**Journal of Information Systems Engineering and Management**
2025, 10(2s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

# SVM-RBN Model with Attentive Feature Culling Method for Early Detection of Fruit Plant Diseases

Piyush Sharma[1], Devi Prasad Sharma[2], Sulabh Bansal[3]
[1]*Department of Information Technology, Manipal University Jaipur, India*
*piyush9.sharma@gmail.com*
[2]*Department of Information Technology, Manipal University Jaipur, India*
*deviprasad.sharma@jaipur.manipal.edu*
[3]*Department of Information Technology, Manipal University Jaipur, India*
*sulabh.bansal@jaipur.manipal.edu*

**ARTICLE INFO**

Received: 02 Oct 2024

Revised: 01 Dec 2024

Accepted: 14 Dec 2024

**ABSTRACT**

Accurate disease identification and early disease management strategies are required in India to achieve high production standards and good quality in fruits and vegetables. Image-based evaluations approaches have evolved nowadays as a result of technical developments. However, producing the wrong decision may have a negative impact on productivity. Thus, this study offered hybrid SVM-RBN model with attentive feature culling method for automatically recognizing diseases in apple fruits with high accuracy. As a result, the model generated more effective outcomes with 96% accuracy, 99% precision, 94% recall, and 93% F1 Score. Thus, by employing this technology, one may detect fruit plant illnesses at an early stage, thereby increasing fruit yield.

**Keywords:** Machine Learning, Crop Disease, hybrid SVM-RBN model, leaf diseases, Bacteria, Fungi

## INTRODUCTION

Agriculture has been the foundation of all civilizations. The focus is on expanding productivity, ignoring the environmental consequences that have become obvious in the environment's deterioration. Plant diseases, which can damage both the quality and quantity of plants, can particularly have an impact on the development of agriculture. In general, plant ailments are brought on by fungi, bacteria, viruses, mould, etc. Typically, farmers or experts can identify plant illnesses only by looking at them [1].

The farmer in rural areas could believe it's difficult to distinguish the disease that might be present in their harvests. They cannot easily visit the agricultural office to find out what the illness may be. Our main goal is to identify the disease introduced in a plant by observing its shape using image processing and machine learning [2]. Pests and diseases cause crops or parts of plants to be destroyed, which lowers food output and increases food insecurity. Additionally, little is known about illnesses and pest management or control in many less developed nations. One of the main causes of declining food production is the presence of toxic pathogens, inadequate disease control, and dramatic climate change [3].

The traditional method of identifying plant diseases involves direct eye observation and memory, but these techniques are time-consuming and imprecise. The most recent approaches require laboratory tests, knowledgeable personnel, and well-equipped laboratories, which are not always accessible in rural regions. Automatic disease detection is advantageous because it lessens the laborious task of keeping an eye on vast agricultural farms and identifies disease signs at an early stage [4]. The primary method for identifying plant disorders is a trained professional's naked eye inspection. The symptoms that sick plants present might vary, which can result in a misdiagnosis. An automated system created to diagnose plant illnesses by the look of the plant and visual symptoms could be used as a verification method in disease diagnostics [5].

AI is the development of artificial intelligence based on prior learning experiences. Modern inventions are produced by computerising machine activities using machine learning (ML) approaches. Machine learning is a numerical method for building intelligent machines. Based on information on water pressure, supplement content, harvest photographs, atmosphere, and soil moisture content, AI assists in the prediction of illness and its treatment [6]. Since the plant disease really impacts harvest output and hence lessens the creation nature of products, it represents a severe danger to global food security. Testing is used to accurately and completely diagnose yield illness. The regular arranging of plant disease differentiation evidence requires human input [7]. Plant diseases are discovered by visually inspecting plants and by using a rancher's knowledge, instinct, and thinking. In modern times, picture-based diagnostic methodologies have emerged as a result of technological advancements and the cost of image acquisition falling. Making the wrong decision or hesitating might hurt profitability [8]. The PC architecture, however, found it challenging to process an image because of the volume of data it carried.

In this proposal, we propose to build a hybrid machine learning technique to more reliably and precisely identify and categorise the plant disease.

