International Journal of Knowledge Based Computer Systems
13 (1) 2025, 01-09
www.publishingindia.com//ijkbcs/

Potato Leaf Disease Detection using Machine Learning

Palak Jasrotia

M.Tech. Student, Department of Computer Science and IT, University of Jammu, Jammu and Kashmir, India. Email: palak9799@gmail.com

Abstract: The Paper "Potato Leaf Disease Detection using Machine Learning" focuses on the accurate identification of diseases in potato leaves to support early intervention and enhance crop yield. The process begins with Data Collection, where a dataset of 2,170 potato leaf images is sourced from Kaggle. This dataset is categorized into three classes: Early Blight, Late Blight, and Healthy. Next, Data Pre-Processing is undertaken to ensure the images are cleaned, resized, and normalized, preparing them for effective analysis. During Feature Extraction, relevant features from the images are identified to represent the data meaningfully. The dataset is then split into two subsets: the Train Image Set with 1,736 images, and the Test Image Set with 434 images, to facilitate model training and evaluation respectively. Various Machine Learning Techniques are applied, including k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes, Decision Tree, and Random Forest. Each trained model undergoes rigorous testing to evaluate its performance. To assess the effectiveness of each model, Performance Metrics such as accuracy, precision, recall, and F1-score are computed. The model exhibiting the highest accuracy is subjected to Feature Selection using Neighborhood Component Analysis (NCA) to enhance its performance further. Ultimately, a Hybrid Model is developed by combining the strengths of the individual models, aiming to improve overall accuracy and robustness in disease detection. This comprehensive approach integrates multiple machine learning techniques and feature selection methods, offering a robust solution for potato leaf disease detection. The project demonstrates the potential of machine learning in agricultural applications, contributing to more efficient and precise disease management.

Keywords: Decision tree, Feature extraction, Feature selection, K-nearest neighbor, Machine learning, Naive bayes, Potato leaf disease, Random forest, Support vector machine.

I. Introduction

Agriculture productivity plays a very significant role in the Indian economy. Agriculture is the major occupation in India. The economy of our country highly depends on agriculture and its associated products. India stood first in the world with the highest net cropped area followed by the US and China. Various pests and diseases affect the plant growth, quantity and quality of the product. So, it is very necessary to detect the disease at an early stage of the growth of the plant. India is a cultivation nation, more than 60% of the population of India has agriculture and its productivity as their main occupation. In India, agriculture is dependent on the monsoon. When the monsoon is good, agriculture productivity is good; when the monsoon is less, productivity is less or not in good condition [1].

Potato is one of the most cultivated crops. Worldwide potatoes have their cultivation priority as a staple food. For successful potato production, a strong food security system can be developed as it is a great source of vitamins and minerals. However, several diseases affect potato production and degrade agricultural development. Therefore, disease detection in an early stage can provide a better solution for successful crop cultivation [2].

II. Literature Review

Bhuyar (2014) [3] proposed an approach where different classification algorithms such as J48, Naïve Bayes, and Random forest algorithm were applied to soil dataset to predict its fertility. J48 algorithm gave better result with an accuracy of 98.17% than other algorithms.

Singh *et al.* (2015) [4] performed a review of the literature on various image processing techniques for detecting leaf disease. The writers wanted to speed up identification and detection of plant diseases while lowering the subjectivity that comes with naked-eye observation. They presented an algorithm that uses picture segmentation to automatically identify and categorize plant leaf diseases. The impact of HSI, CIELAB, and YCbCr color spaces on disease spot detection was examined by the authors. To identify the disease spot, the Otsu technique was applied to the color component after an image was smoothed using a median filter. The suggested method was not put to the test on any datasets.

He *et al.* (2015) [5] deep residual networks were discussed in relation to image recognition challenges. As shallow representations for image retrieval and classification, VLAD and Fisher Vector are cited in the paper's literature review as related concepts. The writers also covered the advantages of encoding residual vectors over original vectors for vector quantization. The Multigrid method was also mentioned as a method for solving partial differential equations (PDEs) by reformulating systems as sub-problems at multiple scales, where each subproblem is accountable for the residual solution between a coarser and a finer scale. This method is used in low-level vision and computer graphics.

