International Journal of Computer Techniques – Volume 12 Issue 2, March - April - 2025

Statistical Analysis of RGB and Edge detection model in Classification of Plant Diseases

Shruti Saini[1], Sugandh Raghav[1], Dr. Shivani Saxena[2]
[1,] Department of Physical Sciences, Banasthali Vidyapith, Banasthali, Rajasthan
[2,] Assistant Professor, Department of Physical Sciences, Banasthali Vidyapith, Rajasthan

Abstract: The proposed work aims to analyse model trained by features extracted from RGB and edge detection method for classification of plant leaf diseases. Image dataset of Apple, Bell Pepper, Cherry, Corn, Peach, Potato, Rice, Strawberry, Tomato, Wheat are taken from Kaggle datasets. For simulation and extracting features from plant leaf images MATLAB is used, then models are trained using deep learning toolbox. The performance of models is evaluated using confusion matrix and ROC curve, which helps to determine which model is more efficient for classification and detection of plant leaf diseases. Accuracy of the models is determined after training.

Key Words: Confusion Matrix, Classification of Plant Leaf Diseases, Edge Detection Model, Red Green Blue (RGB) model, Receiver Operating Characteristic (ROC) curve.

1 Introduction

Large number people in India are still dependent on agriculture for their livelihood and food security. Disease and pests are major problems that affects the productivity and quality of the crops, diagnosis of disease require timely examination and experience. Plants disease can cause health issues if spoiled crops are consumed [1]. Fig 1, shows comparison between traditional and precision farming based on factors like disease prevention, environmental effect, economical, use of resources and cultivation.

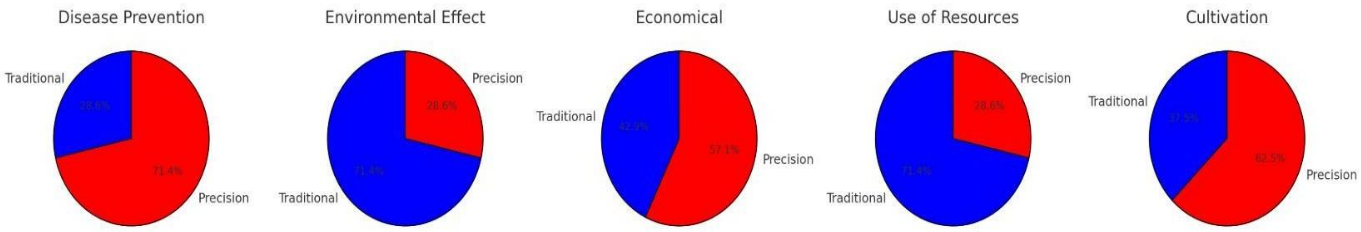

Fig 1: Comparison between traditional and precision farming based on different parameters

Manual inspection, the method used for traditional disease identification, is labour-intensive, erratic, and out of reach for small-scale farmers. Accuracy is decreased by differences in symptoms between growth stages and environmental factors, which results in postponed action and financial losses [2]. Fig 2 shows overall comparison between traditional and precision farming which concludes that precision farming is more beneficial for farmers.

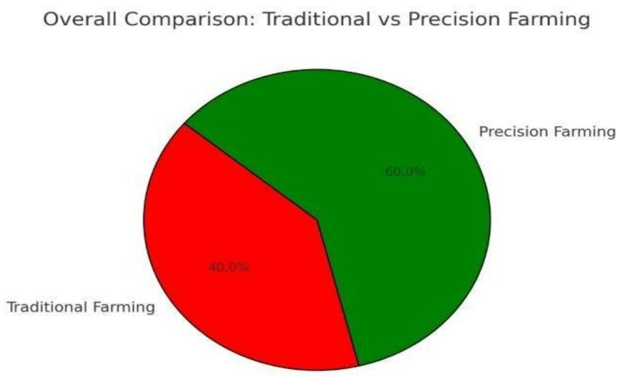

Fig 2: Traditional farming vs precision farming

Machine learning and image processing analyse leaf pictures and reliably identify disease patterns to deliver responses. An AI-based solution can monitor in real-time and provide farmers with timely feedback without expert help. This study develops a machine learning-based algorithm to detect and classify diseases in different plants. It uses segmentation, colour transformation, and texture analysis to extract features and train a model in MATLAB for disease classification. Comparison between models trained by RGB and edge detection features are done to know which model should be used for integration. The method is tested using a Kaggle dataset [3], that includes images from various stages of growth in an effort to provide farmers and agricultural professionals with an easy-to-use tool [4].