### *1.1 Organization of the Paper*

The paper in this research project is structured as follows: Introduction, review of the literature, suggested technique, findings, explanation, and conclusion.

## LITERATURE SURVEY

Hafiz Tayyab Rauf et al [9] Due to its high vitamin C content and other essential nutritional components, citrus is a huge plant that is grown in tropical regions of the world. Citrus illnesses have impacted it, lowering the fruit's quality and harming the growers' bottom lines. To identify and categorise plant illnesses, image processing and computer vision algorithms were used, and in Pakistan's Sargodha region, citrus berries, leaves, and stems were photographed in December from orchards.

S.M. Jaisakthi et al [10] an autonomous method based on image processing and machine learning approaches has been proposed to detect these disorders. The system uses grab cut segmentation to separate the leaf from the background picture, and global thresholding and semi-supervised methodology to segregate the sick zone. To distinguish between healthy, rot, and leaf blight, machine learning techniques including Support Vector Machine (SVM), adaboost, and Random Forest tree are applied. Using SVM, we increased our testing precision to 93%.

Divyanshu Varshney et al [11] Agriculture is an important source of income in India, but is seriously affected by different illnesses. Machine Learning is used to identify impacted leaf photos and different machine learning algorithms are employed to determine whether a plant is afflicted with a disease. The aim of this endeavour has the following goal, to detect plant illnesses and the name of the fertiliser to be applied using image analysis. The bugs and insects responsible for the epidemic are also discussed.

Abu Sarwar Zamani et al [12] To identify infected leaf disease photographs, Precision Agriculture's automated leaf disease detection system employs picture capture, image processing, image segmentation, feature extraction, and machine learning techniques. As an input, a leaf picture is utilised, and leaf shots are pre-processed to reduce noise. The image is divided into sections using the K-Means algorithm, the features are retrieved using principal component analysis, and the classification of the pictures is done using methods like RBF-SVM, SVM, random forest, and ID3.

Md Rahmat Ullah et al [13] This project uses Convolutional Neural Network (CNN) as a training method to compare leaf physiognomy changes to their real color, shape, structure, and so on. This will help identify and differentiate plant diseases such as aster yellows, bacterial wilt, scab, and more.

Deepa et al [14] Plant diseases are a major cause of decreased yield and farmer revenue, and Researchers are striving to create a system that will automatically recognise them. This research aims to build a unique strategy using machine learning techniques, which may help identify a cure as soon as possible.

## PROPOSED METHADOLOGY

Plant diseases have a detrimental effect on agricultural production, leading to food insecurity. Diseases including Apple Scab, Apple Rot, Apple Blotch, Cedar Apple Rust, and Powdery Mildew are particularly harmful to apple plants. To identify these diseases, a quick and effective recognition system must be created. There are conventional approaches for diagnosing diseases, but the image processing strategy must be a non-invasive way that can give farmers an accurate, economical, and dependable answer. Early diagnosis is essential for the effective prevention and control of plant diseases, and they can be a major factor in managing and making decisions regarding agricultural productivity.

Traditional ML models have a bias in favour of the feature with larger magnitude values, so a top-notch ML technique must be created to accurately identify diseased fruit plants and help farmers take preventative action. This study suggests a unique SVM-RBN model with attentive feature culling to address the shortcomings in the currently used conventional approach

*3.1 Ternate Impeccable Pre-Processing Technique*

In image processing, image acquisition is considered as the first step in the pre-processing phase. The acquired images are often noisy, and may also suffer from other distortions such as uneven lighting or colour balance. Therefore, pre-processing techniques are applied to enhance the quality of the acquired images before further analysis. For our proposed model we undergo the following steps for pre-processing, they are

1. Step-1= Grey scale conversion
2. Step-2= Adaptive median filter (Noise removal)

### ***3.1.1. Reading Input Image and Grey Scale Conversion***

The initial stage of picture acquisition in image processing is reading an input image. The input image can be read from various sources, such as a digital camera, scanner, file, or network. The most common way to read an input image in image processing is to load it from a file. The image file types JPEG, PNG, TIFF, BMP, and GIF are just a few that are accessible. The choice of file format depends on the application and the type of image being acquired. For our work we have taken the data from kaggle an open source data base for datset. In most programming languages, there are libraries available to read and manipulate image files. In Python, the OpenCV library is commonly used for image processing tasks.