Rajeshwari and Arunesh (2016) [6] performed a comparative analysis of ML algorithms i.e., Naive Bayes, JRIP and J48 for prediction of soil types. The experiments were performed on soil data consisting of 110 samples using data analytics tool R. The experimental results predicted that JRIP algorithm performed better as it gave highest accuracy of 98.18% with kappa statistic of approximate 1.0.

Sujata (2016) [7] proposed a model to estimate the crop yield in order to improve the value and gain of farming area using data mining techniques.

Islam *et al.* (2017) [8] suggested a method to identify diseases from images of potato plant leaves by combining image processing and machine learning. The authors classified diseases using the 'Plant Village' database of openly accessible plant images. Using the suggested method, the research classified 300 images of diseases with a 95% accuracy. The paper also highlights how crucial contemporary phenotyping and plant disease detection are to assuring food security and sustainable agriculture.

Tiwari *et al.* (2020) [9] proposed a methodology in which VGG19 was used for feature extraction, and multiple classifiers, including logistic regression, SVM, KNN, and neural networks, were applied for classification. Logistic regression, in combination with VGG19, achieved a classification accuracy of 97.8%. The proposed approach outperformed previous models, achieving a higher classification accuracy (97.8%). Various approaches, including image segmentation and traditional image processing techniques, have been used for disease detection.

Tarik *et al.* (2021) [10] proposed a methodology that involves the use of Convolutional Neural Networks (CNNs) for detecting various types of potato diseases. The authors collected a dataset consisting of 2034 images of healthy and diseased potato leaves. The images were preprocessed using OpenCV, and a CNN model architecture was employed for disease classification. The CNN model achieved a high accuracy rate of 84% for disease detection. The training, testing, and validation processes were discussed, indicating successful results. The authors mention the use of transfer learning to enhance the model's performance, adjusting pre-trained weights to match the desired output dimensions.

Afzaal *et al.* (2021) [11] explored the use of artificial intelligence, specifically convolutional neural networks (CNNs) and deep learning, for the early detection of early blight disease in potato plants. They collected a dataset of images of healthy and diseased potato plants from four fields, using different lighting conditions. The CNNs used in the study were Google Net, VGG Net, and Efficient Net, trained using the PyTorch framework. The images were classified into three classes (2-class, 4-class, and 6-class) to accurately identify the disease at different growth stages. Results showed that Efficient Net and VGG Net outperformed Google Net in terms of disease identification accuracy.

Rashid *et al.* (2021) [12] proposed PDDCNN method, applied to the Potato Leaf Disease (PLD) dataset, achieved an impressive accuracy of 99.75%, demonstrating high precision, recall, F1-score, and an excellent ROC curve. With minimal parameters, it surpasses existing methods in computational efficiency. Future extensions include multi-disease detection on a single leaf, disease localization, severity estimation, dataset enhancement, an IoT-based real-time monitoring system, and a website/mobile application for broader accessibility. Cross-dataset evaluations on Plant Village and PLD datasets, with and without augmentation, further validated the superiority of the PDDCNN method.

Singh and Kaur (2021) [13] proposed a methodology that includes a description of the necessity of image segmentation in analyzing colored images. Implementation of the K-means algorithm for segmentation. Steps involved in the K-means algorithm. Mentioning of distance metrics like Minkowskis Manhattan, and Euclidean distances. Explanation of the importance of feature extraction in reducing dimensionality. Adaptation of the binary classification for multiclass classification in the context of plant disease detection. Use of the proposed framework on the 'Plant Village' dataset. Achievement of an overall accuracy of 95.99%. dataset for three popular crops (tomato, corn and potato) and the accuracy was 90.01%.