2 Related Work

All nations are some or the other way dependent on agriculture. The goal of agricultural research is to produce crops with higher quality and quantity at lower cost and higher profit. Plant diseases may cause the agricultural product's quality to deteriorate. Pathogens, which include bacteria, viruses, and fungi, are the cause of these illnesses. Thus, identifying and categorizing plant diseases at an early stage is a crucial undertaking. Experts must constantly supervise farmers, which can be excessively costly and time-consuming. Utilizing image processing and some automatic classification methods, numerous systems have been proposed to either solve or at least lessen the issues, depending on the applications [5].

Table 1: List of some of the existing disease detection methods

Author and Year	Classification Methods	Plant	Features Extracted	Result
Mokhtar Ali (2016) [6]	SVM with Cauchy kernel, Invmult Kernel and Laplacian Kernel	Tomato	Colour, Texture, Shape-Based, Statistical, Frequency domain, Morphological.	99.5 %.
Singh and Mishra (2016) [7]	Minimum Distance Criteria with K mean and GA, SVM with GA	Banana, Beans, Lemon, Rose	Colour, Texture, Shape-Based, Statistical, Frequency domain, Morphological, GA Feature Selection, K-Mean Clustering.	86.54, 93.63% and 95.71%
Jayamoorthy and Palanivel (2017) [8]	Spatial Fuzzy C Mean (SFCM), SVM	Plants	Colour, Texture, Shape, Statistical, Frequency domain, spatial, SFCM Specific Features, SVM Classification	Better accuracy
Rothe and Ksirsagar (2015)[9]	Neuro-Fuzzy Inference System	Cotton	Colour, Texture, Shape, Morphological, Spatial, Statistical, Neuro-Fuzzy, Dynamic Learning.	85%
Rastogi et al. (2015) [10]	ANN, Fuzzy logic	Plant	Colour, Texture, Shape, Statistical, Morphological, ANN-Specific Features, Fuzzy Logic Specific Features, Combined ANN and Fuzzy Logic.	Good
Yadav and Verma (2016) [11]	BPNN and GA	Tomato	Colour, Texture, Shape, Statistical, Morphological, BPNN specific Features, GA specific features.	Better accuracy
Suresha et al. (2017) [12]	KNN	Plant	Colour, Texture, Shape, Statistical, Morphological, Fourier Transform, Statistical Metrics.	76.59%

Hossain et al (2019) [13]	KNN	Plant	Colour, Texture, Shape, Statistical, Morphological, Fourier Transform, Statistical Metrics.	96.76%
Abdhulridha et al (2019)[14]	KNN, MLP	Avocado	Colour, Texture, Shape, Statistical, Morphological, KNN and MLP specific features.	MLP achieved better accuracy than KNN
Hasan et al. (2020) [15]	Deep Belief Networks (DBNs)	Multiple plant types	Colour, Texture, Shape, Statistical, Morphological, high-level features by DBNs, Contextual and Spatial features.	96% - 97.5%
Lee et al. (2021) [16]	Deep Denoising Autoencoders (DDA)	Multiple plant types	Colour, Texture, shape, statistical, Morphological, Deep features by DDAs, Contextual and spatial features.	98.3%
Shoaib et al. (2022) [17]	Convolutional Neural Networks (CNNs)	Tomato	Colour, textures, shape, structural, pattern recognition, statistical, high level abstract features.	99-99.2%
Vadivel and Suguna (2022) [18]	Fast Enhanced Learning	Tomato	Colour, texture, shape, morphological, statistical, pattern recognition, high level abstract features.	Improved accuracy, exact percentage not specified.
Varur et al. (2023) [19]	Deep Learning on embedded platforms	Chili	Colour, texture, shape, morphological, statistical, pattern recognition, high level, abstract features.	Better detection accuracy using embedded systems.
Wani et al. (2022) [20]	Machine Learning and Deep Learning methods	Multiple plants	Colour, texture, shape, statistical, morphological, high level abstract features in DL, Temporal features, Ensemble features.	CNN was identified as the most accurate method.

3 Mathematical Background

To determine the disease present in a leaf the features of the leaf are extracted from the image provided. The following are the methods used to extract the features from the image.

3.1 Red Green Blue (RGB) Model

The process of identifying and analysing diseases by using the colour information present in plant leaves is called "colour feature detection of plant diseases". The colour of leaves plays a great role in identifying plant diseases as it changes with disease, or deficiency of any nutrient. RGB is a simple method to expresses colour information of images which are represented by red, green, and blue channels. This colour information is used identify plant disease. The values are extracted in the form of mean and standard deviation [21]. Fig 3 shows the red green and blue channels of the apple leaf image which is taken from Kaggle dataset [3].