The captured image is transformed to grey scale after collection. Image for Image Enhancement is the process of modifying a photo such that the results are more appropriate for a certain use. Examples of this subtask include sharpening or deblurring an out-of-focus image, accentuating borders, improving image contrast, or brightening an image. Each pixel ranges in grayscale from 0 (black) to 255 (white), typically. A pixel can be represented by eight bits, or exactly one byte, within this range. This range for managing picture files is quite natural. Red, Green, and Blue (RGB) images are kept as an m-by-n data array that specifies each pixel's 15 distinct red, green, and blue colour components. Pixel numbers for grey scale images vary from 0 to 255. A Mat Lab function named rgb2gray (), which functions by creating a weighted sum of the R, G, and B components, is used to convert colour images to greyscale. The fig.1 shows the conversion of Original image to Grey Scale Image.


**Figure 1.** Original Image to Grey Scale image

### ***3.1.2. Adaptive Median Filter***

An image filter known as an adaptive median filter is used to reduce noise in pictures after the original image has been transformed to a greyscale image. The filter can effectively remove various kinds of noise in various areas of the image because it alters its size and shape depending on the local characteristics of the image. The adaptive median filter analyses the pixels in the picture that fall within a predetermined window size, which is typically a square or rectangular area surrounding the pixel being filtered. The filter then calculates the local median of all the pixel values contained within the frame and contrasts it with the pixel value being filtered. The pixel value is kept unchanged if the distance between it and the local median falls within a predetermined range. If the disparity, however, is outside of the acceptable range, the pixel value is changed to the neighbourhood median value.

**Figure 2.** Image after Adaptive Filtering

The adaptive median filter uses an adaptive range, which means it adapts to the local properties of the picture. If there is little noise in the area, the range is set to a lower number, which enables the filter to preserve more of the image's details. In contrast, if there is a lot of noise in the area, the range is increased, enabling the filter to filter out more noise. Because it can adjust to various types of noise in different regions of the image, the adaptive median filter is especially helpful in scenarios where the size and intensity of the noise in the image varies. The fig.2 shows the output of the image after adaptive filter.

*3.2. Attentive Feature Culling Method*

Feature extraction and segmentation are two important steps in image processing that are often used together to analyse and interpret images. The process of locating and extracting important characteristics or patterns from a picture, such as edges, corners, textures, and forms, is known as feature extraction. These features are then used to represent the image in a more concise and informative manner for further analysis. In future extraction at first we are going to do standardization of the image. The purpose of standardizing the features is to remove any potential biases or variations caused by differences in the units, scales, or measurement methods used to obtain the features. This is important because many machine learning algorithms and statistical models rely on the assumption that the input features are independent, identically distributed, and have a similar scale and distribution. In order to simplify its representation or extract valuable information from it, segmentation is the process of separating a picture or a signal into various segments or areas. Segmentation can be used in the context of feature extraction to distinguish the object of interest from the foreground or other unimportant elements of the image.

In order to convert data so that it has a mean of zero, a standardisation technique called mean normalisation is employed in image processing and other domains. This method entails deducting the dataset's mean value from each individual data point. Data from the result will have a mean of zero and a standard deviation that represents the variance of the starting data. Mean normalisation is a common technique used in image processing to uniformly scale the intensity values of the pixels in a picture. This method can assist in removing any general brightness or contrast changes that may be present in the image as a result of varying lighting conditions or other factors. By eliminating any differences in brightness or contrast that might exist between various photos, mean normalisation can also aid in improving the comparability of images.

The mean normalisation formula is as follows:

$$x' = \frac{x-\mu}{\sigma}$$

Where x is the original data, μ is the mean value of the data, $\sigma$is the standard deviation of the data, and x′ is the normalized data.

