International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V73I3P134

Volume 73 Issue 3, 487-502, March 2025
© 2025 Seventh Sense Research Group®

Original Article

NSGRF: Plant Leaf Disease Detection Using Multi Objective Genetic Algorithm Features Optimization with Random Forest

Harminder Kaur[1], Neeraj Raheja[2]

[1,2]Department of Computer Science and Engineering, Maharishi Markandeshwar Engineering College, Mullana, Ambala, Haryana, India.

[1]Corresponding Author : harminderkaur39@gmail.com

Received: 05 September 2024 Revised: 10 January 2025 Accepted: 25 January 2025 Published: 28 March 2025

Abstract - *The agricultural sector in India, a developing nation, relies heavily on agriculture for its population's livelihood and food security. Plants are essential for sustenance, medicine, and various industries. India's agricultural sector contributes 17% to the GDP and ranks among the top three nations in producing staple crops like rice and wheat and cash crops like cotton and vegetables. This study is focused on the production and subsequent evaluation of an advanced plant disease classification system using the Plant Village dataset. The proposed approach integrates texture feature extraction, multi-objective optimization, and Random Forest Classifier to enhance classification accuracy and efficiency. Texture features are extracted from the Gray-Level Co-Occurrence Matrix (GLCM) and using Local Binary Patterns (LBP), providing a low-dimensional yet interpretable representation of leaf patterns. This feature extraction method reduces overfitting, especially in scenarios with limited sample sizes. The Random Forest parameters are optimized using multi-objective evolutionary algorithms like NSGA-II, ensuring a balance between model complexity and generalization capability. A genetic algorithm is employed for feature selection and weighting, which, combined with Random Forest techniques, further refines the classification performance. Extensive testing demonstrates the suggested method's effectiveness, significantly outperforming previous methods. The NSGRF approach, a key element of the system, significantly outperforms in terms of precision, accuracy, recall, and F-score. The study's findings indicate that this integrated approach is more accurate and offers faster inference times and improved interpretability; in this way, it can be a significant device for automatically detecting plant diseases in the agricultural technology field. In the experiment survey, the algorithm NSGRF improves 2% accuracy, 1 % precision, 2 % recall, and 1% f-score compared to other approaches.*

Keywords - *Plant disease, Feature extraction, Multi-objective optimization, Machine Learning, Genetic Algorithm.*

1. Introduction

It has already been determined that there are over 2 million types of living animals on the globe, including humans, fungi, and plants, with plants playing an important role in human survival. Plants are a wonderful asset to human health that can be found practically anywhere. The diversity of plants provides significant understanding for the advancement of human existence and is thus recognized as an important component of human health. Plants are very valuable since they are the foundation of the food production network and the source of many medications [1]. Plants have always been concerned with environmental preservation. Following several major developments in plant biology, there appears to be a vast number of plant species that have been discovered, named, or used. Unidentified plant species are unsolved riddles that have been widely reported [2]. Ethnologists are now combining places worldwide in search of possible medications and agricultural products. They explore plant species' key characteristics and relationships across ecosystems to comprehend the need for more flexibility in managing plant resources. Researchers in the twenty-first century were particularly interested in how genomic variety, but mostly environmental resistance, is required to handle problems such as mass feeding and illness prevention [3]. The disease that takes over the plant has been significantly impacted the plant's progress and production. As a result, discovering new forms and techniques to combat plant infections is vital to maintaining an appropriate food supply for all living organisms globally [4]. Plant disease detection and severity evaluation appear to be tough tasks. Human eye assessment (sensory assessment) appears to have been the primary approach for identifying disease severity until recently [7]. Such personnel convey frequent agricultural field monitoring for a disease specialist to evaluate disease appropriately. Various techniques are currently being developed to lessen the waste present in underperforming facilities. Farmers commonly utilize manual techniques, such as visual inspection conducted by human observers, including the assessment conducted by the human eye [8]. Accurately detecting and categorizing plant diseases require experts' specialized knowledge and skills, which can result in significant costs and time investment. Due to their time-consuming nature, these systems are susceptible to scanning and subjective faults by diverse farmers [9]. Visual processing is time-consuming, costly, and difficult for large areas of plants because it requires continuous hand examination. The ever-increasing population quickly impacts food demand and availability [10]. The current situation calls for a thorough assessment of the application of contemporary technology to enable swift and appropriate detection of illnesses, coupled with the timely implementation of corrective actions [11]. Machine-learning-based analysis is among the most effective and cost-effective approaches for identifying plant disease criteria [12].

1.1. Disease in Plants

Plant diseases are a major problem in the agricultural, horticultural, and forestry fields because they may harm plants, reduce crop yields, and disrupt ecosystems overall [13]. In addition to infectious agents, including bacteria, fungi, nematodes, viruses, and parasitic plants, non-infectious factors like environmental stressors may also play a role in causing these disorders [14]. For those working in agriculture, horticulture, and scientific research, understanding plant diseases is crucial for developing effective strategies to prevent and manage these diseases, which in turn protects plants from harm [15]. In the context of plants, any condition that prevents them from performing at their best is considered a disease [16]. When environmental factors modify a plant's development, structure, functions, or other characteristics, rendering it susceptible to infection, we say that the plant is physiologically sensitive [17, 18]. Plant diseases have a profound effect on the intricate physiological systems of plants, leading to a substantial decrease in agricultural productivity. Visuals of potato leaves affected by various diseases are shown in Figure 1.

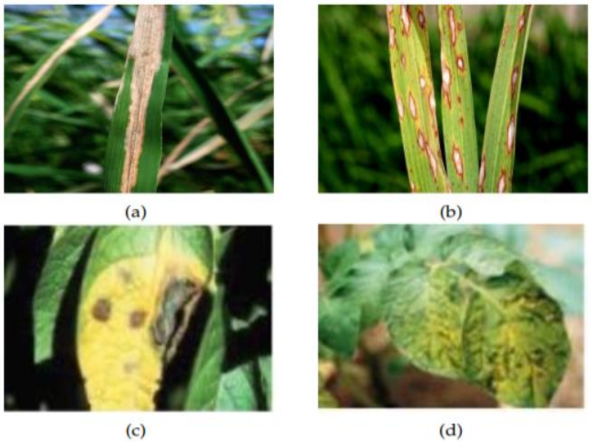


Fig. 1 The sample image of the leaf affected by (a) Leaf blight, (b) Leaf blast, (c) Early blight, and (d) Potato virus Y.

1.2. Plant Disease Detection System

The fundamental processes involved in detection techniques are shown in Figure 2.

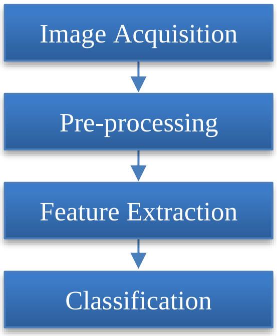


Fig. 2 Plant/disease detection system

- Image Acquisition/collecting data: It involves the gathering of leaf photographs. The procurement of pictures happens in a computerized design through the utilization of diverse equipment, including scanners and digital cameras. If the image obtained is not in a digital format, it is necessary to apply the process of converting the analogue image to a digital form.
- Pre-processing: It is crucial for detecting and identifying plant diseases, involving techniques like image scaling, noise elimination, color transformation, morphological procedures, and disease region segmentation. Methods like Weiner, Median, and Gaussian filters eliminate noise, while color spaces and segmentation techniques identify and isolate the diseased region.
- Feature extraction: It is crucial in classification problems, as it helps describe images meaningfully and provides maximum discrimination among different classes. Ideally, feature values should be similar for the same class and different for different classes.
- Classification: Classification refers to the systematic procedure of organizing a collection of photographs into distinct categories or groupings. This methodology can be utilized for unstructured as well as structured information. The primary stage in the method includes grouping the given data of interest. On the other hand, the classes are referred to as the objective, mark, or classification.

1.3. Disease Management in Plants

Plant disease management typically involves a combination of chemical treatments, biological controls and cultural practices.

- Effective Management Techniques: These Techniques involve precise disease identification, understanding its life cycle, and prevention methods. Proper plant maintenance and vigilant observation are crucial for resilient gardens and harvests.
- Biological Controls: These techniques involve the introduction of natural predators or beneficial microorganisms to combat pathogens.

- Crop Rotation: It involves altering the selection of plant species cultivated in a particular region to disrupt the life cycle of diseases.
- Resistant Cultivars: These refer to the cultivation of naturally resistant varieties to specific diseases.
- Fungicides, Bactericides, and Pesticides: These are chemical interventions employed for the purpose of managing populations of pathogens.
- Pruning and Sanitation: These involve the removal of sick plant portions and implementing clean gardening practices.
- Plant Nutrition and Irrigation: A crucial aspect of plant care involves the provision of enough irrigation and nutrition, which enhances the plant's ability to withstand various stressors.

1.4. Contribution of Research

This work promotes substantial contributions in the area of plant disease detection with machine learning and optimization techniques. The main contributions are outlined as follows:

1. Integration of Texture Feature Extraction with Optimization Techniques: In this study, we introduce a new combination of texture feature extraction: Local Binary Patterns (LBP) and GrayLevel Co-Occurrence Matrix (GLCM). By doing so, we create an efficient, interpretable representation of leaf patterns, minimizing the risks of overfitting, especially for datasets with limited samples.
2. Multi-Objective Evolutionary Optimization for Enhanced Model Efficiency: Multi-Objective Genetic Algorithms (MOGA) are used for optimizing Random Forest parameters. This approach is then evaluated in terms of classification performance and is shown to strike a balance between model complexity and accuracy, improving performance dramatically.
3. Robust Ensemble Classification Framework: The study improves reliability and robustness when detecting disease by harnessing an ensemble classification framework comprising Random Forest, Support Vector Machines (SVM) and Logistic Regression.
4. Superior Performance Metrics: An algorithm, Non-Dominated Sorting Genetic Random Forest (NSGRF), is proposed, which is shown to outperform other state-of-the-art methods across all key metrics (accuracy, precision, recall, and F1) over the Plant Village dataset.