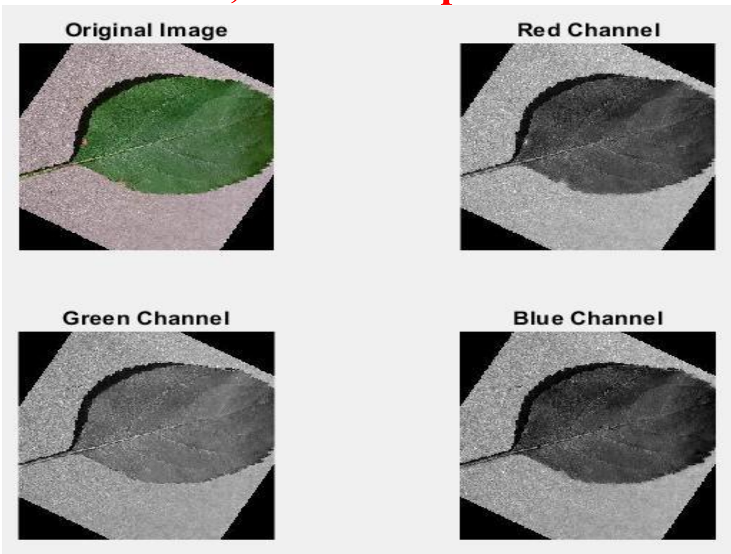

Fig 3: RGB analysis on apple leaf

The values are calculated by the following formulas.

$$\text{Mean}=\frac{1}{}\sum_{=1}^{N} \quad \ldots\ldots\ldots\ldots (1)$$

Where N is the number of pixels in an image Where X is the colour channel (Mean is calculated for red green and blue channels).

$$\text{Standard Deviation}=\sqrt{\frac{1}{}\sum_{=1}(Xi - X)} \ldots\ldots\ldots\ldots\ldots\ldots(2)$$

Where N is the number of pixels in an image. Where X is the colour channel (standard deviation is calculated for red, green and blue channels).

3.2 Edge Detection Model

Boundaries present in an image can be determined using edge detection method. It aids in simplification of image data by lowering the quantity of the data that needs to be processed while maintaining its structural characteristics. This simplification is important for tasks like object recognition, segmentation, and picture enhancement. Fig 4 shows edge detection analysis of the apple leaf image which is taken from Kaggle datasets[3].

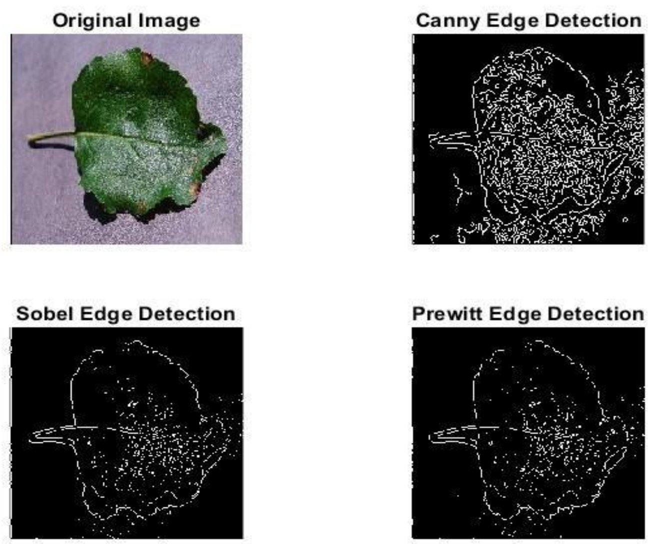

Fig 4: Edge analysis on apple leaf

The following parameters are calculated in edge detection method.

$$\text{Total edges} = \sum_{=1} \sum_{=1}^{H} E(i, j) \qquad \text{............. (3)}$$

where H, W are the height and width of the image E (I, j): Binary edge image after applying the Canny filter (1 for edge pixels, 0 for non- edge pixels).

$$\text{Edge density} = \frac{}{} \qquad \text{............. (4)}$$

A higher value means the image has more detected edges relative to its size.

$$\text{Mean Edge density} = \frac{1}{} \sum_{(i,j)\epsilon E} I(i, j) \qquad \text{............. (5)}$$

where (I, j) is the Intensity of the grayscale image position E is Set of all edge pixels.

4 Performance Indices

The performance of the trained models can be measured using the following methods

4.1 Confusion Matrix

Confusion matrix is used to ensure how efficiently a model is performing by comparing the values predicted by model to the actual results. which helps to know if the model can be further used. The confusion matrix has, True Positive: The model correctly predicted a positive outcome, True Negative: The model correctly predicted a negative outcome, False Positive: The model incorrectly predicted a positive outcome. Also known as a Type I error, False Negative: The model incorrectly predicted a negative outcome. Also known as a Type II error. The features like accuracy precision recall F1-score and specificity can be determined using confusion matrix by the following formulas [22].