The intensity levels of the pixels in an image can be made comparable and have similar ranges and distributions by performing mean normalisation. This can aid in enhancing the precision and reliability of image processing methods or machine learning algorithms that depend on the intensity values of pixel data.

Thresholding, one of the standardisation techniques, is employed after the image has been normalised. A technique called thresholding is used to divide a picture into various areas or objects depending on a predetermined threshold value. To distinguish the foreground items from the background, a threshold value is often chosen depending on the intensity or colour values of the pixels. There are several ways to execute thresholding, including global thresholding, adaptive thresholding, and Otsu's thresholding. We utilise Otsu's threshold standardisation in this work. The fig.3 shows the apple fruits and leaf healthy and unhealthy image for training purpose.

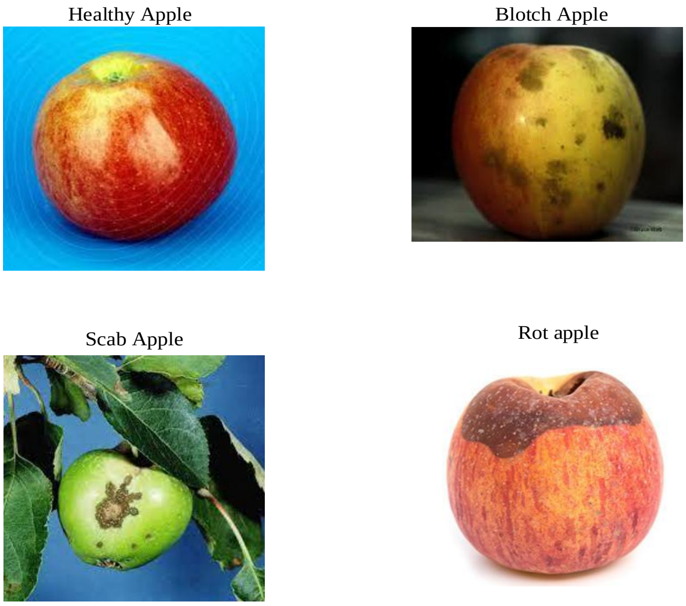


**Figure 3(a).** Healthy and Unhealthy Fruit

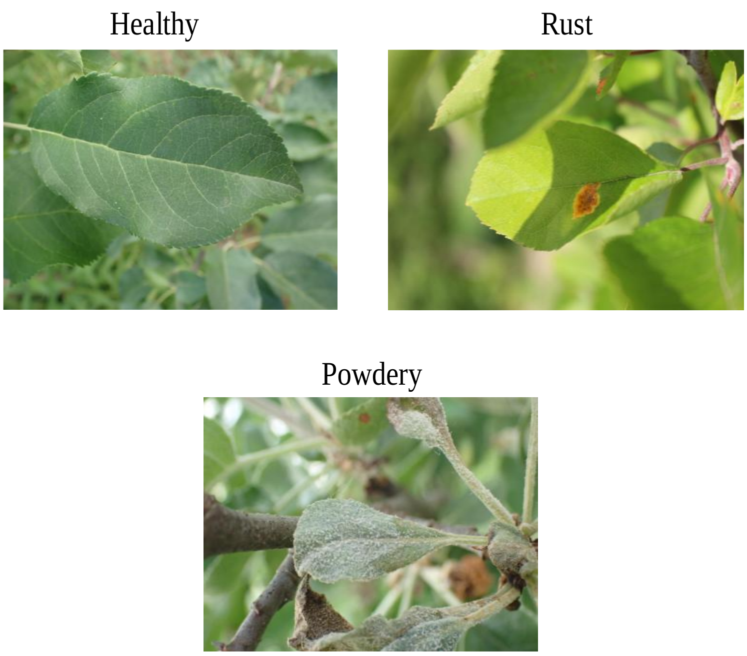


**Figure 3(b).** Healthy and Unhealthy Leaf

*3.3. SVM-RBN Classifier*

A classifier in image processing is a machine learning model that has been taught to identify and categorise different items or properties in a photograph. A label or group of labels that describe the content of the input image are generated from the input image. In image processing, classifiers are frequently employed in a wide range of applications, including autonomous cars, object detection, facial recognition, and medical imaging. The learning method involves identifying patterns and features in the photos that correlate to particular categories. They are trained using sizable datasets of tagged images. In order to identify and categorise the apple fruit and plant disease in our suggested model, we are using a hybrid SVM-RBN classifier.