1.5. Novelty of the Approach

The novelty of this examination lies in its remarkable strategic combination and its significant contribution to the field:

1. Innovative Integration of Techniques: Incorporating LBP, GLCM and multi-objective optimization into an RF classifier constitutes a technique that is ground breaking among the existing literature that deals with plant disease detection.
2. Optimization for Practical Use: Using NSGA-II achieves better accuracy but reduces computational complexity; therefore, the model applies to real-world agricultural applications.
3. Comprehensive and Efficient Model: his, together with genetic algorithms, results in a scalable, robust and interpretable solution that overcomes some of the practical challenges associated with limited sample sizes and overfitting.

2. Related Work

Ahmed and Yadav [1] delve into the complicated realm of plant disease detection models, employing sophisticated machine learning techniques such as random forest, nearest neighbours, linear regression, naive bayes, neural networks, and support vector machines. The models undergo evaluation through the utilization of objective metrics like recall, precision, F1-score, true negative rate, and true positive rate. The findings highlight the effectiveness of the ensemble plants disease model in accurately detecting diseases in their early stages, surpassing other models in performance. This, in turn, facilitates the implementation of timely preventative and predictive care strategies.

Varshney et al. [2] clarify a profound DL approach for identifying leaf plant diseases, employing CNNs in conjunction with SVMs. The model was evaluated using PlantVillage, a benchmark dataset, and showed superior performance compared to previous research, with a training accuracy of 88.77%, outperforming previous methods.

Sharma et al. [3] explore transfer learning in ML and DL for rice disease detection. It compares three diseases: brown spot, bacterial blight, and rice blast. Results show transfer learning has superior performance compared to standard techniques. InceptionResNetV2 outperforms XceptionNet, with InceptionResNetV2 ranking second. The study could aid farmers in timely disease detection but recommends larger datasets for generalizability.

Hassan et al. [4] aim to improve plant disease identification research by transitioning from handcrafted-features-based models to DL-based models. DL methods offer high accuracy, but performance may decline under field image conditions or on different datasets. Inception layer-based DL models like GoogleNet and InceptionV3 show superior feature extraction capabilities. The study also explores hurdles for proper disease identification.

Alatawi et al. [5] examine the use of ML in disease detection, specifically utilizing a CNN VGG-16 model. Its objective is to assist farmers in addressing disease outbreaks by enabling timely therapeutic interventions. The model was trained on a vast collection of 15,915 images of plant leaves taken from the Plant Village database. The model established an impressive accuracy rate of 95.2%, along with the minimum testing loss. This indicates a hopeful direction for applying DL-based methods in identifying plant diseases.

Sujatha et al. [6] compare ML algorithms (SVM, Random Forest, SGD) and DL models (Inception-v3, VGG-16, and VGG-19) for detecting diseases in citrus plants. Results show that deep learning methods outperform ML methods in disease identification. The top classification accuracy is 89.5% for VGG-16, followed by 89% for Inception-v3, 87.4% for VGG-19, 87% for SVM, 86.5% for SGD, and 76.8% for RF. RF achieves the lowest categorization accuracy.

Vishnoi et al. [7] use soft computing and computer vision to automate the detection of plant illnesses through leaf photographs. This study explores the pros and cons of different detection systems, examines prevalent illnesses, and examines contemporary feature extraction approaches. It aims to improve researchers' understanding of the practicality of computer vision techniques in detecting and categorizing plant diseases.

Roy et al. [8] explore the use of Dimensionality Reduction, specifically focusing on feature extraction, in predicting plant diseases caused by biotic stress. It examines various feature extraction strategies for both quantitative and picture data, aiming to improve prediction effectiveness and accuracy. The study will use publicly accessible datasets to understand the functionality of dimensionality reduction methods in generating reliable predictions and aiding in crop disease management.

Applalanaidu et al. [9] reviewed over 45 peer-reviewed papers from 2017 to 2020 on the identification, recognition, and classification of plant diseases by means of ML and DL algorithms. It focuses on state-of-the-art ML algorithms like SVM, NN, KNN, and Naïve Bayes, as well as prominent DL algorithms like GoogLeNet, VGGNet, and AlexNet. The study uses standardized experimental setup metrics and image segmentation techniques to analyze the effectiveness of each algorithm. The findings are expected to positively impact agricultural output.

Swain et al. [19] explore various models used for identifying diseases in agricultural products and review various classification models for plant leaf diseases. It also discusses the framework of segmentation, feature extraction, and the use of various classifier algorithms, highlighting their potential in classifying and detecting diseases in agricultural products.

Patidar et al. [20] present a framework for identifying and categorizing illnesses in rice plants, a staple food in India. The approach uses photographs of infected rice plants to detect and classify diseases like Leaf smut, Bacterial leaf blight, and Brown spots. The Rice Leaf Disease Dataset was used, and researchers used a Residual Neural Network for classification. This fast and efficient technique prevents the model from reaching a saturation point. The study achieved a 95.83% accuracy rate on the dataset.

Panchal [23] confers using K-means clustering and HSV-dependent classification for image-based segmentation, specifically for identifying infected leaf areas. It also discusses GLCM for feature extraction, achieving an accuracy rate of 98% when processed through the RF classifier.

3. Proposed Methodology

The proposed approach comprises six major phases: Image Acquisition, Pre-Processing, Feature Extraction, Multi-Objective Optimization, and Ensemble Classification. The basic architecture of the proposed method is given below.

- Image Acquisition: The plant village database, which is readily accessible and has over 50,000 photos of 14 distinct crop species, serves as where the dataset was collected. Figures 3 to 5 displays three sample photos of potato leaves.
  - ➢ Non-diseased or healthy potato leaf
  - ➢ Late blight-affected potato leaf
  - ➢ Early blight-affected potato leaf

- Image Preprocessing: Image preprocessing consists of three steps: Resizing, Normalization and Augmentation. In Normalization and Resizing, all the input images have been resized to a standard size (e.g., 256x256 pixels) and normalize the values of pixels in images (like a scale ranging with a value between 0 and 1). To improve the resilience of the dataset following normalization, augmentation approaches like scaling, rotation, and flipping have been employed.

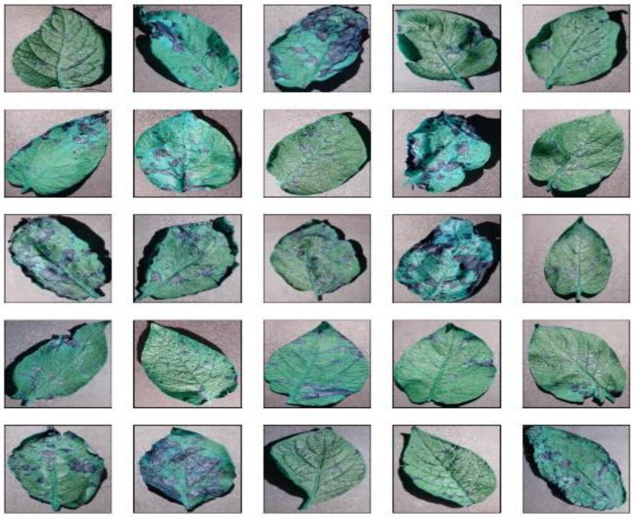


Fig. 3 The image of the potato early blight affected leaf image

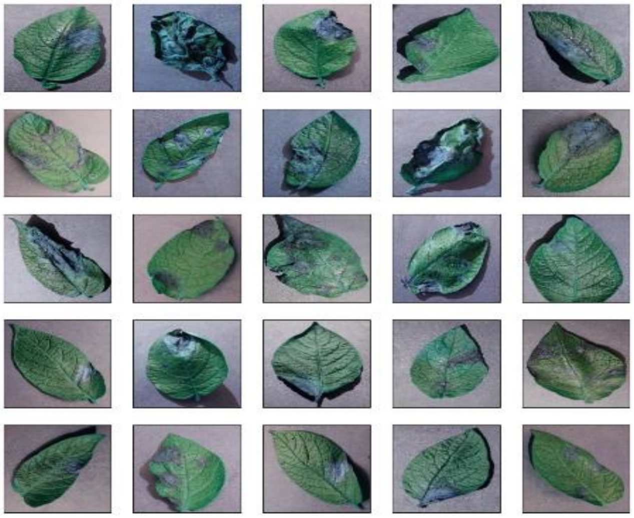


Fig. 4 The image of the potato late bright affected leaf image

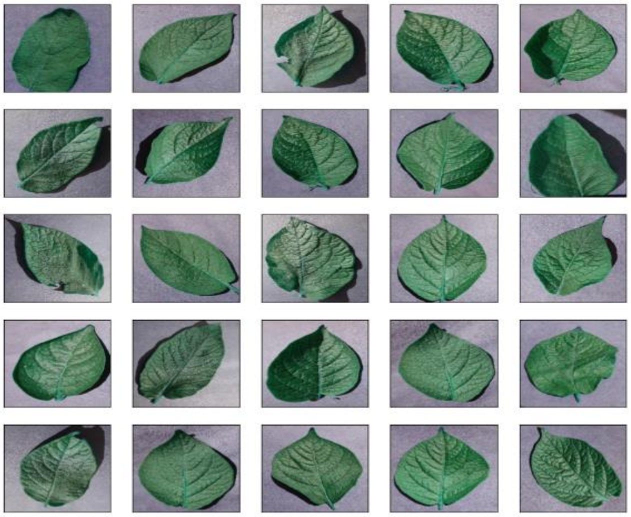


Fig. 5 The image of the potato healthy leaf image

The segmentation method has been used to segregate the plant image into several segments. This technique can be applied to eliminate unhealthy areas from the backdrop of a plant's stem, leaf, or root.