1. Accuracy= $\frac{+}{+\;+\;+}$, ... (6)

   Where, the model's total accuracy indicates how frequently its forecasts come true.
2. Precision (Positive Predictive Value) = $\frac{}{+}$ .... (7)

   It measures how many of the predicted labels are actually correct.
3. Recall (Sensitivity or True Positive Rate) = $\frac{}{+}$ .... (8)

   It measures how many actual class instances were correctly classified.
4. Specificity (True Negative Rate) = $\frac{}{+}$ .... (9)

   It measures how well the model avoids false alarms.

4.2 Receiver Operating Characteristic Curve

A receiver operating characteristic (ROC) curve is a graph that shows how well a machine learning model can distinguish between classes. It is an important measure of model accuracy. The ROC curve displays the trade-off between a classifier's overall performance and true positives and false positives. Assessing the effectiveness of various classification models is beneficial. It compares the performance of two networks and makes the process of selecting the optimal threshold for a particular model simpler [23].

5 Methodology

Fig 5 shows the workflow of the proposed model, firstly image dataset of various plants is taken from Kaggle[3], then features are extracted from the plant leaf images using RGB and Edge detection methods at last models are trained using machine learning toolbox by providing the features extracted by the images.

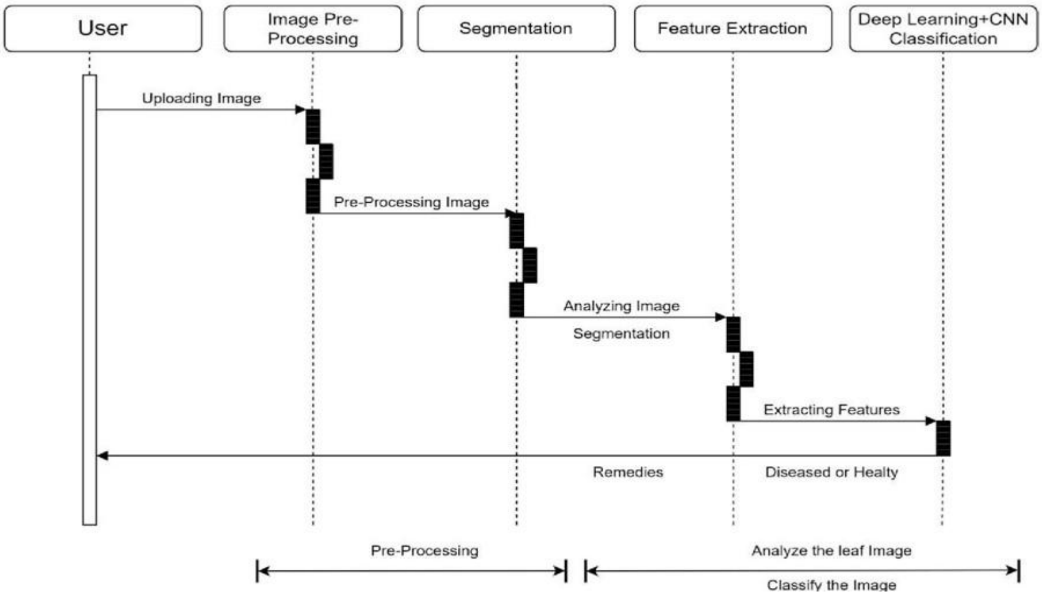

Fig 5: Workflow of proposed model

5.1 Image Acquisition from Dataset

Various plant species and disease kinds at different growth stages are covered by the dataset of photos of healthy and diseased plant leaves [3].The dataset comprising photos is downloaded in.jpg format so that it may be loaded into MATLAB for additional training and processing. Before training the model, we have to subdivide the dataset images in test, train and valid dataset with ratio 70:15:15 respectively. In this paper, leaves of various plant and fruits are taken into consideration. It includes potato, tomato, rice, wheat, strawberry, grape, bell pepper, corn, cherry, apple, peach, dataset of the following plants is taken from Kaggle[3]. Fig 6 and 7 shows images present in dataset of apple leaves.

