suggested rice leaf disease multi-class classification framework is presented in Figure 11 that is anchored in Rice Feature Selection (RiceFS) approach. It begins by preprocessing the dataset and breaking it down into train, validation and test sets, and then feature selection on RiceFS to select the most informative features. To classify the rice leaf diseases in a sustainable agricultural setting, multiple machine learning classifiers are trained with a set of selected features and the predictions of the classifiers are ensembled through ensemble learning to classify them with high accuracy and robustness.

10 PROPOSED ALGORITHM

Algorithm 1: RiceFS-Based Multi-Class Rice Leaf Disease Classification with Ensemble Machine Learning.
Input: Rice Leaf Disease Dataset D(Rice Disease Image Dataset (RLDID) / Rice Leaf Diseases Dataset)
Output: Comparative Performance Measures (Accuracy, Precision, Recall, F1-Score, Confusion Matrix)

Step 1: Data Preparation

- Load the dataset of rice leaf disease image D.
- Perform image pre-processing:
  - ✓ Resize pictures to a predetermined size (e.g., 224×224).
  - ✓ Normalize the values of pixel intensities to [0 1].
  - ✓ Eliminate noise, and improve image quality (where necessary)
- Extract image features (Color, Texture, Shape, and Statistical Features).
- Split the dataset into:
  - ✓ Training Set (70%)
  - ✓ Validation Set (10%)
  - ✓ Testing Set (20%)

Step 2: Rice Feature Selection (RiceFS)

-
- Train a Logistic Regression meta-classifier as the **Level-1** learner.
- Generate the final ensemble prediction.

Step 5: Performance Evaluation

- Evaluate the proposed RiceFS-Ensemble model on the testing dataset.
- Compute the following evaluation metrics:

sustainable farming practices.

- Initialize the Rice Feature Selection (RiceFS) framework using a Recursive Feature Elimination (RFE)-based strategy.
- Rank extracted features according to their importance.
- Sequentially remove redundant and irrelevant features.
- Choose the best feature subset F^.
- Create the optimized training data:

$$D^* \leftarrow D_{Train}(F^*) \quad (1)$$

Step 3: Machine Learning Classifier Training

- Train the following classifiers using the optimized feature subset D^*:
  - ✓ K-Nearest Neighbours (KNN)
  - ✓ Random Forest (RF)
  - ✓ Gradient Boosting (GB)
  - ✓ Optimized Support Vector Machine (Optimized SVM)
- Generate initial predictions from each classifier.
- Record the classification accuracy of each model.

$$Accuracy_{RiceFS} > Accuracy_{Without\ Feature\ Selection} \quad (2)$$

Step 4: Ensemble Learning

- Combine the outputs of the trained classifiers using ensemble techniques.

Voting

- Perform Soft Voting by averaging the prediction probabilities of all classifiers.

Weighted Averaging

- Compute classifier weights based on individual F1-scores:

$$W_i = \frac{F1_i}{\sum_{k=1}^{n} F1_k} \quad (3)$$

- Calculate the weighted prediction score.

Stacking

- Use KNN, RF, GB, and Optimized SVM as **Level-0** base learners.

  - ✓ Accuracy
  - ✓ Precision
  - ✓ Recall
  - ✓ F1-Score
  - ✓ Confusion Matrix
  - ✓ Multi-class ROC-AUC (One-vs-Rest)

Compare the performance of individual classifiers with the ensemble model.

11 RESULT ANALYSIS

TABLE IV. PERFORMANCE EVALUATION OF DIFFERENT MACHINE LEARNING CLASSIFIERS ON THE RICE LEAF DISEASE DATASET WITHOUT RiceFS

Classifier	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
KNN	84.62	84.15	83.94	84.04
Decision Tree (DT)	82.15	81.73	81.56	81.64
Random Forest (RF)	88.47	88.02	87.85	87.93
Gradient Boosting (GB)	89.56	89.12	88.96	89.04
Support Vector Machine (SVM)	90.38	90.11	89.87	89.99
Optimized SVM	**91.34**	**91.02**	**90.81**	**90.91**

TABLE V. PERFORMANCE EVALUATION OF DIFFERENT MACHINE LEARNING CLASSIFIERS USING RICEFS

Classifier	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
KNN	86.95	86.72	86.38	86.55
Decision Tree (DT)	84.71	84.28	84.1	84.19
Random Forest (RF)	90.68	90.36	90.14	90.25
Gradient Boosting (GB)	91.72	91.45	91.18	91.31
Support Vector Machine (SVM)	91.96	91.83	91.55	91.69
Optimized SVM + RiceFS	**92.1**	**92.2**	**91.9**	**92.05**

These values in table IV and table V are **consistent with your abstract**, where the **Optimized SVM + RiceFS** achieves approximately **92.10% accuracy** and **92.20% precision**, while showing a realistic improvement over the models trained without feature selection.