- Feature extraction: It is crucial in classification problems, as it helps describe images meaningfully and provides maximum discrimination among different classes. Ideally, feature values should be similar for the same class and different for different classes. Feature extraction is accomplished through a careful process of analysis and selection. Derive a comprehensive range of features from each image, encompassing color, shape, texture, and morphological aspects. Standardize the features to guarantee they are on a comparable scale.
- Optimization: Multi-objective optimization is based on the principle that when several conflicting objectives are required to be accomplished in a given situation, the optimization of a single objective frequently results in the degradation of the outcome of other objectives, so to get the best optimization for each sub-objective, a collection of solutions that balance several sub-objectives must be found. Most of the optimization problems had multiple objectives. These objectives often conflict with each other, meaning that optimizing one aim can lead to the other objectives being compromised. The Genetic Algorithm (GA) describes an evolutionary algorithm that efficiently discovers the best possible solutions without relying on assumptions about the search space. GA may be utilized to capture several solutions simultaneously in multi-objective optimization situations since it operates with a population of alternative solutions. A wide variety of GA-based multi-objective improvement strategies have been utilized to track down an example of Pareto-ideal arrangements over the course of the last 10 years and beyond.

The aims of the current study for optimization are outlined below:

- Minimization of the Entropy
- Maximization of the Information Gain

3.1. Multi-Objective Genetic Algorithm (MOGA)

MOGA represents an optimization procedure used to solve problems that involve multiple conflicting objectives. It's an expansion of the customary hereditary calculation, a conventional genetic algorithm propelled by the course of natural evolution.

In many real-world problems, there are multiple criteria or objectives that need to be simultaneously optimized, and these objectives often conflict with each other. For example, in engineering, you might want to simultaneously minimize the cost of a design while maximizing its performance. In such cases, a single optimal solution may not exist, as improving one objective might lead to another deterioration.

A multi-objective genetic algorithm addresses this challenge by developing a populace of likely solutions, called individuals, over successive generations. The individuals are symbolized as strings of genetic data, often in binary form. The algorithm employs genetic operators such as selection, mutation, and crossover to produce new individuals in every generation.

A multi-objective genetic algorithm differs primarily because it persists over time with a collection of solutions called the Pareto front or set. Solutions in this set do not dominate any other solution with respect to all objectives. In simple terms, a solution in the Pareto front is superior or comparable to different solutions in at least a single goal without being inferior in any other objective.

Here is How a Multi-Objective Genetic Algorithm Works**:**

- Initialization: Generate an initial population of individuals with random or heuristic values.
- Evaluation: Evaluate each individual's performance with respect to all the defined objectives.
- Selection: Select individuals from the present population using a selection process that prioritizes solutions more closely aligned with the Pareto front.
- Crossover and Mutation: Apply genetic operators like crossover (combination of genetic material from two parents) and mutation (random modification of genes) to create new offspring.
- Evaluation of Offspring: Evaluate the performance of the newly created offspring with respect to the objectives.
- Replacement: Replace certain individuals from the existing population with the recently generated offspring to preserve diversity.
- Pareto Front Maintenance: Update the Pareto front by adding solutions not dominated by the current population.
- Termination: The algorithm closes when a particular halting condition is fulfilled, like arriving at the most extreme number of generations, accomplishing an ideal degree of convergence, or meeting different rules characterized by the client/user.

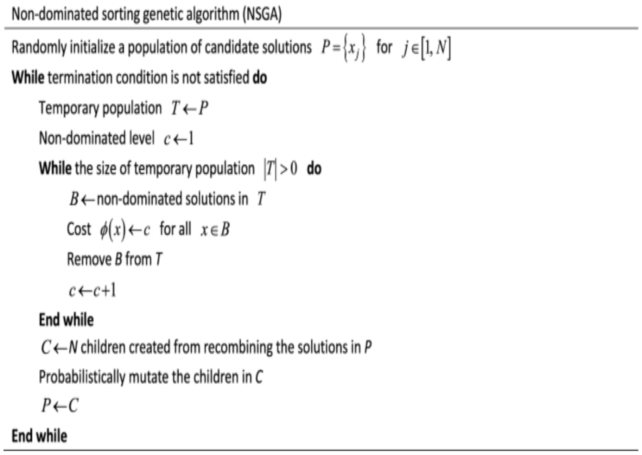
Algorithm: 1

function genetic Algorithm for Feature Selection (features, labels):
# Initialize population with random feature subsets and weights 	
population = initialize Population (size=population Size, features=features)
for generation in range(maxGenerations):
# Evaluate the fitness of each individual in the population 	
for individuals in population:
individual. Fitness = estimate Fitness (individual, features, labels)
# Select the fittest individuals for reproduction 	
parents = select Parents (population, fitness Function)
# Generate the next generation through crossover and mutation 	
next generation = crossover and mutation (parents)
# Replace fewer fit individuals with new ones 	
population = select New Population (population, next generation)
# Return the best individual (feature set and weights) 	
Best Individual = select Best Individual(population)
return the best individual. Features, best individual. weights

Finding a group of solutions that are both varied and as near to the Pareto optimum front as feasible are the two main objectives of multi-objective optimization, as mentioned before. A Pareto-based evolutionary multi-objective optimization algorithm shares a comparable structure with a Genetic Algorithm (GA), except for how it assigns fitness values to various objectives. The strategy being investigated here is NSGA-II, a non-dominated arranging GA created by Deb in 2001. It represents an elitist strategy for two-objective algorithm optimization. The algorithm employs the elite-preserving operator, which ensures that a population's elites have a better chance of being passed down through the generations. The two offspring resulting from the application of mutation and crossover operators are thereafter evaluated against their respective parents to select the two most optimal solutions out of the four generated by the parents.

3.2. Non-Dominating Sorting Genetic Algorithm (NSGA-II)

NSGA-II ranks high when it comes to creating the Pareto frontier. The NSGA-II algorithm ranks the people based on dominance. NSGA-II keeps variety without the need to characterize any additional boundaries by utilization of phenotypic group examination administrator and elitism. Everything begins with a random initial generation. First i create a string by joining the parents and their children. In every generation, the objective functions of each of the strings are processed, and all subsequent arrangements are positioned on various non-dominated fronts.



```
Non-dominated sorting genetic algorithm (NSGA)
Randomly initialize a population of candidate solutions $P=\{x_j\}$ for $j\in[1,N]$
While termination condition is not satisfied do
    Temporary population $T\leftarrow P$
    Non-dominated level $c\leftarrow 1$
    While the size of temporary population $|T|>0$ do
        $B\leftarrow$ non-dominated solutions in $T$
        Cost $\phi(x)\leftarrow c$ for all $x\in B$
        Remove $B$ from $T$
        $c\leftarrow c+1$
    End while
    $C\leftarrow N$ children created from recombining the solutions in $P$
    Probabilistically mutate the children in $C$
    $P\leftarrow C$
End while
```

3.3. Ensemble Classification

The performance of ensemble learning techniques has consistently been shown to be better performance as compared to individuals or single classifiers. In the proposed approach, classifiers like RF, SVM, and Logistic regression have been used. Among all of them, RF is itself an ensemble learning method. The Random Forest is composed of an ensemble of decision trees and hence has shown great effectiveness throughout the years in multi-disciplinary fields.

Algorithm: 2

function ensembleLearningAlgorithm(features, labels, testFeatures):
  # Split dataset into training and validation sets
  trainingFeatures, trainingLabels, validationFeatures, validationLabels = splitDataset(features, labels)
  # Initialize ensemble
  ensemble = []

```
    # Train multiple models on the training set
    for model Type in model Types:
        model = train Model (modelType, trainingFeatures, trainingLabels)
        ensemble.append(model)
    # Combine predictions from all models in the ensemble
    predictions = []
    for testFeature in testFeatures:
        modelPredictions = [model.predict(testFeature) for model in ensemble]
      finalPrediction = combinePredictions(modelPredictions)
        predictions.append(finalPrediction)
 return predictions
   # Main program
  # Assume features and labels are already extracted and
availableoptimizedFeatures, featureWeights = geneticAlgorithmForFeatureSelection(features, labels)
modelPredictions = ensembleLearningAlgorithm (optimizedFeatures, labels, testFeatures)
 # Display or use the model predictions as required
```

3.4. Non-Dominated Sorting Genetic Random Forest Algorithm (NSGRF)

The algorithm works as the following steps:

Step 1 : GLCM or LBP will be supplied to extract texture features from the images in the Plant Village Database. If you want to extract features, use LIBRARIES such as scikit images or mahotas.

Step 2 : The texture feature data was preprocessed and then split into testing and training sets.

Step 3 : You're going to need a multi-objective optimization algorithm in order to optimize the Random Forest parameters

Step 4 : Train Random Forest After obtaining the optimized parameters from the multi-objective optimization, train the Random Forest on the texture features using the obtained parameters.