III. Research Methodology

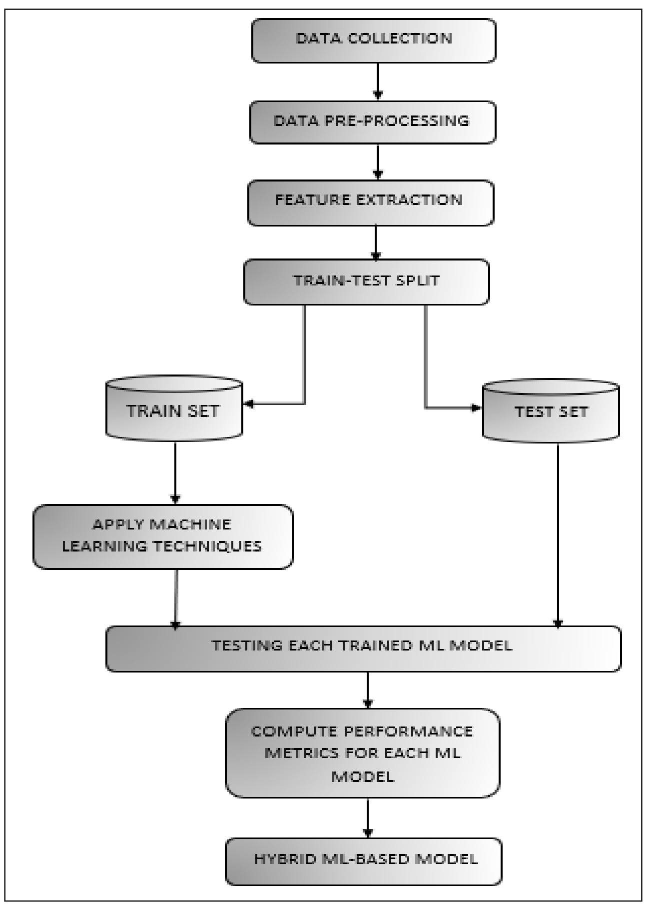

Fig. 1: Methodology for the Current Study

A. Data Collection

The dataset contains images of potato leaves in .jpg format and has been sourced from Kaggle (https://www.kaggle.com/datasets/emmarex/plantdisease) [14].

TABLE I: NUMBER OF IMAGES IN DATASETS

Class	No. of Images
Early Blight	750
Late Blight	693
Healthy	727
Total	2170

It comprises a total of 2,170 images, which are categorized into three distinct classes. The "Early Blight" class contains 750 images, representing the highest number of images among the classes. The "Late Blight" class has 693 images, making it the class with the fewest images. The "Healthy" class includes 727 images. Together, these classes form a comprehensive dataset for analyzing and distinguishing between early blight, late blight, and healthy conditions in plants. Samples of Potato Leaf dataset are shown in Fig. 2.

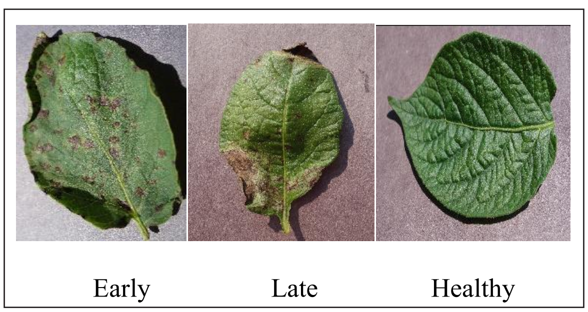

Fig. 2: Sample of Collected Potato Leaf Images

B. Data Pre-Processing

After data collection, preprocessing operations and filtering were performed on the collected images of the dataset. The images were filtered using median filtering to smooth the edges and remove noise. Median filters are particularly useful in reducing random noise, especially when the noise amplitude probability density has large tails and periodic patterns [15]. The dataset was then divided into an 80:20 ratio for training and testing purposes. The Train Image Set comprised 1,736 images, while the Test Image Set included 434 images.

RGB image is converted into grayscale image using grayscale () function.

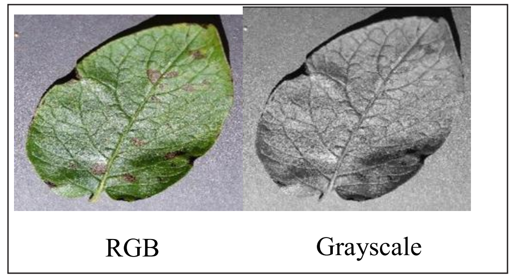

Fig. 3: RGB to Grayscale

TABLE II: CLASS LABELS OF POTATO LEAVES

Class Labels	Potato Leaves
Class 0	Early Blight
Class 1	Late Blight
Class 2	Healthy

C. Implementation of the Framework

The framework begins with the conversion of RGB images into grayscale images using the grayscale () function. This step simplifies the image data and prepares it for feature extraction. In the feature extraction phase, features of the Potato Leaves dataset are extracted using the Gray-Level Co-Occurrence Matrix (GLCM). The GLCM is a statistical method used to examine texture by considering the spatial relationship of pixels within an image. Specifically, it calculates how often pairs of pixels with specific values occur in a specified spatial relationship, thereby creating a GLCM [16].