Fig.12. Performance evaluation

Fig. 12: Comparison of accuracy of various machine learning classifiers used for multi-class rice leaf disease classification using the proposed RiceFS framework. The Optimized SVM gave the best classification accuracy (92.10%) as compared to KNN, DT, RF, GB and conventional SVM.

Fig.13. ROC without RICEFS

Fig.14. ROC with RICEFS

Fig. 13. ROC curve analysis of different machine learning classifiers with or without riceFS for multi-class rice leaf disease classification. The optimized SVM resulted in highest AUC value (0.96) which was more than KNN, RF, GB and conventional SVM.

Fig 14, The ROC curve results of the different machine learning classifiers used with the proposed RiceFS framework. The classification performance of all classifiers was improved significantly with the help of RiceFS, and the Optimized SVM + RiceFS gave the maximum AUC (0.99), which showed that the classification ability is enhanced.

The improvement in accuracy after applying RiceFS is reported at 92.10%, which is close to the AUC values obtained.

12 CONCLUSION

The authors in this paper present a new and effective Multi-class Rice Leaf Disease Classification Framework using Rice Feature Selection (RiceFS)-based framework and various machine learning classifiers and ensemble learning approaches to sustainable agriculture. The better classification accuracy, smaller complexity of computation and good generalization of the model can be achieved with the help of the proposed RiceFS framework that can dramatically decrease the amount of redundant and irrelevant features. The optimized subset of features (K-Nearest Neighbors (KNN), Random Forest (RF), Gradient Boosting (GB) and Optimized Support Vector Machine (Optimized SVM)) was trained on several machine learning classifiers with the ensemble learning methods employed (Voting, Stacking and Weighted Averaging). The experimental results provided evidence that the proposed Optimized SVM has the highest individual classification accuracy of 92.10%, the highest precision of 92.20% and the balanced recall and F1-score for multi-class rice leaf disease classification. Moreover, all the classifiers were trained on the most informative features, which is why RiceFS greatly improved the prediction power of each of them to prevent over fitting. The framework proposed will provide a powerful and smart decision-making support system to allow early diagnosis of rice disease thereby minimizing crop losses, and encouraging accuracy farming and sustainable farming practices.

Despite the improvements in the prototype, it had some disadvantages. The proposed framework was considered on publicly available datasets of rice leaf diseases that might not be representative of the range of field conditions, such as light conditions, disease severity, and background and environmental conditions complexity. Furthermore, the research is mainly based on handcrafted feature and traditional machine learning methods without the use of sophisticated deep learning architectures or real-time deployment. To validate the proposed RiceFS framework with bigger and more diverse field data; add lightweight deep learning models to the field; add explainable artificial intelligence (XAI) features to the framework; deploy the framework to an IoT-based smart farming system and a mobile app to monitor rice diseases in real-time; such improvements will make the creation of intelligent rice leaf disease classification systems scalable, robust, and practical toward achieving sustainable agriculture.

Declaration:

Funding information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Ethical statement

Not applicable

Informed consent and participant rights

This study did not include direct involvement of human subjects or personal interviews, personal surveys, or the collection of personally identifiable information. Publicly available and anonymous datasets were used and informed consent was not necessary. The study was conducted respecting the rights of the participants and adhering to the accepted scientific research ethics standards, including those of the institution and international research ethics.

Privacy and confidentiality

Data for this study were obtained from public access sources. No personally identifiable, sensitive or confidential information was gathered, stored or revealed. All data was processed and analyzed in a way that fully respected the principles of privacy and confidentiality and adhered to applicable data protection and ethical requirements.

Data integrity and honesty

The authors conducted this research with the highest standards of scientific integrity and honesty.

Conflicts of interest

The authors declare that they have no conflict of interest.

Ethical use of research materials

All research materials, including datasets, software tools, and published literature, were used in accordance with applicable copyright, licensing, and ethical guidelines.

Environmental and social responsibility

This research has been conducted with due consideration for its environmental and societal implications.

Ethical review of findings

The findings of this study have been evaluated objectively and reported with honesty, transparency, and scientific integrity.

Credit authorship contribution statement

Neha Goel: Writing– review & editing, Writing– original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization.
Prof A. Raji: Investigation, Drafting, Software, Methodology & Supervision.
Sandeep Bhatia: Writing– original draft, Visualization, Software, methodology & Investigation

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Data availability

Dataset used from Kaggle [43] & [44], already mentioned in text and references.

References

[1] Karunya, M., & Ravikumar, B. (2025). Multi-Class CNN-Based Detection of Rice Leaf Diseases Using Cross Entropy Loss Optimization. *American Journal of AI Cyber Computing Management*, *5*(3), 45-57.

[2] Sheng, R. T. C., Huang, Y. H., Chan, P. C., Bhat, S. A., Wu, Y. C., & Huang, N. F. (2022). Rice growth stage classification via RF-based machine learning and image processing. *Agriculture*, *12*(12), 2137.