Fig 6: datasets of apple leaves affected by disease

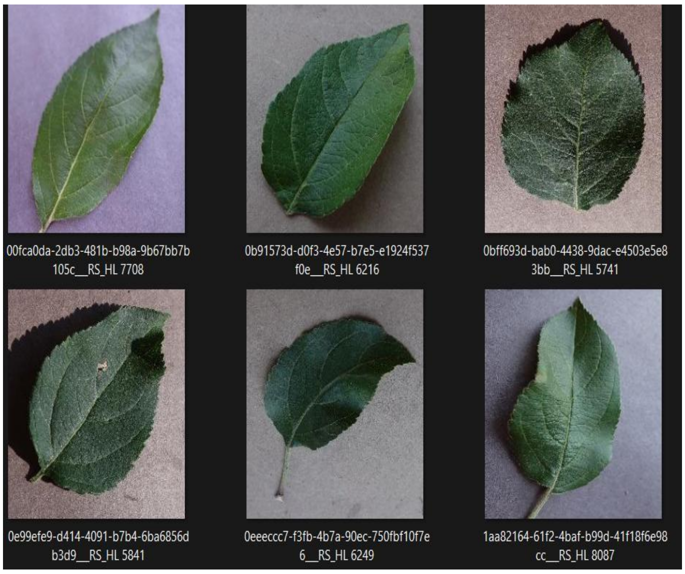

Fig 7: datasets of healthy leaves of apple

The code below is used to load the image dataset and display the classes in MATLAB

```
% load_dataset.m dataDir ='C:\Desktop\dataset\Apple';
 % Specify the path to the dataset folder imds = imageDatastore(dataDir, ...
% Load images using imageDatastore
  'IncludeSubfolders', true, ...          'LabelSource',
'foldernames'); disp("Classes in the dataset:");
% Display class labels disp(categories(imds.Labels));
```

5.2 Extraction of Image Features

The proposed model extract RGB features, mean and standard deviation of red, green and blue channel. Table 2 shows features, R_mean, R_std, G_mean, G_std, B_mean and B_std.

Table 2: Feature extracted by RGB model from apple dataset

R_Mean	R_Std	G_Mean	G_Std	B_Mean	B_Std	Label
112.0606864	59.83683448	117.5185547	48.91878441	99.24463887	61.24310678	Apple Black Rot
90.16193001	66.22293121	97.29888791	60.39078297	78.87521923	64.69022451	Apple Black Rot
119.3953683	41.62844134	129.4378787	33.94347022	124.2407526	50.09470845	Apple Black Rot
118.9206792	42.29904837	128.9892578	34.96269696	123.7339365	50.41321828	Apple Black Rot
97.99864477	57.3823821	108.9568917	56.68074348	99.96785316	62.89026565	Apple Black Rot
150.751694	57.12661788	157.7354313	42.37530536	150.6351443	62.53746366	Apple Black Rot
115.26704	54.36089646	124.3797034	48.3517336	113.5703524	59.82262499	Apple Black Rot
92.56582828	63.56352041	103.3215083	63.06003868	89.69941805	66.47446803	Apple Black Rot
107.4595026	52.10080332	120.6878587	40.62413548	104.2927894	59.16138563	Apple Black Rot

In Table 2, each row represents the values extracted from the different images present in the data, mean values indicates the average colour intensity, which helps in identifying dominant colours and standard deviation shows how much the colour intensities vary in the image. here higher std values indicates more variation in colour which could be sign of spots, discoloration, or disease patch. healthy leaf will have moderate colour intensity and variation which will result in lower value of mean and std.

Edge detection-based feature extraction, extracting structural and texture-based features, are taken into consideration. Edge detection extracts Edge density and mean edge density of plant leaf image.

Table 3: Features extracted by edge detection model from apple dataset

TotalEdges	EdgeDensity	MeanEdgeIntensity	Labels
11722	0.178863525	126.7635216	Apple Black Rot
10231	0.156112671	120.2097547	Apple Black Rot
10505	0.160293579	120.4123751	Apple Black Rot
11538	0.176055908	121.9981799	Apple Black Rot
9814	0.149749756	117.4948033	Apple Black Rot
9508	0.145080566	141.3884098	Apple Black Rot
10417	0.158950806	114.8419891	Apple Black Rot
9747	0.148727417	112.5717657	Apple Black Rot

Table 3 shows the features extracted from apple leaves, here the features extracted are total edges, edge density and mean edge density, which helps in examining the health of the plant. Total edges indicate the sharp transitions present, the proportion of the image occupied by edges are extracted using edge density, mean edge density is calculated for average intensity of the detected images. Healthy leaves have fewer edges as they are smoother and more uniform. High edge density indicate that the leaf might be heavily infected or damaged. A high mean edge intensity indicates the leaf is infected.
Similarly, features are extracted from the following plants corn, tomato, grapes, wheat, rice, cherry, bell pepper, potato, strawberry, peach.

5.3 Training of Model

Models are trained using pattern recognition method present in deep learning toolbox of MTLAB. The factors which determine the efficiency of the trained models are ROC curve and confusion matrix, Fig 8 and 9 shows the confusion matrix of models trained by the features extracted from leaf images of apple by RGB and Edge detection method.