To characterize the texture of the images, the greycomatrix () function is used to generate the GLCMs, and the greycoprops () function extracts specific features from these matrices. The features calculated from the GLCM include Angular Second Moment (ASM), Contrast, Dissimilarity, Homogeneity, Energy, and Correlation. These features are extracted at four different angles: 0°, 45°, 90°, and 150° [17].

In addition to the GLCM features, the framework computes additional statistical features such as mean, standard deviation, variance, skewness, and kurtosis. In total, 29 features are extracted by applying GLCM feature extraction at the specified angles. This comprehensive set of features provides a detailed representation of the texture and statistical properties of the Potato Leaves dataset, which is crucial for subsequent analysis and modeling steps.

D. Technique Used for Feature Extraction

In this section, Features of the Potato Leaf datasets have been extracted using the Gray-Level Co-Occurrence Matrix (GLCM).

E. Implementation of Gray-Level Co-Occurrence Matrix (GLCM)

The framework for feature extraction begins with converting RGB images to grayscale using the grayscale () function. This step simplifies the data and prepares it for the next phase. The feature extraction process involves the use of the Gray-Level Co-Occurrence Matrix (GLCM), a statistical method that examines texture by considering the spatial relationship of pixels. The GLCM, also known as the gray-level spatial dependence matrix, characterizes the texture of an image by calculating how often pairs of pixels with specific values occur in a specified spatial relationship within the image. The GLCMs are created using the greycomatrix () function. Subsequently, the greycoprops () function extracts specific features from the GLCM. These features include Angular Second Moment (ASM), Contrast, Dissimilarity, Homogeneity, Energy, and Correlation. To capture comprehensive texture information, the features are extracted at four different angles: 0°, 45°, 90°, and 150°. Additionally, statistical features such as mean, standard deviation, variance, skewness, and kurtosis are computed. In total, 29 features are extracted by applying GLCM feature extraction at the specified angles, providing a detailed representation of the texture and statistical properties of the Potato Leaves dataset.

F. Texture Features

At first level decomposition of GLCM, texture features like ASM, contrast, homogeneity, correlation, dissimilarity and energy have been calculated for each coefficient. A total of 29 features Classification of Potato Leaves using Machine Learning Techniques for each image were calculated.

Angular Second Moment (ASM), also known as Energy, is a feature derived from the Gray-Level Co-Occurrence Matrix (GLCM) used to characterize the texture of an image. It measures the uniformity or smoothness of the texture.

ASM = ∑i∑jP(i, j)2

Contrast is a measure of intensity contrast between a pixel and its neighbor over the whole image.

Contrast = | *iji* − j| 2 *p*(*i*, *j*)

Homogeneity measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

Homogeneity = 1 1+|*i*−*j*| *ij* 2 *p*(*i*, *j*)

Correlation is a measure of how to correlate a pixel to its neighbor over the whole image.

Correlation = *i*−μ*ij*−μ*jpi*,*j*σ*i*σ*jij*

Dissimilarity is a measure of distance between pairs of pixels in the region of interest.

Dissimilarity = *i* − *jjpi*,*ji*

Energy provides the sum of squared elements in the GLCM.

Energy = *Pi*,*ji*,*j* 2 [18]

After the implementation of GLCM, all the computed features have been saved in .csv file. The resulting size of .csv file becomes 2170*30 since the dataset contains 2170 images and 29 are the input extracted texture features and 01 is the output class label. Later, the resulting dataset is divided in the ratio 80:20, into two sets, train set and test set. Thus, train set contains 1736 images and test set contains 434 images.

G. Technique Used for Feature Selection

To further enhance the performance of the Random Forest model, Neighborhood Component Analysis (NCA) feature selection was applied, aiming to achieve better results by identifying and utilizing the most relevant features for classification.

H. Implementation of Neighborhood Component Analysis (NCA)

Neighborhood Component Analysis (NCA) was applied to the dataset for feature selection. NCA is a supervised dimensionality reduction method that enhances the predictive power of the model by selecting the most relevant features. Neighborhood Component Analysis for Feature Selection (NCAFSR) is implemented to select the most relevant features from the dataset before training a classification model. The process begins by importing necessary libraries such as pandas, numpy, and scikit-learn modules. Three datasets, representing different stages or conditions, are loaded from CSV files using pd.read_csv. Any missing values in these datasets are filled with the mean of the respective columns. The data is then normalized using the MinMaxScaler from scikit-learn's preprocessing module to scale the features between 0 and 1. After normalization, the data is combined into a single array and shuffled to ensure random distribution.