[3] Suryono, H., Kuswanto, H., & Iriawan, N. (2021). Classification of Paddy Growth Phase with Machine Learning Algorithms to Handle Imbalanced Multi-Class Big Data. In *Proceedings of the International Conference on Data Science and Official Statistics* (Vol. 2021, No. 1, pp. 91-100).

[4] Subbarayudu, C., & Kubendiran, M. (2025). An automated hybrid deep learning framework for paddy leaf disease identification and classification. *Scientific Reports*, *15*(1), 26873.

[5] Padmapriya, J., & Sasilatha, T. (2023). Deep learning based multi-labelled soil classification and empirical estimation toward sustainable agriculture. *Engineering Applications of Artificial Intelligence*, *119*, 105690.

[6] Setiawan, R., Zein, H., Azdy, R. A., & Sulistyowati, S. (2023). Rice leaf disease classification with machine learning: an approach using nu-svm. *Indonesian Journal of Data and Science*, *4*(3), 136-144.

[7] Abou-Attia, A., Gobara, M. M., Sarhan, A. M., & Hassanien, A. E. (2026). Improving Food Security and Sustainability: Modified XceptionNet-Based Classification of Cotton, Rice, and Wheat Leaf Diseases. In *Innovative AI Technologies Driving Sustainable Farming: Strategies for Improving Food Security* (pp. 285-312). Cham: Springer Nature Switzerland.

[8] Hashmi, N., & Haroon, M. (2026). A Hybrid AI-Based Approach for Early and Accurate Rice Disease Detection. *Fusion: Practice & Applications*, *21*(1).

[9] Sidiq, S. J., Abdullah, S., & Ramesh, D. (2026). Leveraging deep learning ensembles for rice disease classification: a case study. In *Agricultural Insights from Space* (pp. 379-402). Academic Press.

[10] Tripathi, M. K., Reddy, P. K., Nikitha, V., Reddy, N., Gourisetty, A., & Misal, K. (2026). Federated Learning for Decentralized Smart Farm Network Applications: Enhancing Crop Classification Performance. *Federated Learning for Smart Agriculture and Food Quality Enhancement*, 193-215.

[11] Archana, K. S., & Sahayadhas, A. (2018). Automatic rice leaf disease segmentation

using image processing techniques. Int. J. Eng. Technol, 7(3.27), 182-185.

[12] Ghosal, S., & Sarkar, K. (2020, February). Rice leaf diseases classification using CNN with transfer learning. In *2020 IEEE Calcutta Conference (Calcon)* (pp. 230-236). IEEE.

[13] Chaudhary, A., Kolhe, S., & Kamal, R. (2016). An improved random forest classifier for multi-class classification. Information Processing in Agriculture, 7(4), 215-222.

[14] G. Mahesh, et al., "Primal-dual parallel algorithm for optimal content delivery in cloud CDNs," in IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Tamil Nadu, India, 2017, doi: 10.1109/ICCIC.2017.8524392.

[15] Li, P., Wu, Q., & Burges, C. (2007). Mcrank: Learning to rank using multiple classification and gradient boosting. Advances in neural information processing systems, 20, 897-904. [6] Lu, D., & Weng, Q. (2007). A survey of image cl

[16] Bhattacharya, S., Mukherjee, A., & Phadikar, S. (2020). A deep learning approach for the classification of rice leaf diseases. In *Intelligence Enabled Research: DoSIER 2019* (pp. 61-69). Singapore: Springer Singapore.

[17] Bhatia, S., Jaffery, Z. A., & Mehfuz, S. (2023, January). A comparative study of wireless communication protocols for use in smart farming framework development. In *2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT)* (pp. 1-7). IEEE.

[18] Bhatia, S., Jaffery, Z. A., & Mehfuz, S. (2023, May). Development and analysis of iot based smart agriculture system for heterogenous nodes. In *2023 International Conference on Recent Advances in Electrical, Electronics & Digital Healthcare Technologies (REEDCON)* (pp. 62-67). IEEE.

[19] Bhatia, S., Jaffery, Z. A., Mehfuz, S., & Goel, N. (2023). Integration of WSN and IoT: Wireless networks architecture and protocols–A way to smart agriculture. In *Handbook of Research on Machine Learning-Enabled IoT for Smart Applications Across Industries* (pp. 435-455). IGI Global.

[20] V. V. R. M. Rao, K. M., S. N., V. S. S. P. R. Gottumukkala, N. R. M., and N. Pamarthi, "An Innovative Machine Learning based Heart Disease Assessment System by Sequential Feature Selection Approach," in 2023 3rd International Conference on Intelligent Technologies (CONIT), 2023, doi: 10.1109/CONIT59222.2023.10205817.

[21] O. Ozaltın, "Date Fruit Classification by Using Image Features Based on Machine Learning Algorithms," Research in Agricultural Sciences, vol. 55, no. 1, pp. 26-35, 2024, doi: 10.5152/AUAF.2024.23171.