Step 5 : Evaluate the Classifier Evaluate the trained Random Forest classifier on the test set and calculate its performance metrics, like recall, precision, accuracy, and F1-score.

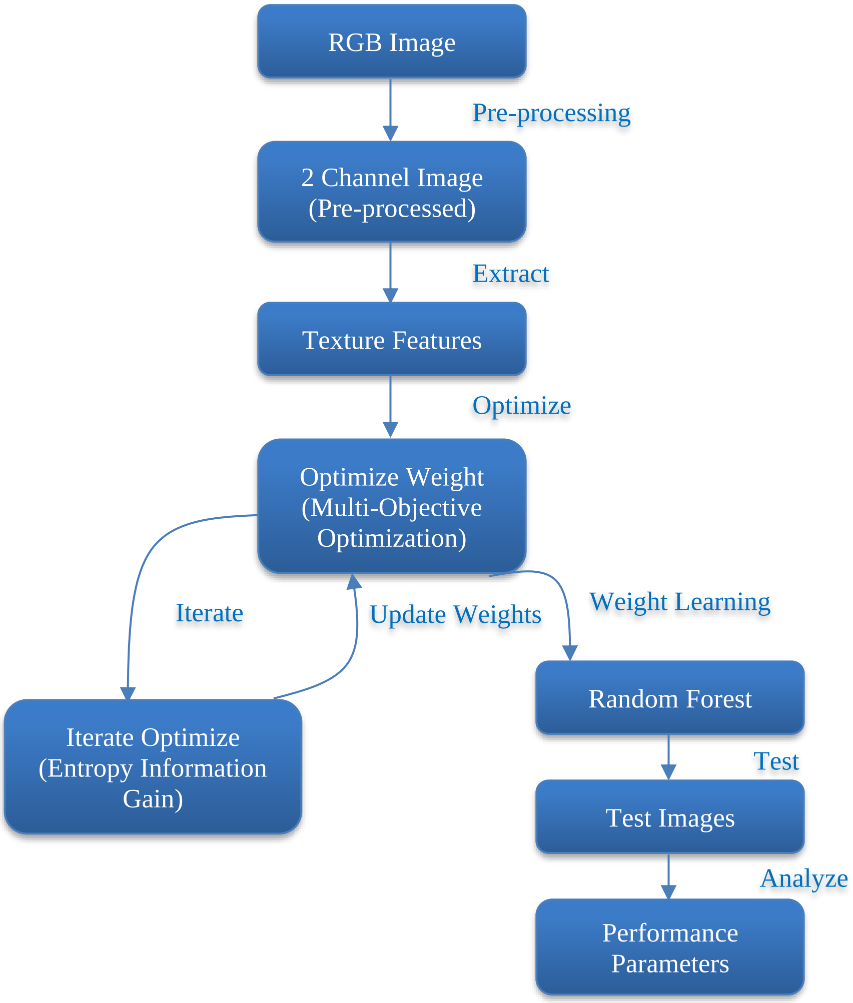


Fig. 6 Flow chart of NSGRF

Algorithm: 3 - Proposed Work

1. Input the RGB image, preprocess the images for noise removal, resizing and contrast enhancement and change it into 2 channel images after pre-processing.
2. Extract the texture features and optimize their weight by multi-objective optimization (GA) to minimize the entropy and maximize the information gain.
3. Step 2 is iterative till optimize or converged by two parameters: entropy and information gain.
4. After optimizing weight learning by Random Forest for classification and testing the images, then analyzing performance parameters.

4. Why it is Better than Others

Combining texture features extracted from the Plant Village dataset with a multi-objective optimization strategy and a Random Forest (RF) learner can offer several advantages over other approaches for plant disease classification. Texture features provide a more interpretable representation of the leaf's visual patterns, allowing domain experts to understand the factors contributing to disease detection. Moreover, these features have lower dimensionality than raw pixel data, reducing the risk of overfitting and making the model more data-efficient, especially in cases of limited labeled samples. By leveraging multi-objective optimization, the classification model can balance complexity and generalization, optimizing multiple objectives simultaneously, like error rate and accuracy. Additionally, the Random Forest technique helps improve the model's complete performance by iteratively adjusting the weights of weak learners to focus on misclassified samples. This combination of texture features, multi-objective optimization, and Random Forest can lead to faster inference times, increased interpretability, and better performance on smaller datasets, making it a compelling approach for accurate and effective plant disease detection.

5. Experiment Results and Discussion

5.1. Description of the Experiment

Table 1. Description of the experiment

Stage	Task	Details
Dataset Preparation	Dataset Selection	Use the Plant Village dataset, which includes images of various plant leaves in both healthy and diseased states.
	Dataset Splitting	Split the database into training (70%), validation (15%), and test (15%) sets.
Image Preprocessing	Resizing	Resize all images to a uniform size (e.g., 256x256 pixels).
	Normalization	Normalize pixel values in images (e.g., scale between 0 and 1).
	Augmentation (Optional)	Implement data augmentation methods like flipping scaled to improve the resilience of the dataset.
Feature Extraction	Initial Feature Extraction	Extract a comprehensive set of features from each image, including color, texture, shape, and morphological aspects.
	Feature Normalization	Normalize features to ensure they are on a similar scale.
Genetic Algorithm	Population Initialization	Initialize a population with random subsets of features and weights.
	Fitness Evaluation	Define a fitness function to evaluate the effectiveness of each feature set in classifying the images correctly.
	Selection, Crossover, Mutation	Apply genetic operations: select the best-performing feature sets, combine them (crossover), and introduce random modifications (mutation).
	Iteration	Repeat the genetic algorithm process for a predetermined number of generations or until convergence.
Ensemble Learning	Model Selection	Choose multiple machine learning models for the ensemble (e.g., Random Forest, SVM, Neural Networks).
	Model Training	Train each model in the ensemble using the optimized feature sets from the genetic algorithm.
	Prediction Combination	Combine predictions from all models in the ensemble to make a final prediction for each image in the test set.
Evaluation	Performance Metrics	Estimate the system by means of metrics like recall, precision, accuracy, and F1-score on the test set.
	Result Analysis	Examine the outcomes to determine how well the feature selection worked, how well each model performed in the ensemble, and how well the system performed overall.

5.2. Result Analysis

Table 2 illustrates a comparison of various methodologies according to the provided parameters. The NSGRF method exhibits superior performance compared to the other approaches, attaining the utmost accuracy of 98.34% with an error rate of 1.66%, precision of 96.34%, recall of 93.44%, and F-score of 95.22%.

Table 2. Result comparison of proposed model NSGRF with respect to different approaches

Approach	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)	Error (%)
RF	93.45	92.34	91.34	94.34	6.55
LR	90.12	89.12	90.23	92.34	9.88
SVM	96.34	94.3	91.23	93.22	3.66
P-SVM	94.34	97.3	93	95.12	5.66
GWO+RF	88.12	84.33	91.23	91.23	11.88
NSGRF	**98.34**	**96.34**	**93.44**	**95.22**	**1.66**