The combined data is split into features (X) and labels (Y), and then into training and testing sets using train_test_split. Feature selection is performed using Neighborhood Component Analysis for Feature Selection and Reduction (NCAFSR), which fits the model to the training data to learn the weights of the features. These weights help in transforming the training and testing sets by emphasizing more relevant features. A Random Forest Classifier is then trained on the transformed training set, and predictions are made on the transformed testing set. The performance of the classifier is evaluated using a confusion matrix, classification report, and accuracy score, providing insights into the model's effectiveness.

The confusion matrix illustrates the performance of a classification model across three classes. Class 0 has 138 true positives, 7 false negatives, and 11 false positives. Class 1 shows 124 true positives, 11 false negatives, and 15 false positives. Class 2 exhibits 134 true positives, 2 false negatives, and 3 false positives. The classification report provides detailed metrics for precision, recall, and F1-score for each class. Class 0 achieves a precision of 0.93, recall of 0.94, and F1-score of 0.93, based on 147 instances. Class 1 has a precision of 0.85, recall of 0.90, and F1-score of 0.87, encompassing 138 instances. Class 2 shows perfect precision (1.0), recall of 0.96, and F1-score of 0.98, covering 149 instances. The overall accuracy of the model is 0.91, with macro-averaged precision, recall, and F1-score also at 0.91, calculated across all 434 instances. The weighted averages for precision, recall, and F1-score are consistently 0.91. The "Accuracy" score provided is 0.9124423963133641, indicating the overall effectiveness of the model in correctly predicting classes across the dataset. It provides a complete workflow for detecting potato leaf diseases using machine learning. It demonstrates the importance of data preprocessing, feature selection, and the application of an effective classifier to achieve accurate and reliable results. By applying NCA for feature selection and using a Random Forest classifier, the model is optimized for better performance in detecting various potato leaf diseases.

I. Performance Metrics Used

In order to evaluate the performance of all the ML trained models for crop yield prediction, four important measures namely *accuracy*, *recall*, *precision*, and *F-score* has been computed. These measures can be calculated from confusion matrix described as under:

As multiple (three) class labels (Low, Mid, High) are there in both the wheat and mustard crop yield prediction problem, thus 3*3 confusion matrix is used. It is a technique for summarizing the performance of classification algorithms.

TABLE III: CONFUSION MATRIX

	Predicted Class			
Actual Class		Low (L)	Mid (M)	High (H)
	Low (L)	TP_L	E_{LM}	E_{LH}
	Mid (M)	E_{ML}	TP_M	E_{MH}
	High (H)	E_{HL}	E_{HM}	TP_H

- *True Positive (TP):* These are the correctly predicted positive values which mean value of both actual and predicted class is yes. True positive will be diagonal values for every class TP_L, TP_M, TP_H.
- *True Negative (TN):* These are the correctly predicted negative values which mean value of both actual and predicted class is no. True negative for a certain class will be the sum of all columns and rows excluding that class's column and row.

  TN for class *Low*, $TN_L = TP_M+E_{MH}+E_{HM}+TP_H$

  TN for class *Mid*, $TN_M = TP_L+E_{LH}+E_{HL}+TP_H$

  TN for class *High*, $TN_H = E_{LM}+E_{LH}+TP_M+E_{MH}$
- *False Positive (FP):* When value of actual class is no but the predicted class is yes. False positive (FP's) for a class is the sum of values in the corresponding column (excluding the TP).

  FP for class *Low* = $E_{ML}+E_{HL}$

  FP for class *Mid* = $E_{LM}+E_{HM}$

  FP for class *High* = $E_{LH}+E_{MH}$
- *False Negative (FN):* When value of actual class is yes but the predicted class is no; False Negative (FP's) for a class is the sum of values in the corresponding row (excluding the TP).

  FN for class *Low* = $E_{LM}+E_{LH}$

  FN for class *Mid* = $E_{ML}+E_{MH}$

  FN for class *High* = $E_{HL}+E_{HM}$
- *Accuracy:* Accuracy is the ratio of correctly predicted observations to the total observations. Accuracy is great measure but only when one has symmetric datasets where value of false positive and negative are almost same.