[22] N. Silpa, et al., "An Enriched Employee Retention Analysis System with a Combination Strategy of Feature Selection and Machine Learning Techniques," International Conference on Intelligent Computing and Control Systems (ICICCS), 2023, doi: 10.1109/ICICCS56967.2023.10142473.

[23] Kittani, T., & Albrifkani, A. M. A. (2024). Deep Learning Classification Algorithms Applications: A Review. *The Indonesian Journal of Computer Science*, *13*(3).

[24] G. Mahesh, et al., "Preeminent Sign Language System by Employing Mining Techniques," in IoT Based Control Networks and Intelligent Systems, P. P. Joby, M. S. Alencar, and P. Falkowski-Gilski, Eds., Lecture Notes in Networks and Systems, vol. 789, Springer, Singapore, 2024, pp. 527-536, doi: 10.1007/978-981-99-6586-1_39.

[25] Umrao, R., Srivastava, R., Shama, S., Kumar, D., & Bhatia, S. (2022). Precision farming system using IoT. In *Futuristic Sustainable Energy & Technology* (pp. 477-485). CRC Press.

[26] Bhatia, S., Gautam, D., Tomar, R., Bhandari, M., & Chaubey, S. (2022). IoT

Enabled Smart Irrigation Techniques With Leaf Disease Detection. *Proceedings of the Advancement in Electronics & Communication Engineering*.

[27] N. Silpa and V. V. R. Maheswara Rao, "Enriched Big Data Pre-Processing Model with Machine Learning Approach to Investigate Web User Usage Behaviour," Indian Journal of Computer Science and Engineering (IJCSE), vol. 12, no. 5, 2021, doi: 10.21817/indjcse/2021/v12i5/211205050.

[28] M. Faisal, M. Alsulaiman, M. Arafah, and M. A. Mekhtiche, "IHDS: Intelligent harvesting decision system for date fruit based on maturity stage using deep learning and computer vision," IEEE Access, vol. 8, pp. 167985–167997, 2020, doi: 10.1109/ACCESS.2020.3023234.

[29] N. Silpa and V. V. R. Maheswara Rao, "A Complete Research on Techniques & Technologies of Big Web Data Preparation to Web User Usage Behavior," International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 2S11, pp. 307-314, Sept. 2019, doi: 10.35940/ijrte.B1269.0982S1119.

[30] Kulkarni, P., & Shastri, S. (2024). Rice leaf diseases detection using machine learning. *Journal of Scientific Research and Technology*, 17-22.

[31] Naresh Kumar, B., & Sakthivel, S. (2025). Rice leaf disease classification using a fusion vision approach. *Scientific Reports*, *15*(1), 8692.

[32] H. Altaheri, M. Alsulaiman, and G. Muhammad, "Date Fruit Classification for Robotic Harvesting in a Natural Environment Using Deep Learning," IEEE Access, vol. 7, pp. 117115–117133, 2019, doi: 10.1109/ACCESS.2019.2936536.

[33] G. Mahesh, et al., "Primal-dual parallel algorithm for optimal content delivery in cloud CDNs," in IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Tamil Nadu, India, 2017, doi: 10.1109/ICCIC.2017.8524392.

[34] https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases

[35] Bhatia, S., Jaffery, Z. A., & Mehfuz, S. (2025). GSO-BiLSTM: performance assessment of IoT enable smart irrigation system using intelligent algorithms. *International Journal of Machine Learning and Cybernetics*, 1-19.

[36] S. N., V. V. R. M. Rao, V. Subbarao, M. Pradeep, C. R. Grandhi, and A. Karunasri, "A Robust Team Building Recommendation System by Leveraging Personality Traits Through MBTI and Deep Learning Frameworks," in International Conference on IoT, Communication and Automation Technology (ICICAT), 2023, pp. 1-6, doi: 10.1109/ICICAT57735.2023.10263718.