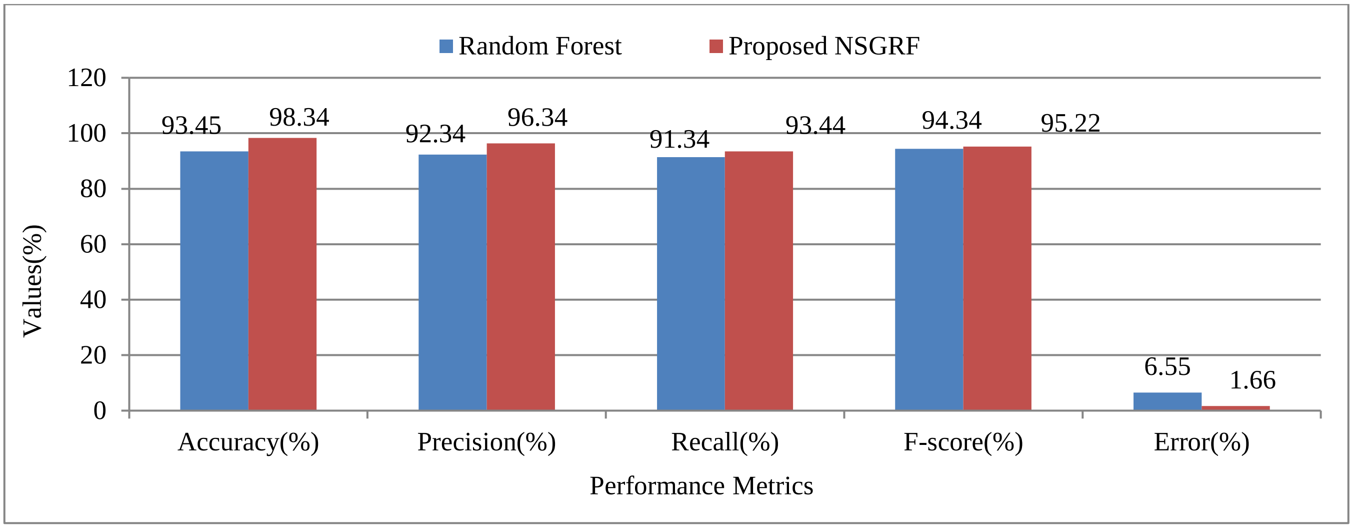


Fig. 7 Performance comparison of random forest and proposed NSGRF

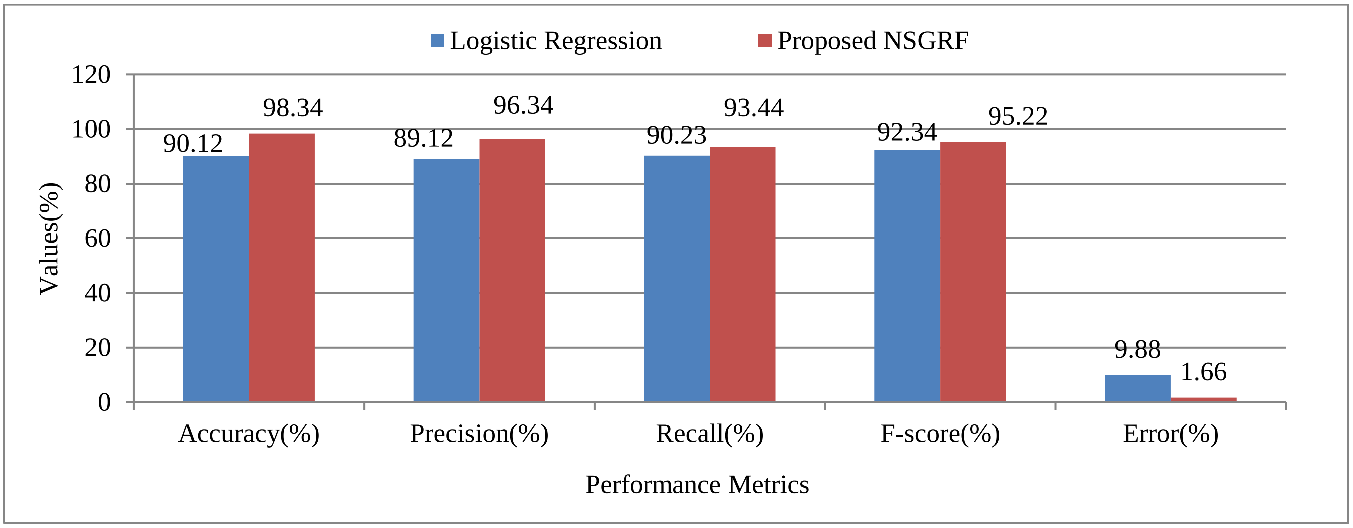


Fig. 8 Performance comparison of logistic regression and proposed NSGRF

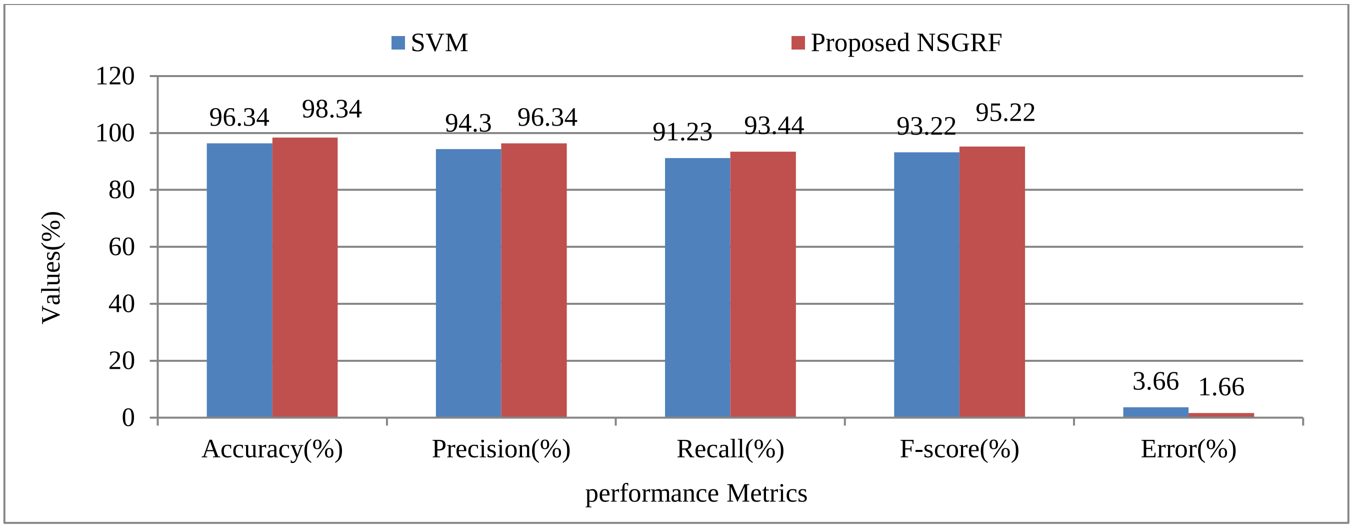


Fig. 9 Performance comparison of SVM and proposed NSGRF

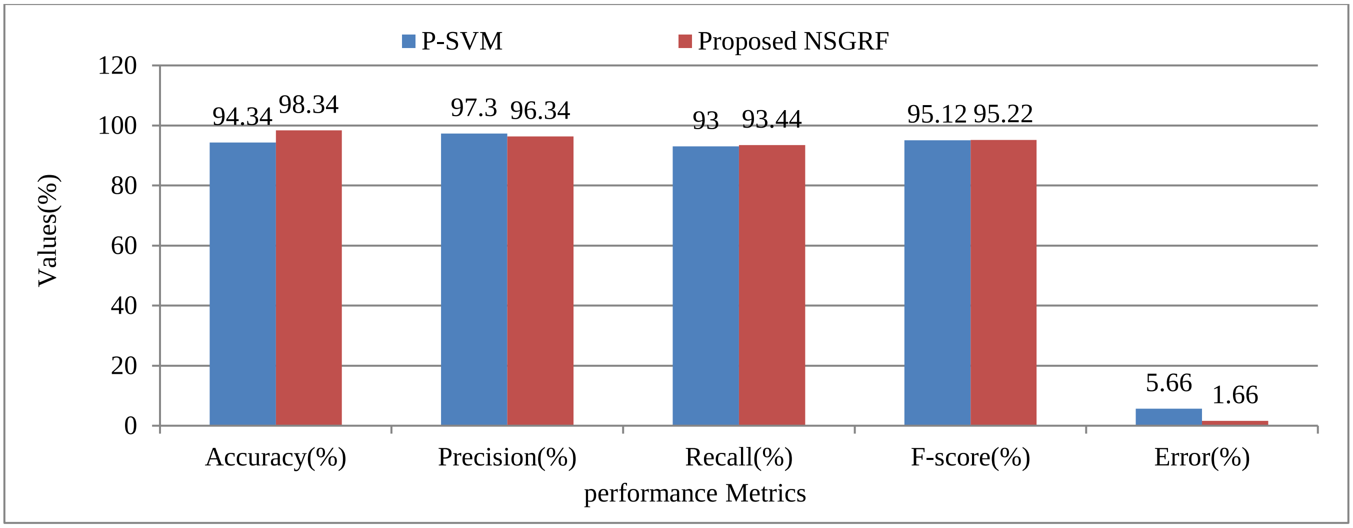


Fig. 10 Performance comparison of P-SVM and proposed NSGRF