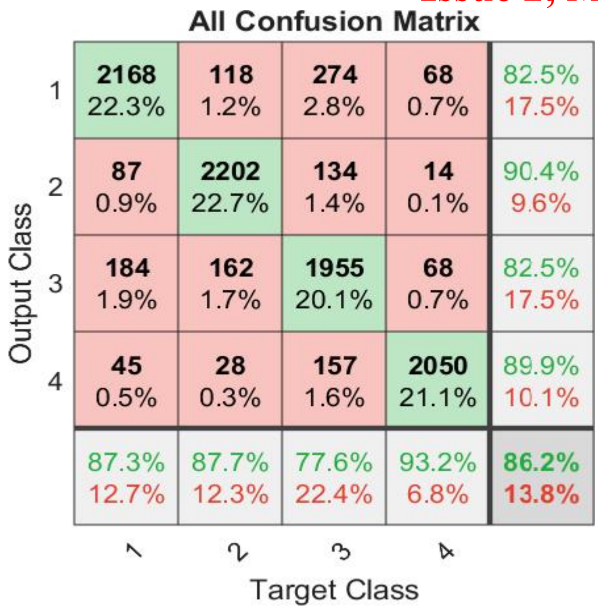

Fig 8: Confusion matrix of model trained by features extracted from RGB model

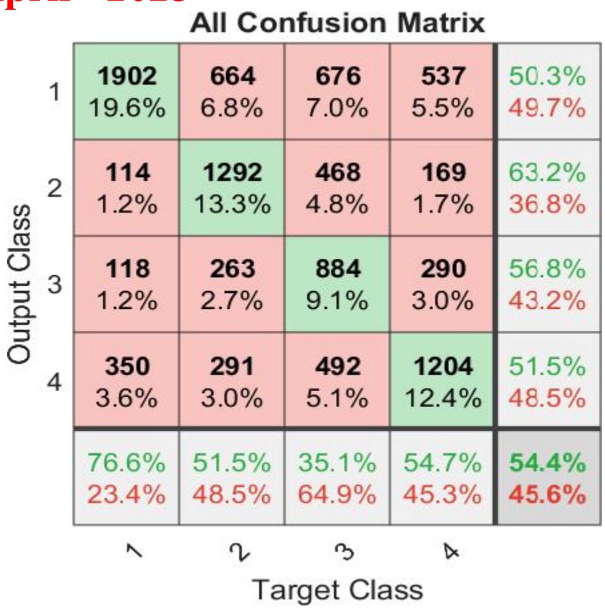

Fig 9: Confusion matrix of model trained by features extracted from edge detection model

In Fig 8, class 1 has accuracy of 87.3% which suggest that most samples were correctly classified and there is a significant misclassification rate of 12.7%, similarly class 2, 3, 4 have accuracy of 87.7%, 77.6%, 93.2% respectively and misclassification rate of 12.3%, 22.4%, 6.8% respectively. In Fig 9 the accuracy of model trained by features extracted from the edge detection method are less accurate for apple leaves compared to the model trained by RGB features, here the accuracy of class1 is 76.6% which is less than the RGB model and misclassification rate is 23.4% similarly the accuracy of class 2, 3, 4 for edge detection are 51.5%, 35.1%, 24.7% which is less than the accuracy obtained by RGB.

Another parameter to measure the performance of the trained model is ROC curve, given below is the roc curve of the trained model obtained from RGB and edge detection features extracted from apple leaves dataset.