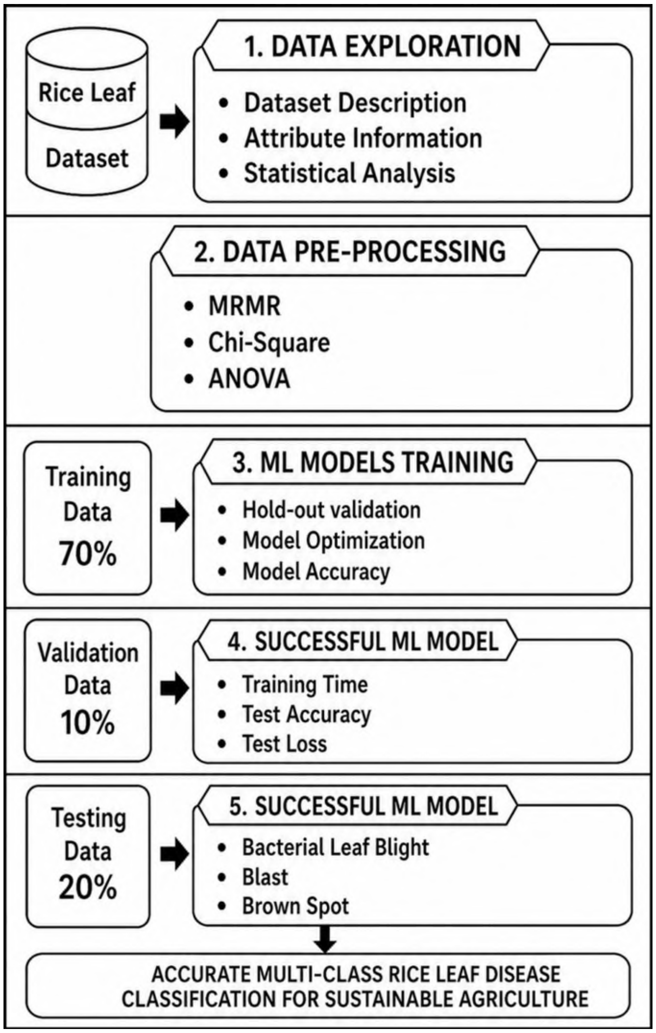
[37] Padmanabhuni, S., & Gera, P. (2025). Bayesian optimized deep learning and ensemble classification approach for multiclass plant disease identification. *Discover Sustainability*, *6*(1), 709.

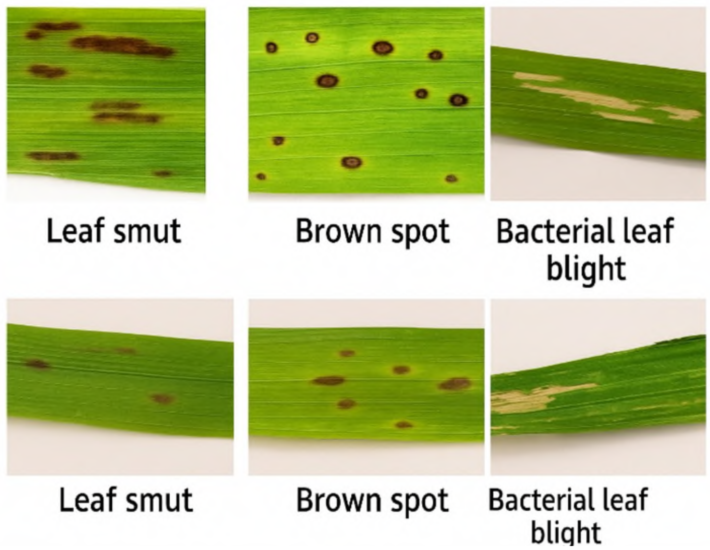
[38] Mukherjee, R., Ghosh, A., Chakraborty, C., De, J. N., & Mishra, D. P. (2025). Rice leaf disease identification and classification using machine learning techniques: A comprehensive review. *Engineering Applications of Artificial Intelligence*, *139*, 109639.

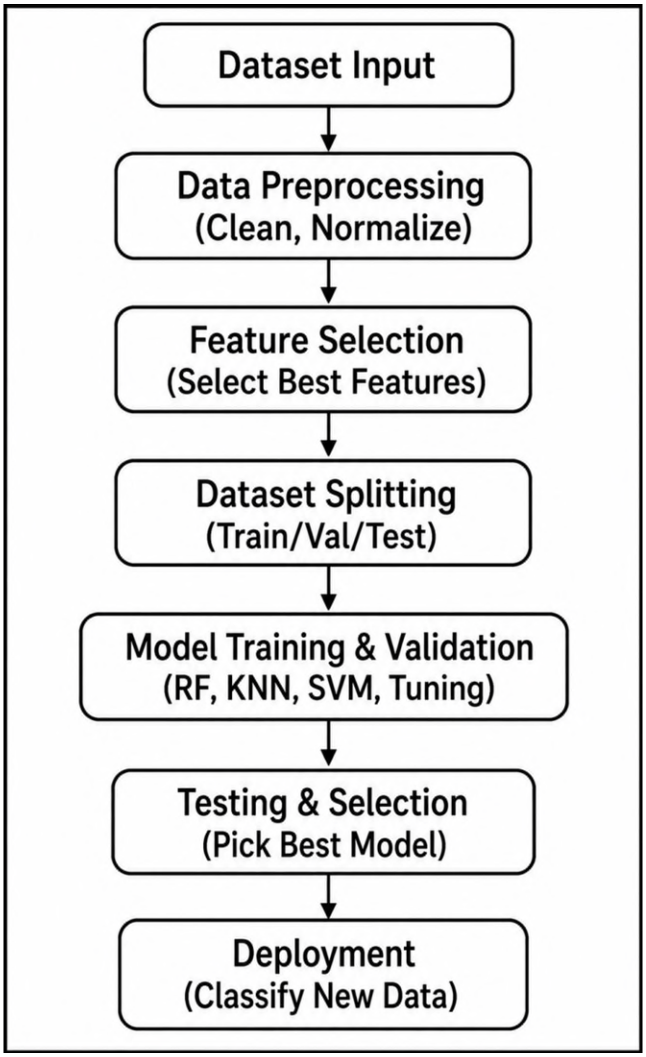
[39] Kazi, S. S., Palkar, B., & Mishra, D. (2025). VCNet: Optimized Deep Learning framework with deep feature extraction and genetic algorithm for multiclass rice crop disease detection. *MethodsX*, 103551.