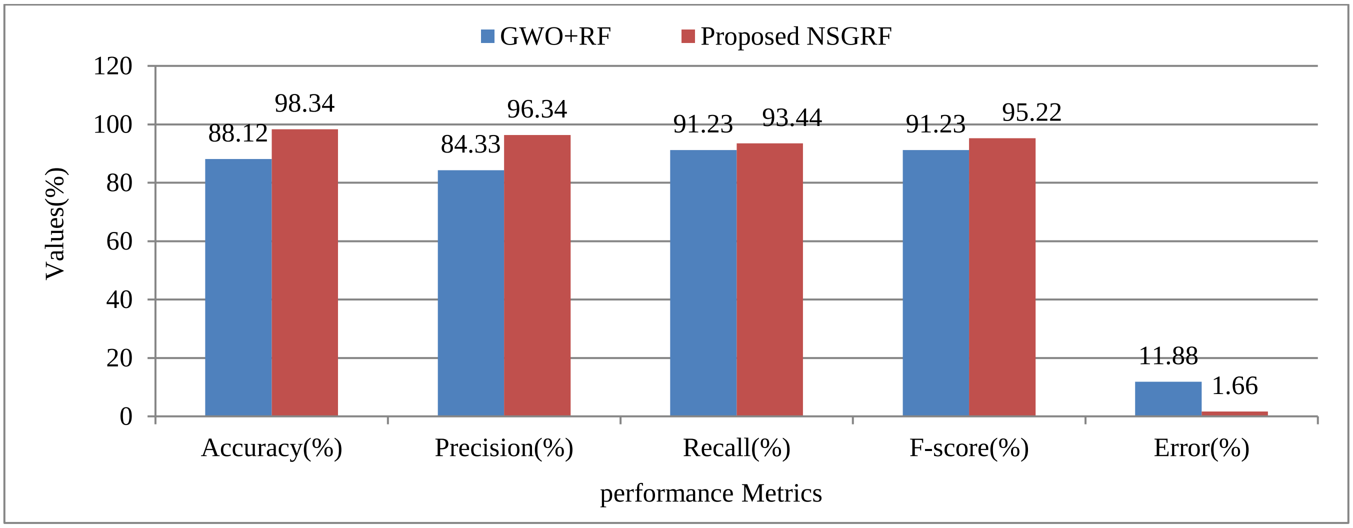


Fig. 11 Performance comparison of GWO-RF and proposed NSGRF

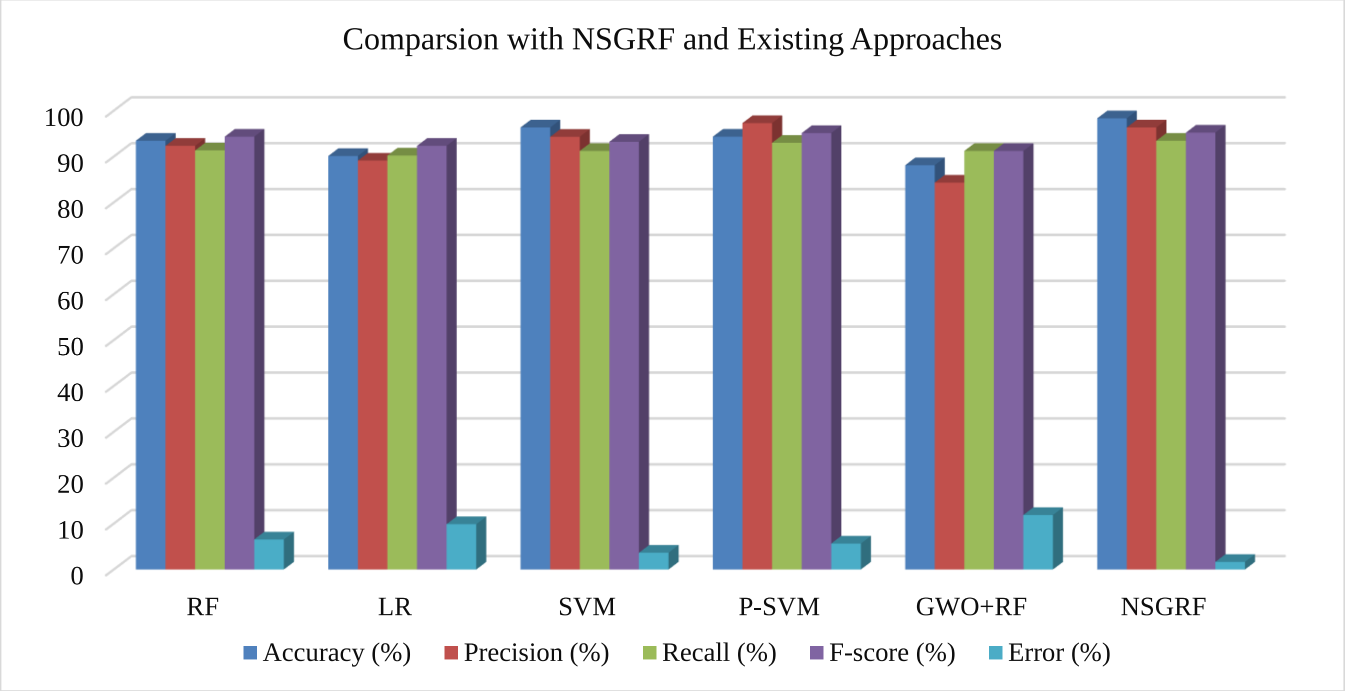


Fig. 12 Performance comparison of different classifiers with proposed NSGRF

A comparison of several approaches is shown in Figure 12, with an emphasis on how well they perform in terms of precision, accuracy, recall, and F-score. It is clear from the visual evidence that the NSGRF performs better than the other methods.