The Hybrid SVM and RBN model is proposed in this research which combines the Support Vector Machines (SVM) and Radial Basis Network approaches (RBN). Support Vector Machines are a well-liked machine learning classification algorithm that work well for both linearly and non-linearly separable data. SVMs divide data into various classes using a boundary or a hyper plane. On the other hand, Radial Basis Network are a sort of dynamic system made up of connected Boolean variables whose states change in response to a set of rules or functions. RBNs have been utilised in the modelling of intricate systems including brain networks and gene regulatory networks.

The Hybrid SVM and RBN model enhances classification accuracy by combining the benefits of SVMs and RBNs. The data is first classified using an SVM in the algorithm, and the projected class labels are then supplied to an RBN. By recording the dynamic interactions between the features, the RBN helps to further enhance the classification.

## 4. Result and Discussion

*A. 4.1. Dataset*

The image of the apple fruits and leaf, which serves as input for diagnosing diseases, was collected from the Kaggle.com image repository.

*B. 4.2. Epoch Vs Accuracy Curve*

The epoch vs. accuracy graph in image processing is a diagram that displays the correlation between the quantity of training epochs and the task-specific accuracy of a machine learning model. The y-axis shows the model's accuracy, while the x-axis shows the number of epochs. The graph is often drawn during a machine learning model's training phase and is used to track the model's performance over time. The figure 4 shows the graph for epoch vs. accuracy curve.

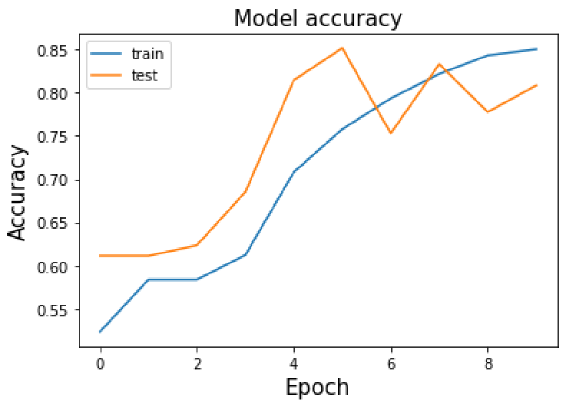


**Figure 4.** Graph between Epochs vs. Accuracy

**USING THE TEMPLATE**

*4.3. Epoch Vs Loss*

The epoch vs. loss graph in image processing is a graphic that displays the correlation between the quantity of training epochs and the failure rate of a machine learning model on a particular task. It may be used to identify the appropriate amount of training epochs and to identify possible issues, such as over- or under fitting. In general, as the loss trend decreases, the model should exhibit little loss on both the training and validation sets of data. If the loss on the training data keeps declining, the model may be over fitting and more training may not be necessary. Figure 5 shows the model Loss Graph.

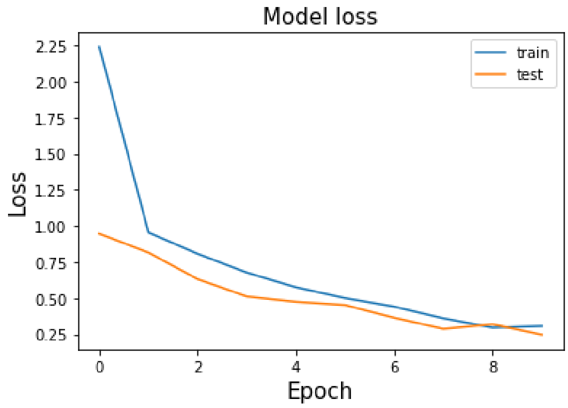


**Figure 5.** Model Loss graph

*A. 4.4. ROC for SVM and RBN Model*

A binary classification system's performance is assessed using the ROC curve. By comparing the true positive rate (TPR) and false positive rate (FPR) at various categorization criteria, it is produced. A point in the top left-hand corner of the ROC curve would represent the ideal classifier, which would have a TPR of 1 and an FPR of 0. A classifier with a ROC curve that runs diagonally from the bottom left to the top right of the plot would perform no better than random guessing. An AUC value of 0.8 or higher is generally considered to be good performance in many applications. Figure 6 shows the ROC for the SVM and RBN.