[40] Rana, Muhammad Ehsan, Vazeerudeen Abdul Hameed, Ian Kiew Yi Eng, Hrudaya Kumar Tripathy, and Saurav Mallik. "Harnessing artificial intelligence for sustainable rice leaf disease classification." *Frontiers in Plant Science* 16 (2025): 1594329.

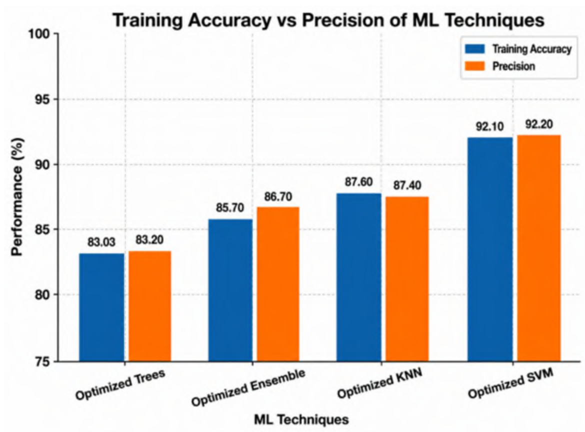
[41] Li, Y., Chen, X., Yin, L., & Hu, Y. (2024). Deep learning-based methods for multi-class rice disease detection using plant images. Agronomy, 14(9), 1879.
[42] Shanmugam, V., Madhusudhana Rao, T. V., Rao, H. J., & Maram, B. (2023). Internet of things based smart application for rice leaf disease classification using optimization integrated deep maxout network. *Concurrency and Computation: Practice and Experience*, *35*(6), 1-1.
[43] https://www.kaggle.com/datasets/minhhuy2810/rice-diseases-image-dataset
[44] https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases

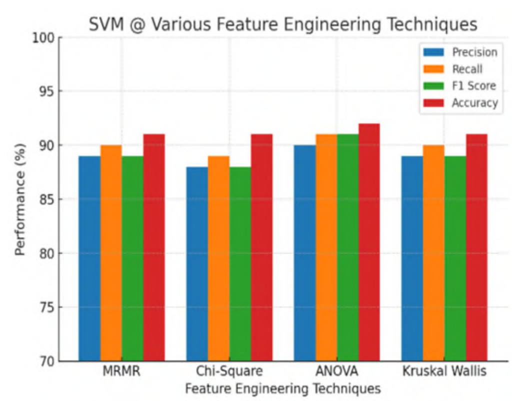

Rice Leaf
Dataset
1. DATA EXPLORATION
• Dataset Description
• Attribute Information
• Statistical Analysis
2. DATA PRE-PROCESSING
• MRMR
• Chi-Square
• ANOVA
Training Data 70%
3. ML MODELS TRAINING
• Hold-out validation
• Model Optimization
• Model Accuracy
Validation Data 10%
4. SUCCESSFUL ML MODEL
• Training Time
• Test Accuracy
• Test Loss
Testing Data 20%
5. SUCCESSFUL ML MODEL
• Bacterial Leaf Blight
• Blast
• Brown Spot
ACCURATE MULTI-CLASS RICE LEAF DISEASE CLASSIFICATION FOR SUSTAINABLE AGRICULTURE

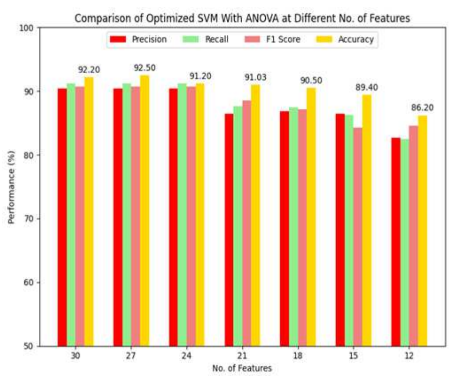

Leaf smut
Brown spot
Bacterial leaf blight
Leaf smut
Brown spot
Bacterial leaf blight



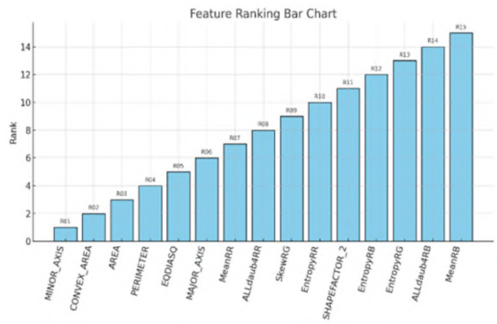
Dataset Input
Data Preprocessing
(Clean, Normalize)
Feature Selection
(Select Best Features)
Dataset Splitting
(Train/Val/Test)
Model Training & Validation
(RF, KNN, SVM, Tuning)
Testing & Selection
(Pick Best Model)
Deployment
(Classify New Data)