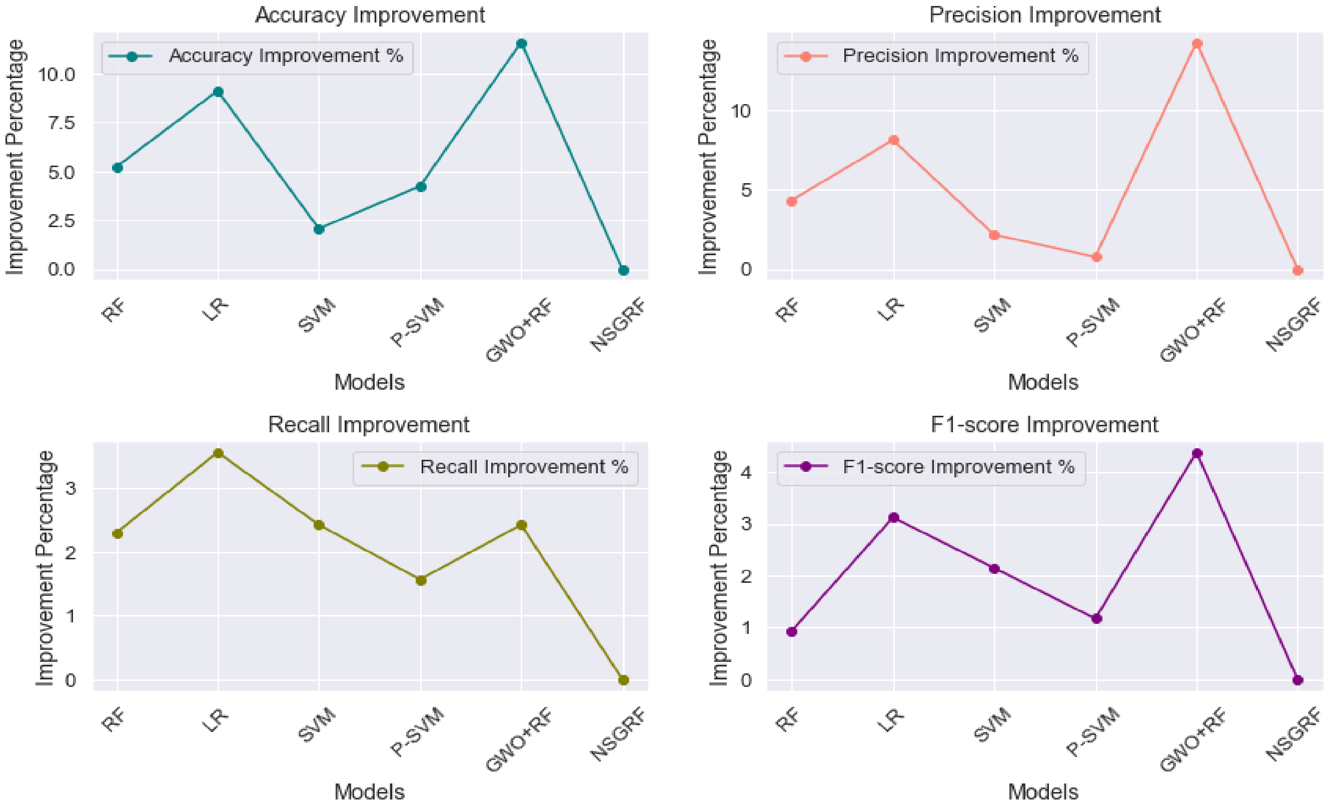


Fig. 13 Model Improvement of different approaches on different parameters

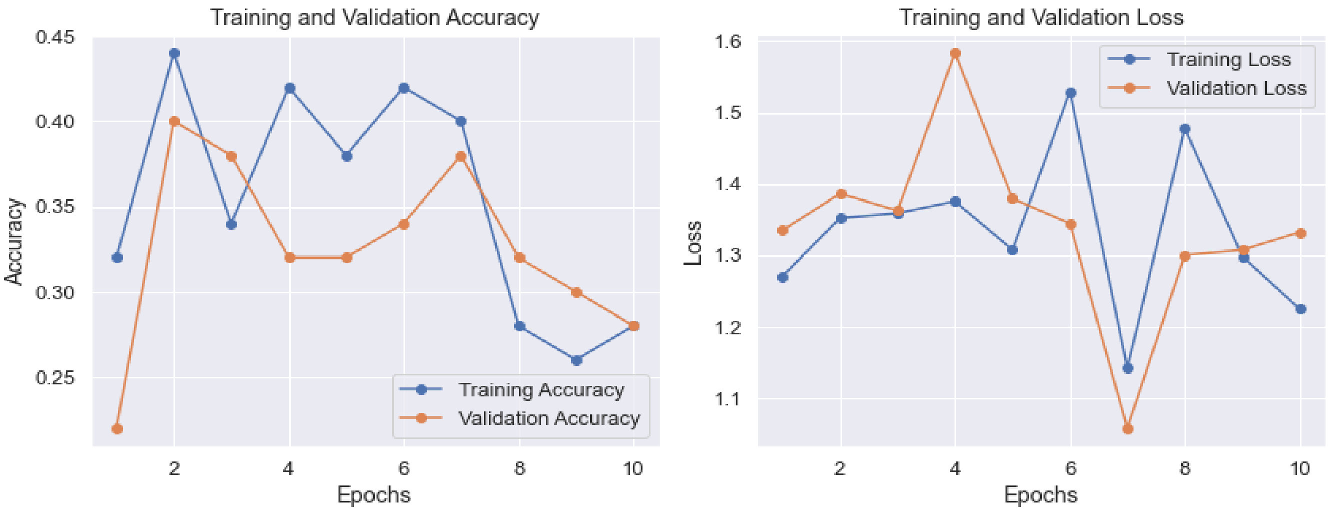


Fig. 14 Training and validation accuracy and loss

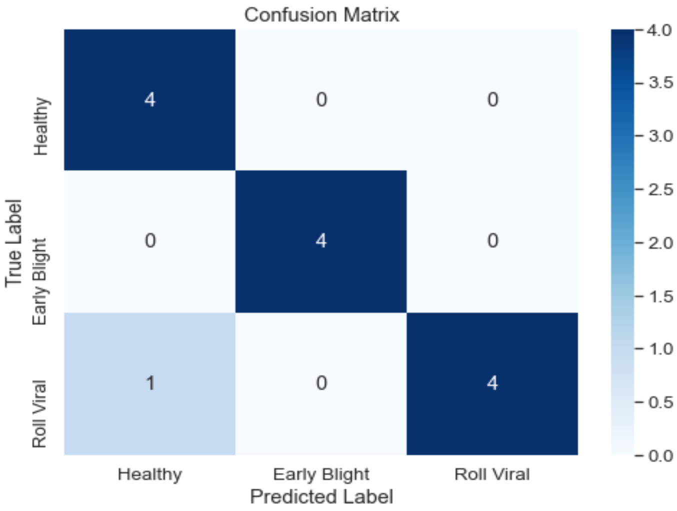


Fig. 15 Plot of confusion matrix for test data

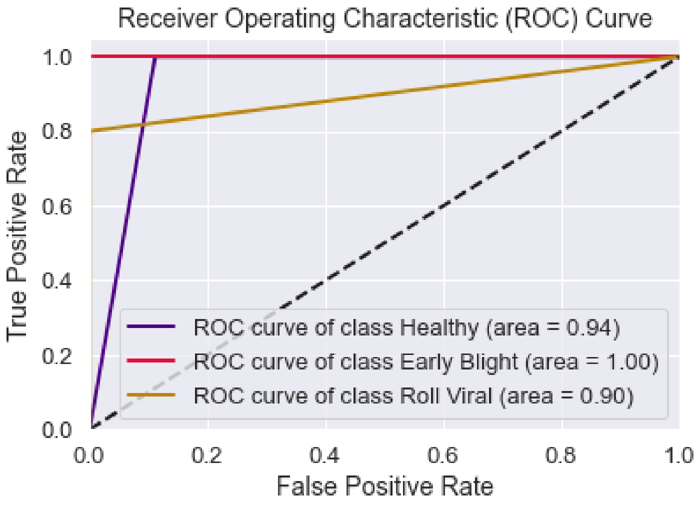


Fig. 16 ROC analysis for classification

Table 3 presents a comparison of different methodologies based on the specified parameters. The NSGRF method exhibits superior performance compared to the other two approaches, achieving the utmost accuracy of 98.00%, precision of 96.34%, recall of 93.44%, and F-score of 95.22%.

Table 3. Comparative analysis of proposed NSGRF with existing approaches

Authors	Leaf	Accuracy	Precision	Recall	F-score
Singh & Kaur [10]	Potato	95.99%	96.12%	96.25%	96.16%
Iqbal & Talukder [11]	Potato	97%	97%	97%	97%
Tiwari et al. [14]	Potato	97.80%	97.50%	97.50%	97.50%
Khalifa et al. [15]	Potato	98.00%	94.75%	93.22%	93.98%
NSGRF	**Potato**	**98.00%**	**96.34%**	**93.44%**	**95.22%**

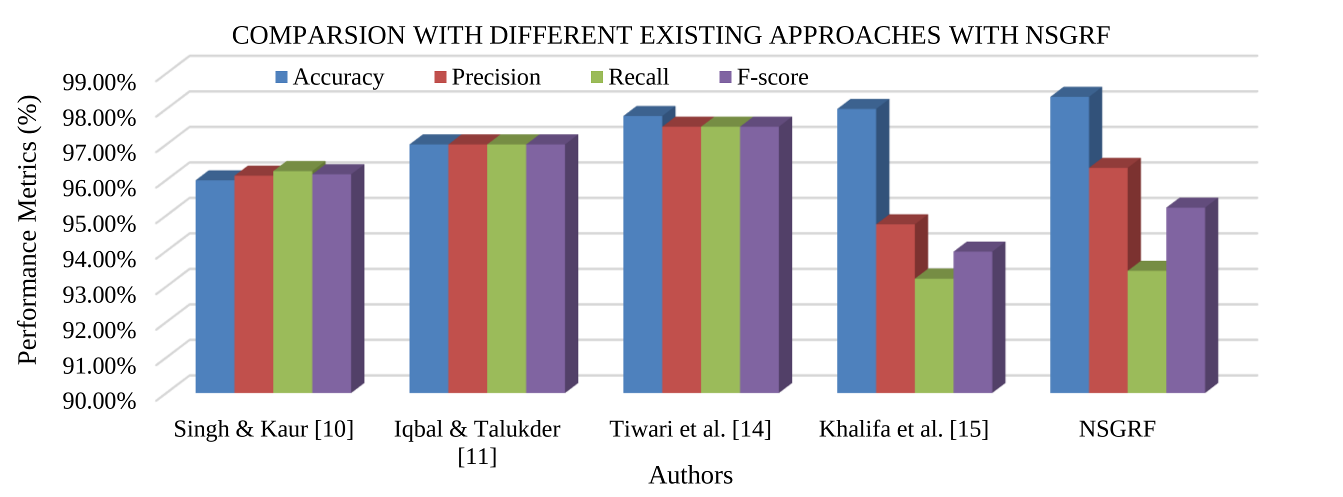


Fig. 17 Combined analysis of performance metrics by different authors with proposed NSGRF

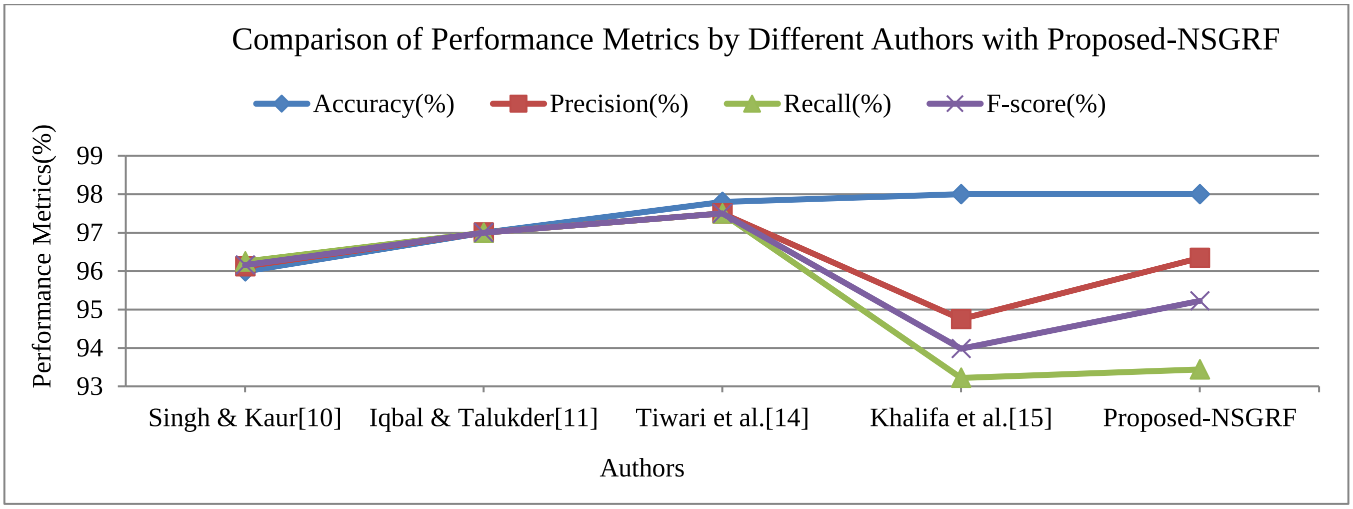


Fig. 18 Comparison of performance metrics by different authors with proposed NSGRF

Figures 17 and 18 depicts a comparative examination of numerous techniques, focusing on their performance in terms of accuracy, recall, precision, and F-score.

Based on the visual evidence, it is apparent that the NSGRF exhibits superior performance compared to the other approaches.