  Accuracy = TP+TN/TP+FP+FN+TN
- *Precision:* Precision is the ratio of correctly predicted observations to the total predicted positive observations. High precision leads to low false positive rate.

  Precision = TP/(TP+FP)
- *Recall:* Recall is the ratio of correctly predicted positive observations to the all observations in the actual class-yes.

  Recall = TP/(TP+FN)
- *F1-Score:* F1 Score is the weighted average of precision and recall. Mostly used when there is an uneven class distribution.

  F1-Score = 2*Recall*Precision / (Recall + Precision) [18]

IV. RESULTS AND DISCUSSION

The performance of various machine learning techniques was evaluated using several metrics. The K-Nearest Neighbors (KNN) algorithm achieved an accuracy of 77.1%, with a precision of 80%, recall of 77%, and an F-Score of 0.777. The Support Vector Machine (SVM) method demonstrated an accuracy of 79.4%, precision of 85%, recall of 79%, and an F-Score of 0.80. Decision Trees (DT) exhibited an accuracy of 80.8%, with both precision and recall at 81%, and an F-Score of 0.81. The Random Forest (RF) classifier showed a significant improvement, attaining an accuracy of 84.1%, precision of 85%, recall of 84%, and an F-Score of 0.84. Lastly, Naive Bayes (NB) presented an accuracy of 79%, precision of 83%, recall of 79%, and an F-Score of 0.78.

TABLE IV: EXPERIMENTAL RESULTS FOR ML TECHNIQUES UNDER STUDY

ML Techniques	Accuracy	Precision	Recall	F-Score
KNN	77.1%	80%	77%	0.77
SVM	79.4%	85%	79%	0.80
DT	80.8%	81%	81%	0.81
RF	84.1%	85%	84%	0.84
NB	79%	83%	79%	0.78

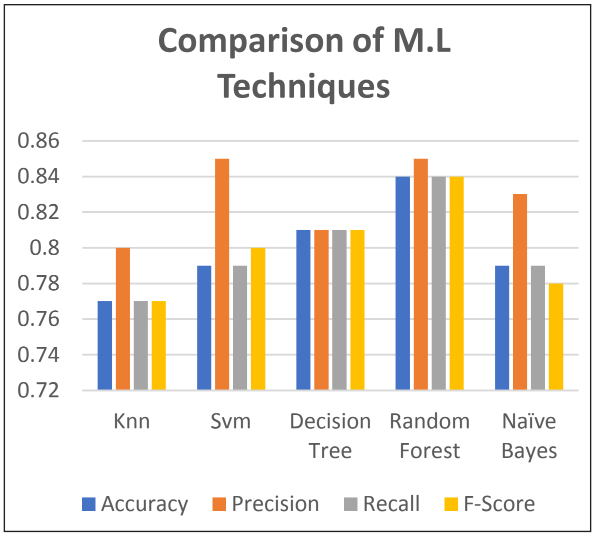

Fig. 4: Comparative Analysis of All ML Techniques

The Hybrid ML-Based Model, specifically the Modified Random Forest (Modified RF), demonstrated superior performance compared to the individual machine learning techniques. The Modified RF model achieved an impressive accuracy of 91.2%, with precision, recall, and F-Score all at 91%. This indicates that the hybrid approach significantly enhances the model's predictive capabilities.

Table V: Experimental Results for Hybrid ML-Based Model under Study

ML Techniques	Accuracy	Precision	Recall	F-Score
Hybrid ML-Based Model	91.2%	91%	91%	0.91

V. Conclusion

In this research, we explored the efficacy of machine learning techniques for detecting potato leaf diseases based on texture features extracted from grayscale images using the Gray-Level Co-Occurrence Matrix (GLCM). The study encompassed three main conditions: Early Blight, Late Blight, and Healthy leaves, each characterized by distinct textural patterns captured through GLCM analysis. Our findings demonstrated that the Random Forest classifier, augmented by Neighborhood Component Analysis for Feature Selection (NCAFS), achieved the highest accuracy of 84.1% among the tested models. This approach not only effectively classified potato leaf diseases but also highlighted the importance of robust feature extraction and selection methods in enhancing classification accuracy. The results underscore the potential of machine learning in precision agriculture, offering a promising tool for early disease detection and proactive crop management strategies. Future research could explore larger datasets, integrate more sophisticated feature extraction techniques, and validate the model in field conditions to further enhance its applicability and impact in agricultural settings.