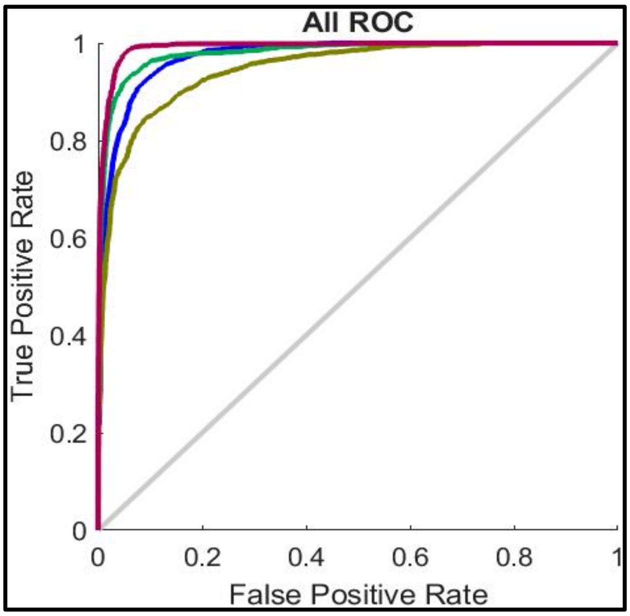

Fig 10: ROC curve of model trained by features extracted from RGB method

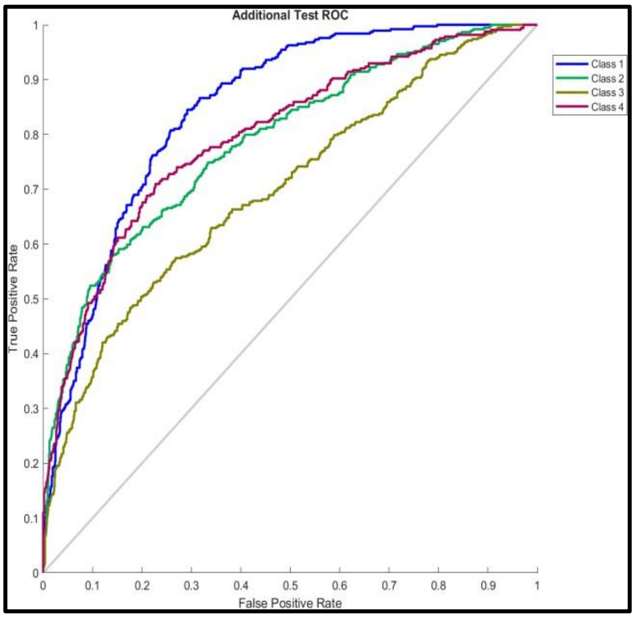

Fig 11: ROC curve of model trained by features extracted from edge detection method

The curves of the various classes in Fig 10 with their corresponding colours touching the upper left corner (0,1) show a high true positive rate and a low false positive rate, indicating that the model is doing a good job of differentiating between classes. A large area under the curve indicates better classification performance. In

Fig 11 smaller area under the curve indicates that the model trained using the edge detection method features is less successful at classifying the photos into the appropriate category. It concludes that model trained from RGB features is more accurate than model trained by edge detection features.

Table 4: Comparison between RGB and Edge detection model on the basis of accuracy in percentage

Plants	RGB model (%)	Edge detection model (%)
Apple	86.2	54.4
Bell pepper	91.4	65.4
Corn	93.9	95.6
Cherry	99.9	95.1
Grape	90.0	92.8
Peach	97.8	77.6
Potato	96.6	67.7
Rice	86.3	88.7
Strawberry	98.3	99.8
Tomato	85.9	92.8
Wheat	93.6	91.2

Table 4 shows accuracy obtained by the trained models of different plants the first column contains the names of the plants, second column have the accuracy obtained by the model trained by the features extracted from edge detection by the plant leaf images, similarly last column shows the accuracy obtained by model trained from RGB features extracted by plant leaf images. The model trained by features extracted from apple dataset from edge detection method is least accurate among all the models (highlighted by red colour), model trained by features extracted from cherry dataset from RGB method is most accurate among all the models (highlighted by green colour).

6 Advantages

It helps in early detection of plant leaf disease which prevents damage and also increase the yield of the crop. Automation in disease detection reduce the need of manual labour and experts. Since the image of the plant leaf is uploaded, farmers don not have to remove plant parts.

7 Limitations

Image quality is affected by environmental factors and lighting which might hinder the detection of disease. Disease having similar visual symptoms can be difficult to detect. It may take time for the farmers to adapt to new technology. Developing and maintaining an accurate model requires research which might not be cost effective.

8 Conclusion

Major source of food for human beings are plants and as we know large number of people in India are still dependent on farming, there is a need of early detection in plant leaf disease. The main aim is to introduce technology in agricultural field, automation in plant disease detection.

Experimental result shows that models trained by edge detection features are more accurate for plants having sharp edges whereas RGB model is effective for all plant leaves. The colour of leaves plays a great role in identifying plant diseases as it changes with disease, or deficiency of any nutrient. Models trained using RGB features are more efficient and accurate as compared to the models trained by edge detection features.

9 References

[1] Arivazhagan S., Newlin Shebia R., (2013). *Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features.* Agricultural engineering institute: CIGR journal, Volume 15, N0.1. pp. 211-217.

[2] Sarkar, C., Gupta, D., Gupta, U., & Hazarika, B. B. (2023). *Leaf disease detection using machine*

[3] *learning and deep learning: Review and challenges*. Applied Soft Computing, Volume 145.

[4] https://www.kaggle.com/datasets/emmarex/plantdisease .

[5] Jafar, A., Bibi, N., Naqvi, R. A., Sadeghi-Niaraki, A., & Jeong, D. (2024). *Revolutionizing agriculture with artificial intelligence: plant disease detection methods, applications, and their limitations*. Frontiers in Plant Science.