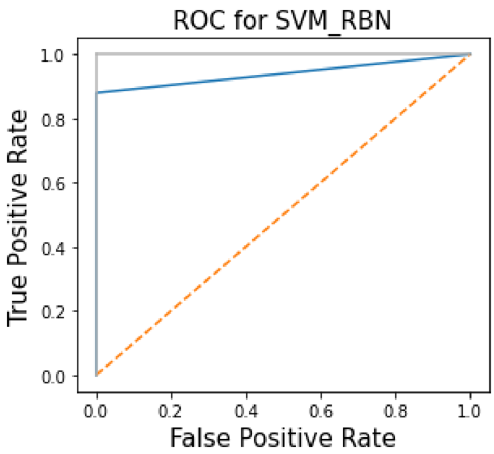


**Figure 6.**ROC for SVM RBN Model

## 5. Comparison of Proposed Model with other Model

*A. 5.1. Accuracy*

A classification algorithm's performance is gauged using a statistic called accuracy. It is determined by dividing the total number of guesses by the number of predictions that were right.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

In simpler terms, the percentage of accurate forecasts to all other guesses is known as accuracy. It is given as a percentage or a decimal number between 0 and 1, where 1 denotes the highest potential accuracy and 0 the lowest.

The accuracy of our suggested model in comparison to other models is shown in Figure 7 below.

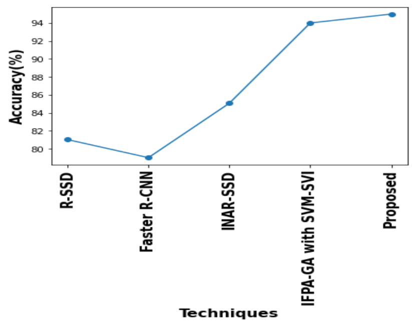


**Figure 7.** Accuracy comparison

*5.2. Precision*

A classification algorithm's precision is a statistic that assesses the proportion of accurate positive predictions among all positive predictions. In other words, it assesses the precision or accuracy of the model's optimistic forecasts.

$$\text{Precision} = \frac{TP}{TP+FP}$$

Where:

- The instances when the model accurately predicted the positive class are known as true positives (TP).
- False positives (FP) are situations in which a model predicted a positive class, but the actual class was negative.

In simpler terms, precision is the proportion of accurate positive predictions to all of the model's positive predictions. It is stated as a proportion or a decimal number between 0 and 1, with 1 being the highest degree of precision and 0 the lowest degree of precision. The figure 8 shows the precision graph with our proposed model with other model.

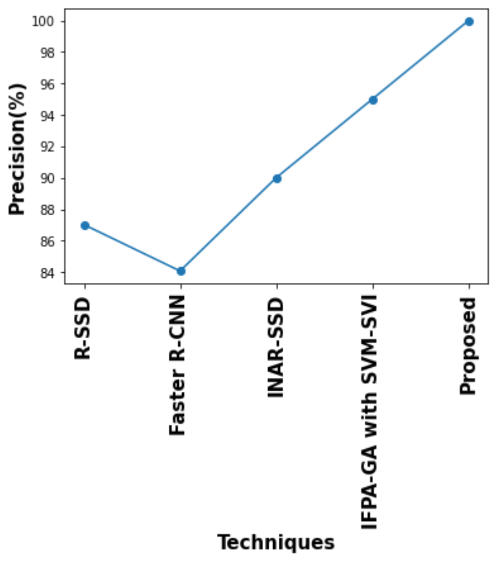


**Figure 8.** Precision graph Comparison

*5.3. Recall*

A classification problem's recall metric counts the proportion of accurate positive predictions among all real positive examples. It assesses the model's capability to recognise the positive situations, in other words.

$$\text{Recall} = \frac{TP}{TP+FN}$$

Where:

- The instances when the model accurately predicted the positive class are known as true positives (TP).
- False negatives (FN) are situations in which a model predicted a negative class, but the actual class was positive.