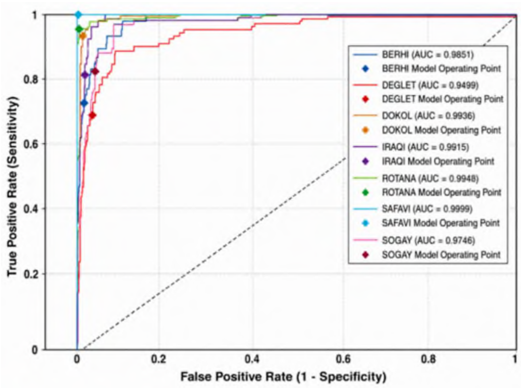

Training Accuracy vs Precision of ML Techniques
Training Accuracy
Precision
Performance (%)
100
95
90
85
80
75
83.03
83.20
85.70
86.70
87.60
87.40
92.10
92.20
Optimized Trees
Optimized Ensemble
Optimized KNN
Optimized SVM
ML Techniques

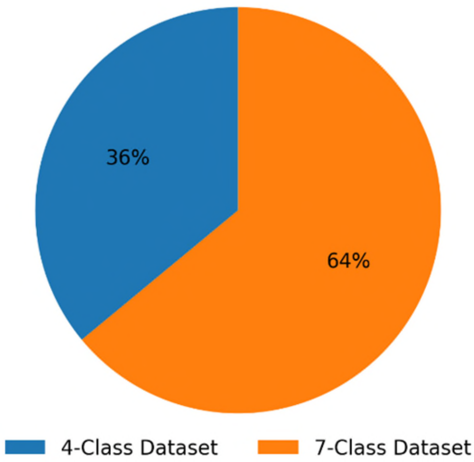

SVM @ Various Feature Engineering Techniques
Precision
Recall
F1 Score
Accuracy
Performance (%)
100
95
90
85
80
75
70
MRMR
Chi-Square
ANOVA
Kruskal Wallis
Feature Engineering Techniques

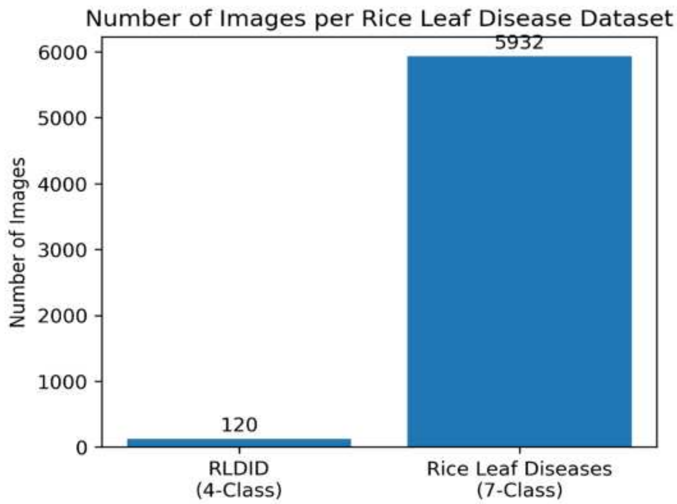


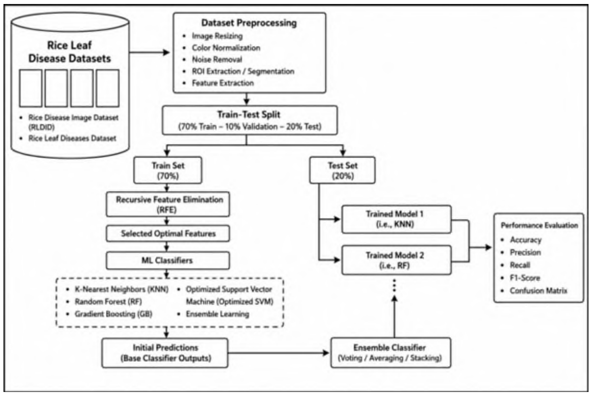
Comparison of Optimized SVM With ANOVA at Different No. of Features
Precision
Recall
F1 Score
Accuracy
92.20
92.50
91.20
91.03
90.50
89.40
86.20
Performance (%)
100
90
80
70
60
50
30
27
24
21
18
15
12
No. of Features


Feature Ranking Bar Chart
Rank
0
2
4
6
8
10
12
14
R01
R02
R03
R04
R05
R06
R07
R08
R09
R10
R11
R12
R13
R14
R15
MINOR_AXIS
CONVEX_AREA
AREA
PERIMETER
EQDIASQ
MAJOR_AXIS
MeanRR
ALLdaub4RR
SkewRG
EntropyRR
SHAPEFACTOR_2
EntropyRB
EntropyRG
ALLdaub4RB
MeanRB