[6] Hrishikesh Kanjalkar P. and Prof. Lokhande S., (2013). *Detection and Clarification of Plant Leaf Diseases using ANN*. International Journal of Scientific and Engineering Research. ISSN: 2229 5518

[7] Mokhtar, U., Ali, M.A., Hassenian, A.E. and Hefny, H., (2015). *Tomato leaves diseases detection approach based on support vector machines*. 11th International computer engineering conference (ICENCO) pp. 246-250. IEEE.

[8] Singh, V. and Misra, A.K., (2017). *Detection of plant leaf diseases using image segmentation and soft computing techniques*. Information processing in Agriculture, 4(1), pp.41-49.

[9] Jayamoorthy, S. and Palanivel, N., (2017). *Identification of Leaf Disease Using Fuzzy C-MEAN and Kernal Fuzzy C-MEAN and Suggesting the Pesticides*. International Journal of Advanced Research in Science, Engineering and Technology, 4(5), pp.3852-3855.

[10] Rothe, P.R. and Kshirsagar, R.V., (2015). *Cotton leaf disease identification using pattern recognition techniques*. International conference on pervasive computing (ICPC), pp. 1-6. IEEE.

[11] Rastogi, A., Arora, R. and Sharma, S., (2015). *Leaf disease detection and grading using computer vision technology & fuzzy logic*. 2nd International conference on signal processing and integrated networks (SPIN), pp. 500-505. IEEE.

[12] Mamta Yadav, M.Y. and Toran Verma, T.V., (2016). *Hybrid approach of neural network and genetic algorithm to recognize Black Mold disease in tomato*. International journal for research in applied science and engineering technology (IJRASET), pp. 269-274.

[13] Suresha, M., Shreekanth, K.N. and Thirumalesh, B.V., (2017). *Recognition of diseases in paddy leaves using KNN classifier*. 2nd International Conference for Convergence in Technology (I2CT), pp. 663-666. IEEE.

[14] Hossain, E., Hossain, M.F. and Rahaman, M.A., (2019). *A colour and texture-based approach for the detection and classification of plant leaf disease using KNN classifier*. International conference on electrical, computer and communication engineering (ECCE), pp. 1-6. IEEE.

[15] Abdulridha, J., Ehsani, R., Abd-Elrahman, A. and Ampatzidis, Y., (2019). *A remote sensing technique for detecting laurel wilt disease in avocado in presence of other biotic and abiotic stresses*. Computers and electronics in agriculture, 156, pp.549-557.

[16] Hasan, R.I., Yusuf, S.M. and Alzubaidi, L., (2020). *Review of the state of the art of deep learning for plant diseases: A broad analysis and discussion*. Plants, 9(10), p.1302.

[17] Lee, W.H., Ozger, M., Challita, U. and Sung, K.W., (2021). *Noise learning-based denoising autoencoder. IEEE Communications Letters*, 25(9), pp.2983-2987.

[18] Shoaib, M. and Sayed, N., (2022). *YOLO Object Detector and Inception-V3 Convolutional Neural Network for Improved Brain Tumour Segmentation*. International information and engineering technology association (IIETA), pp. 371-380.

[19] Vadivel, T. and Suguna, R., (2022). *RETRACTED ARTICLE: Automatic recognition of tomato leaf disease using fast enhanced learning with image processing*. Acta Agriculturae Scandinavica, Section B-Soil & Plant Science, 72(1), pp.312-324.

[20] Varur, S., Mugad, A., Kinagi, A., Shanbhag, A., Hiremath, K. and Kulkarni, U., (2023). *Identification of anthracnose in chillies using deep learning on embedded platforms*. 3rd International Conference on Smart Data Intelligence (ICSMDI), pp. 350-356. IEEE.

[21] Wani, J.A., Sharma, S., Muzamil, M., Ahmed, S., Sharma, S. and Singh, S., (2022). *Machine learning and deep learning based computational techniques in automatic agricultural diseases detection: Methodologies, applications, and challenges*. Archives of Computational methods in Engineering, 29(1), pp.641-677.

[22] Javidan, S. M., Banakar, A., Rahnama, K., Vakilian, K. A., & Ampatzidis, Y. (2024). *Feature engineering to identify plant diseases using image processing and artificial intelligence: A comprehensive review. Smart Agricultural Technology*.

[23] Ahmed, I., & Yadav, P. K. (2023). *A systematic analysis of machine learning and deep learning-based approaches for identifying and diagnosing plant diseases. Sustainable Operations and Computers*. pp.96–104.

[24] GeeksforGeeks. (2025). *AUC ROC curve in machine learning*. GeeksforGeeks.