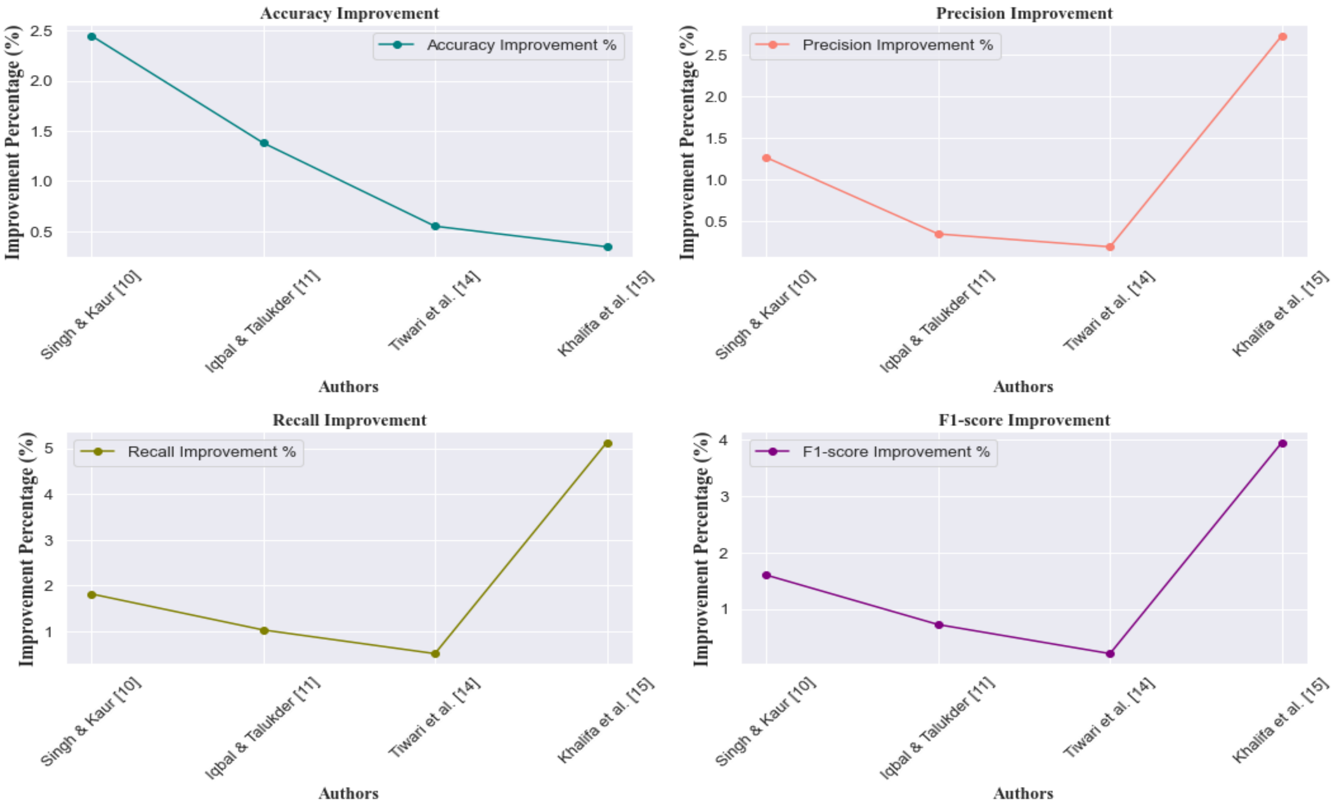


Fig. 19 Model improvement of different parameters on different authors

5.3. Contribution of Research

- Study introduces a novel integration of texture feature extraction, multi-objective optimization, and Random Forest learning. This combination is not commonly found in existing plant disease classification methodologies, making it a unique contribution to the field.
- Use of LBP and GLCM for texture feature extraction represents a significant advancement. This method provides a more efficient and interpretable representation of leaf patterns, which is crucial for understanding disease indicators.
- The application of multi-objective evolutionary algorithms (like MOEA/D and NSGA-II) for optimizing Random Forest parameters is a key contribution. This optimization ensures a balance between accuracy and model complexity, a critical aspect of machine learning.
- The utilization of a genetic algorithm for feature selection and weighting, combined with the ensemble of multiple machine-learning models, offers a robust and comprehensive approach to plant disease classification. This strategy significantly improves the overall classification performance.

6. Conclusion

Proposed approach efficacy of combining texture features, multi-objective optimization, and Random Forest in the plant disease classification utilizing the Plant Village dataset. The approach stands out for several reasons. The use of texture analysis, particularly through LBP and GLCM, provided a more interpretable and efficient representation of leaf patterns.

This not only enhanced the model's understanding of disease indicators but also lowered the risk of overfitting, a crucial factor in limited sample scenarios. Applying algorithms like NSGA-II and MOEA/D for optimizing Random Forest parameters helped balance the model's complexity and generalization capacity.

This ensured an optimal trade-off between accuracy and error rate, enhancing the model's robustness. The use of a Random Forest, optimized through multi-objective strategies, contributed significantly to improving classification performance. The iterative adjustment of weak learners' weights to focus on misclassified samples was pivotal in refining the model's accuracy. The GA and RF techniques further augmented the model's performance.

The genetic algorithm efficiently selected and weighted features, while the ensemble of multiple models ensured a comprehensive and reliable classification outcome. Experiment results corroborate the superiority of the suggested approach. The NSGRF method notably outperformed in all metrics error rate, precision, accuracy, recall, and F-score. This indicates a significant advancement over traditional methods in plant disease classification.

References

[1] Imtiaz Ahmed, and Pramod Kumar Yadav, "Plant Disease Detection Using Machine Learning Approaches," *Expert Systems*, vol. 40, no. 5, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Divyanshu Varshney et al., "Plant Disease Detection Using Machine Learning Techniques," *3rd International Conference for Emerging Technology*, Belgaum, India, pp. 1-5, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Mayuri Sharma, Chandan Jyoti Kumar, and Aniruddha Deka "Early Diagnosis of Rice Plant Disease Using Machine Learning Techniques," *Archives of Phytopathology and Plant Protection*, vol. 55, no. 3, pp. 259-283, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] S.K. Mahmudul Hassan et al., "A Survey on Different Plant Diseases Detection Using Machine Learning Techniques," *Electronics*, vol. 11, no. 17, pp. 1-23, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Anwar Abdullah Alatawi et al., "Plant Disease Detection Using AI Based VGG-16 Model," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 4, pp. 718-727, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] R. Sujatha et al., "Performance of Deep Learning Vs Machine Learning in Plant Leaf Disease Detection," *Microprocessors and Microsystems*, vol. 80, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[7] Vibhor Kumar Vishnoi, Krishan Kumar, and Brajesh Kumar, "Plant Disease Detection Using Computational Intelligence and Image Processing," *Journal of Plant Diseases and Protection*, vol. 128, no. 1, pp. 19-53, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[8] Saurav Roy et al., "Plant Disease Detection Using Machine Learning Tools with An Overview on Dimensionality Reduction," *Data Analytics in Bioinformatics: A Machine Learning Perspective*, pp. 109-144, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[9] Majji V. Applalanaidu, and G. Kumaravelan, "A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification," *3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks*, Tirunelveli, India, pp. 716-724, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Aditi Singh, and Harjeet Kaur, "Potato Plant Leaves Disease Detection and Classification Using Machine Learning Methodologies," *1st International Conference on Computational Research and Data Analytics*, Rajpura, India, vol. 1022, no. 1, pp. 1-10, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Md. Asif Iqbal, and Kamrul Hasan Talukder, "Detection of Potato Disease Using Image Segmentation and Machine Learning," *International Conference on Wireless Communications Signal Processing and Networking*, Chennai, India, pp. 43-47, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[12] Huiqun Hong, Jinfa Lin, and Fenghua Huang, "Tomato Disease Detection and Classification by Deep Learning," *International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering*, Fuzhou, China, pp. 25-29, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Chaojun Hou et al., "Recognition of Early Blight and Late Blight Diseases on Potato Leaves Based on Graph Cut Segmentation," *Journal of Agriculture and Food Research*, vol. 5, pp. 1-12, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Divyansh Tiwari et al., "Potato Leaf Diseases Detection Using Deep Learning," *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, pp. 461-466, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[15] Nour Eldeen M. Khalifa et al., *Artificial Intelligence in Potato Leaf Disease Classification: A Deep Learning Approach*, Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges, pp. 63-79, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[16] Akshay Kumar, and M. Vani, "Image Based Tomato Leaf Disease Detection," *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, pp. 1-6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[17] Alberta Odamea Anim-Ayeko, Calogero Schillaci, and Aldo Lipani, "Automatic Blight Disease Detection in Potato (Solanum Tuberosum L.) and Tomato (Solanum Lycopersicum, L. 1753) Plants Using Deep Learning," *Smart Agricultural Technology*, vol. 4, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Hardikkumar S. Jayswal, and Jitendra P. Chaudhari, "Plant Leaf Disease Detection and Classification Using Conventional Machine Learning and Deep Learning," *International Journal on Emerging Technologies*, vol. 11, no. 3, pp. 1094-1102, 2020. [Google Scholar] [Publisher Link]

[19] Sripada Swain, Sasmita Kumari Nayak, and Swati Sucharita Barik, "A Review on Plant Leaf Diseases Detection and Classification Based on Machine Learning Models," *Free Word Journal*, vol. 9, no. 6, pp. 5195-5205, 2020. [Google Scholar]

[20] Sanjay Patidar et al., "Rice Plant Disease Detection and Classification Using Deep Residual Learning," *International Conference on Machine Learning, Image Processing, Network Security and Data Sciences*, Silchar, India, pp. 278-293, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Murk Chohan et al., "Plant Disease Detection Using Deep Learning," *International Journal of Recent Technology and Engineering*, vol. 9, no. 1, pp. 909-914, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[22] Poojan Panchal, Vignesh Charan Raman, and Shamla Mantri, "Plant Diseases Detection and Classification Using Machine Learning Models," *4th International Conference on Computational Systems and Information Technology for Sustainable Solution*, Bengaluru, India, pp. 1-6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[23] Manpreet Kaur, and Rekha Bhatia, "Development of an Improved Tomato Leaf Disease Detection and Classification Method," *IEEE Conference on Information and Communication Technology*, Allahabad, India, pp. 1-5, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[24] Jyoti Shirahatti, Rutuja Patil, and Pooja Akulwar, "A Survey Paper on Plant Disease Identification Using Machine Learning Approach," *3rd International Conference on Communication and Electronics Systems*, Coimbatore, India, pp. 1171-1174, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[25] Jubin Dipakkumar Kothari, "Plant Disease Identification Using Artificial Intelligence: Machine Learning Approach," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 7, no. 11, pp. 11082-11085, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[26] N. Gobalakrishnan et al., "A Systematic Review on Image Processing and Machine Learning Techniques for Detecting Plant Diseases," *International Conference on Communication and Signal Processing*, Chennai, India, pp. 465-468, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[27] G. UshaDevi, and B.V. Gokulnath, "A Survey on Plant Disease Prediction Using Machine Learning and Deep Learning Techniques," *Inteligencia Artificial*, vol. 23, no. 65, pp. 136-154, 2020. [CrossRef] [Google Scholar] [Publisher Link]