In simpler terms, recall is the proportion of correctly predicted positive outcomes to all of the actual positive cases in the data. A percentage or decimal number between 0 and 1 is used to indicate it, with 1 denoting perfect recall and 0 denoting the least possible recall. The figure 9 represents the recall comparison graph.

- According to the Recall formula, recall is the percentage of accurate positive predictions among all real positive cases in the data.

The F1 score is, to put it simply, the average of accuracy and recall with equal weights. It has a range of 0 to 1, with 0 being the lowest possible performance and 1 denoting flawless precision and recall. The F1 score is often used in situations where precision and recall are equally important, such as in medical diagnosis or fraud detection. The Figure 10 shows the F1 Comparison graph.

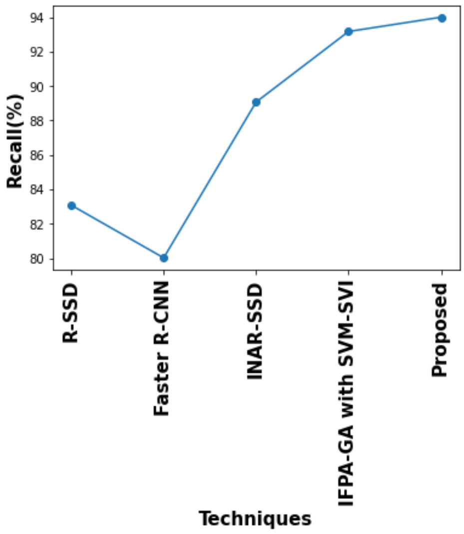


**Figure 9.** Recall comparison graph

*5.4. F-1 Score (%)*

A statistic used to assess the harmony between accuracy and recall in a classification task is the F1 score. It gives a single score that represents the model's overall effectiveness and is the harmonic mean of accuracy and recall.

$$\text{F1 score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Where:

- According to the Precision formula, precision is the proportion of accurate positive predictions among all positive predictions generated by the model.

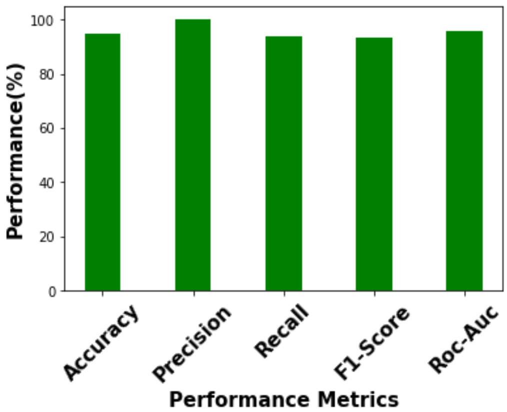


**Figure 11.** Performance metrics of the proposed model

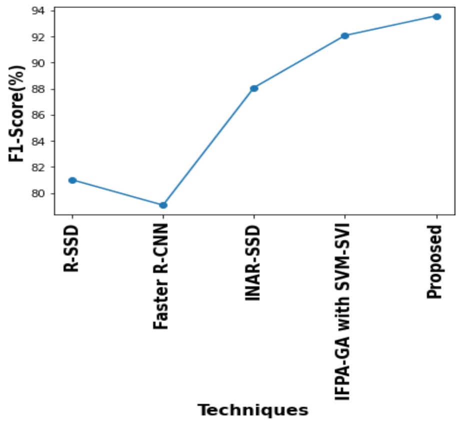


**Figure 10.** F1- Score Comparison graph

*5.5. Performance (%)*

Performance metrics are measurements that are used to assess a system or process's efficacy and efficiency. Several performance indicators are frequently used in image processing to assess how well image processing algorithms perform. The performance of the different parameter has been shown in figure 11