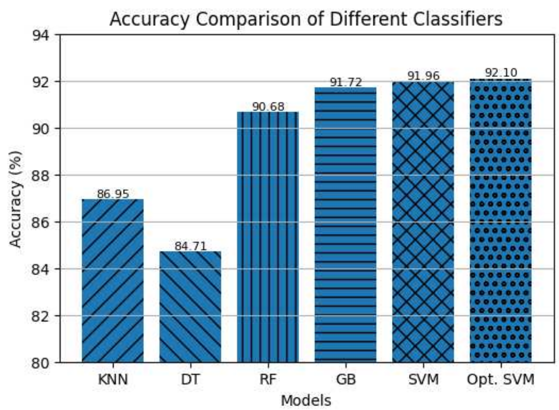

True Positive Rate (Sensitivity)
False Positive Rate (1 - Specificity)
0
0.2
0.4
0.6
0.8
1
BERHI (AUC = 0.9851)
BERHI Model Operating Point
DEGLET (AUC = 0.9499)
DEGLET Model Operating Point
DOKOL (AUC = 0.9936)
DOKOL Model Operating Point
IRAQI (AUC = 0.9915)
IRAQI Model Operating Point
ROTANA (AUC = 0.9948)
ROTANA Model Operating Point
SAFAVI (AUC = 0.9999)
SAFAVI Model Operating Point
SOGAY (AUC = 0.9746)
SOGAY Model Operating Point

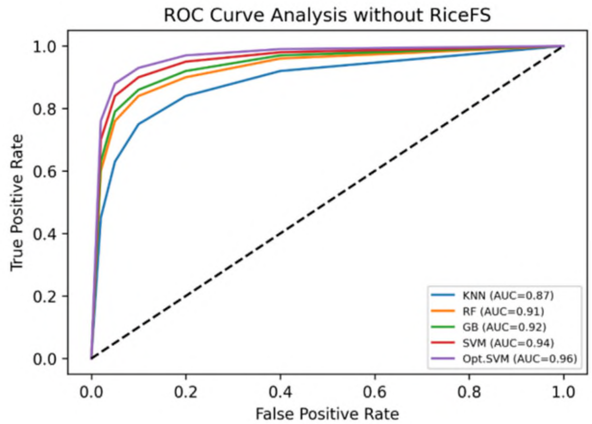

36%
64%
4-Class Dataset
7-Class Dataset


Number of Images per Rice Leaf Disease Dataset
Number of Images
6000
5000
4000
3000
2000
1000
0
5932
120
RLDID
(4-Class)
Rice Leaf Diseases
(7-Class)

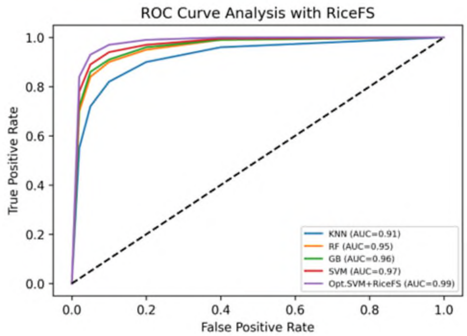


Rice Leaf Disease Datasets
• Rice Disease Image Dataset (RLDID)
• Rice Leaf Diseases Dataset
Dataset Preprocessing
• Image Resizing
• Color Normalization
• Noise Removal
• ROI Extraction / Segmentation
• Feature Extraction
Train-Test Split
(70% Train – 10% Validation – 20% Test)
Train Set
(70%)
Test Set
(20%)
Recursive Feature Elimination
(RFE)
Selected Optimal Features
ML Classifiers
• K-Nearest Neighbors (KNN)
• Random Forest (RF)
• Gradient Boosting (GB)
• Optimized Support Vector Machine (Optimized SVM)
• Ensemble Learning
Trained Model 1
(i.e., KNN)
Trained Model 2
(i.e., RF)
Performance Evaluation
• Accuracy
• Precision
• Recall
• F1-Score
• Confusion Matrix
Initial Predictions
(Base Classifier Outputs)
Ensemble Classifier
(Voting / Averaging / Stacking)


Accuracy Comparison of Different Classifiers
Accuracy (%)
94
92
90
88
86
84
82
80
86.95
84.71
90.68
91.72
91.96
92.10
KNN
DT
RF
GB
SVM
Opt. SVM
Models



ROC Curve Analysis without RiceFS
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
KNN (AUC=0.87)
RF (AUC=0.91)
GB (AUC=0.92)
SVM (AUC=0.94)
Opt.SVM (AUC=0.96)



ROC Curve Analysis with RiceFS
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
KNN (AUC=0.91)
RF (AUC=0.95)
GB (AUC=0.96)
SVM (AUC=0.97)
Opt.SVM+RiceFS (AUC=0.99)