References

[1] V. P. Gaikwad, and V. P. Musande, "Potato plant leaf disease detection using CNN model," *European Chemical Bulletin,* vol. 12, no. 1, pp. 516-527, 2023.

[2] M. A. R. Nishad, M. A. Mitu, and N. Jahan, "Predicting and classifying potato leaf disease using K-means segmentation techniques and deep learning networks," *Procedia of Computer Science,* vol. 212, pp. 220-229, 2022, doi: https://doi.org/10.1016/j.procs.2022.11.006.

[3] V. Bhuyar, "Comparative analysis of classification techniques on soil data to predict fertility rate for Aurangabad district," *International Journal of Emerging Trends & Technology in Computer Science,* vol. 3, no. 2, pp. 200-203, 2014.

[4] V. Singh, Varsha, and A. K. Misra, "Detection of unhealthy region of plant leaves using image processing and genetic algorithm," *International Conference on Advances in Computer Engineering and Applications,* 2015, doi: https://doi.org/10.1109/icacea.2015.7164858.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2016, doi: https://doi.org/10.1109/cvpr.2016.90.

[6] V. Rajeshwari, and K. Arunesh, “Analyzing soil data using data mining classification techniques,” *Indian Journal of Science and Technology*, vol. 9, no. 19, pp. 1-4, 2016.

[7] R. Sujatha, “A study on crop yield forecasting using classification techniques,” *IEEE,* 2016.

[8] M. Islam, K. Wahid, and P. Bhowmik, “Detection of potato diseases using image segmentation and multiclass support vector machine,” *IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2017, doi: https://doi.org/10.1109/CCECE.2017.7946594.

[9] D. Tiwari, M. Ashish, N. Gangwar, A. Sharma, S. Patel, and S. Bhardwaj, “Potato leaf diseases detection using deep learning,” *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020) IEEE Xplore*, 2020, doi: https://doi.org/10.1109/iciccs48265.2020.9121067.

[10] M. I. Tarik, S. Akter, A. A. Mamun, and A. Sattar, “Potato disease detection using machine learning,” *Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 2021, doi: https://doi.org/10.1109/icicv50876.2021.9388606.

[11] H. Afzaal, A. A. Farooque, A. W. Schumann, N. Hussain, A. McKenzie-Gopsill, T. Esau, F. Abbas, and B. Acharya, “Detection of a potato disease (early blight) using artificial intelligence,” *Remote Sensing*, vol. 13, no. 3, 2021, doi: https://doi.org/10.3390/rs13030411.

[12] J. Rashid, I. Khan, G. Ali, S. H. Almotiri, M. Alghamdi, and K. Masood, “Multi-level deep learning model for potato leaf disease recognition,” *Electronics,* vol. 10, no. 17, 2021, doi: https://doi.org/10.3390/electronics10172064.

[13] A. Singh, and H. Kaur, “Potato plant leaves disease detection and classification using machine learning methodologies,” *IOP Conference Series: Materials Science and Engineering*, 2021, doi: https://doi.org/10.1088/1757-899x/1022/1/012121.

[14] Data Collection. [Online]. Available: https://www.kaggle.com/datasets/emmarex/plantdisease

[15] K. Padmavathi, and K. Thangadurai, “Implementation of RGB and grayscale images in plant leaves disease detection - Comparative study,” *Indian Journal of Science and Technology*, vol. 9, 2016, doi: https://doi.org/10.17485/ijst/2016/v9i6/77739.

[16] Implementation of the Proposed Framework. [Online]. Available: https://www.mathworks.com/help/images/texture-analysis-using-the-gray-level-co-occurrence-matrix-glcm.html

[17] Gray-Level Co-Occurrence Matrix (GLCM). [Online]. Available: https://medium.com/@girishajmera/feature-extraction-of-images-using-glcm-gray-level cooccurrence-matrix-e4bda8729498

[18] Y. Gupta, H. Kour, J. Manhas, and V. Sharma, “Classification of spices using machine learning techniques,” *International Journal of Knowledge Based Computer Systems*, vol. 10, no. 1, pp. 27-32, 2022.