**Table 1** Overall comparison of performance

| Model | Accuracy (%) | Precision (%) | Recall (%) | F1- score (%) |
|---|---|---|---|---|
| R-SSD | 81 | 87 | 83 | 81 |
| Faster R-CNN | 79 | 84 | 80 | 78 |
| INAR-SSD | 85 | 90 | 89 | 89 |
| IFPA-GA with SVM-SVI | 94 | 96 | 92 | 92 |
| **Proposed SVM-RBN** | **96** | **99** | **94** | **93** |

## CONCLUSION

In this research, we propose an automated fruit disease recognition system that uses machine learning to detect illnesses in apple fruit and leaves. In order for this to occur, we used the dataset from the Kaggle website and performed a pre-processing procedure in which the raw image was collected, the noise was eliminated, and the image was then converted to a greyscale image. The second use of these pre-processed images is for feature extraction and segmentation, which involves standardising and normalising the image. These photos were transferred to the SVM-RBN classifier after the feature extraction in order to identify the disease and infection in apple fruit. Consequently, we are able to provide results that are more efficient with 96%, 99%, 94%, and 93% in accuracy, precision, recall, and F1 Score, respectively.

## REFERENCES

[1] Chen J, Chen J, Zhang D, Sun Y & Nanehkaran YA 2020 Using deep transfer learning for image-based plant disease identification *Computers and Electronics in Agriculture* **173** 105393.

[2] Barure S, Mahadik B, Thorat M & Kalal A 2020 Disease detection in plant using machine learning International Research Journal of Engineering and Technology (IRJET).

[3] Lestari AD & Nurriski YJ 2022 Application of pest detection on vegetable crops using the cnn algorithm as a smart farm innovation to realize food security in the 4.0 era Journal of Soft Computing Exploration **3**(2) 111-116.

[4] Dhaka VS, Meena SV, Rani G, Sinwar D, Ijaz MF & Woźniak M 2021 A survey of deep convolutional neural networks applied for prediction of plant leaf diseases Sensors **21**(14) 4749.

[5] Sinha A & Shekhawat RS 2020 Review of image processing approaches for detecting plant diseases IET Image Processing **14**(8) 1427-1439.

[6] Khan R, Dhingra N & Bhati N 2022 Role of artificial intelligence in agriculture: A comparative study. In Transforming Management with AI, Big-Data, and IoT Cham: Springer International Publishing 73-83.

[7] He DC, Burdon JJ, Xie LH & Jiasui ZHAN 2021 Triple bottom-line consideration of sustainable plant disease management: From economic, sociological and ecological perspectives Journal of Integrative Agriculture **20**(10) 2581-2591.

[8] Nandhini S & Ashokkumar K 2022 Machine learning technique for crop disease prediction through crop leaf image Appl. Math. **16**(2) 149-158.

[9] Rauf HT, Saleem BA, Lali MIU, Khan MA, Sharif M & Bukhari SAC 2019 A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning Data in brief **26** 104340.

[10] Jaisakthi SM, Mirunalini P & Thenmozhi D 2019, February Grape leaf disease identification using machine learning techniques In 2019 International Conference on Computational Intelligence in Data Science (ICCIDS) 1-6 IEEE.

[11] Varshney D, Babukhanwala B, Khan J, Saxena D & Singh AK 2022, May Plant disease detection using machine learning techniques In 2022 3rd International Conference for Emerging Technology (INCET) 1-5 IEEE.

[12] Zamani AS, Anand L, Rane KP, Prabhu P, Buttar AM, Pallathadka H ... & Dugbakie BN 2022 Performance of machine learning and image processing in plant leaf disease detection Journal of Food Quality **2022** 1-7.

[13] Ullah MR, Dola NA, Sattar A & Hasnat A 2019, November Plant diseases recognition using machine learning In 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART) 67-73 IEEE.

[14] Deepa RN & Shetty C 2021 A machine learning technique for identification of plant diseases in leaves In 2021 6th international conference on inventive computation technologies (ICICT) 481-484).