# Harnessing Machine Learning to Safeguard Crop Yield: Early Detection of Plant Illnesses

**Narmadha Thangarasu[1], Dr. Sumitra Padmanabhan[2], Lokesh Verma[3], Dr. Trapty Agarwal[4], Sourav Rampal[5], Dr. Ashwini Kumar[6]**

*[1]Assistant Professor, Department of Computer Science Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Karnataka - 562112, India, Email Id- narmadha.t@jainuniversity.ac.in, Orcid Id- 0000-0001-9628-4236*
*[2]Associate Professor, Department of uGDX, ATLAS SkillTech University, Mumbai, Maharashta, India, Email Id- sumitra.padmanabhan@atlasuniversity.edu.in, Orcid Id- 0000 0003 4846 080X*
*[3]Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India lokesh.verma.orp@chitkara.edu.in, https://orcid.org/0009-0009-3032-3947*
*[4]Associate Professor, Maharishi School of Engineering & Technology, Maharishi University of Information Technology, Uttar Pradesh, India, Email Id- trapty@muit.in, Orcid Id- 0009-0007-4081-4999*
*[5]Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh- 174103 India sourav.rampal.orp@chitkara.edu.in, https://orcid.org/0009-0000-6270-102X*
*[6]Assistant Professor, Department of Mechanical Engineering, ARKA JAIN University, Jamshedpur, Jharkhand, India, Email Id- dr.ashwini@arkajainuniversity.ac.in, Orcid Id- 0000-0003-3558-8054*

Modern agriculture confronts various obstacles, plant diseases which may have a substantial influence on productivity and quality. The study introduces Integrated Spider Monkey Optimized Efficient Gradient Boosting (ISMO-EGB) as a new method for identifying plant illnesses. The model uses a database of plant leaf disorders, allowing for precise classification based on distinct characteristics. A Gaussian filter is used for noise reduction and feature augmentation, while linear discriminant analysis (LDA) is used for feature extraction. The model combines spider monkey social behaviour with gradient boosting methods, enhancing its effectiveness and precision. The model's efficacy measures include precision, accuracy F1-score and recall compared to conventional techniques like RF, SVM, ANN and KNN. ISMO-EGB demonstrates a higher accuracy rate of 95.05%, demonstrating its potential as a precision agricultural technology for sustainable farming methods and improved food security.

**Keywords:** Machine Learning, Plant Diseases, Crop Yield Safeguarding, Integrated Spider Monkey Optimized Efficient Gradient Boosting (ISMO-EGB).

## 1. Introduction

To protect agricultural output and guarantee global food security, it is critical to use machine learning (ML) for early diagnosis of plant diseases. Visual inspection is a common approach used in traditional ways of diagnosing plant diseases, although it can be laborious and prone to human mistakes. Revolution is facing the agricultural industry. Digitization of agriculture depends on several breakthroughs that originate from outside the agricultural sector, including global positioning system, distributed computing, unmanned aerial vehicles and the Internet of Things (IoT) and interpreted using massive information processing techniques [1]. IoT is a crucial consideration when focusing on IoT applications in farming, a remarkable accomplishment in the realm of agriculture today. Many types of sensors have been developed to identify agricultural items. The IoT changes the plant disease is remotely monitored [2]. The use of "intelligent agriculture practices" through the use of cutting-edge IoT technology holds for improving farming methods and modernizing agriculture. The process of monitoring agricultural diseases is complex, including several phases and unique performers [3]. Problems with leaves are one of the main issues that farmers deal with. The majority of individuals do not want to farm since they do not obtain the desired results, which is another issue. The agricultural disease monitoring process is a complex that involves many different steps and unique performers. One of the main issues that farmers deal with leaves and one of the main explanations for the reason most individuals do not want to farm is that they do not receive the expected results. To obtain significant benefits for the ranchers, plant health condition plays a critical role. Valid plant health checks are necessary at different stages of plant growth to predict diseases that can affect plants [4]. Rats and disease have a major impact on yield estimation and harvest development estimation. The current architecture is dependent on unassisted visual perception, which is a laborious process. To obtain significant benefits for the ranchers, plant health condition plays a critical role. Valid plant health checks are necessary at different stages of plant growth to predict diseases that can affect plants. Rats and disease have a major impact on yield estimation and harvest development estimation. The current architecture is dependent on unassisted sight, which is a laborious process [5]. The best applications for robots seem to be in time-consuming, monotonous and effective tasks, especially in cropping environments with short harvesting periods. Automatic plant security has been investigated in relation to agricultural procedures. It can provide the most challenging test for analysts and engineers since questions about disease identification must be taken into standard robot-related problems agriculture importance to India's economy and the country farmers [6]. Many areas of India are badly afflicted by reduced yield illnesses, which lower production and force inexperienced farmers to commit suicide since they are ignorant of the disease and its late discovery. Farmers have used chemicals to avoid infections, yet the disease is becoming worse every day [7]. The toxicity levels of different foods increase illnesses. For most people in India, agriculture has been their main source of livelihood. The environment has suffered as a result of the widespread commercialization of agriculture. Plant disease identification is one of the main issues in the realm of farming. Early disease identification reduces the risk of the illness spreading to other plants and causing significant financial expenses. The economic impact of crop illnesses could vary from negligible symptoms to the total loss of the entire production of farms. Among the things causing the superiority of agricultural produce to diminish is botanical disorders. A decline in both perspectives can have an impact on a nation's total crop production. The absence of ongoing plant monitoring is the main problem. Occasionally ignorant of illnesses and the times at which they happen. Generally speaking, illnesses can affect any plant at any moment. Ongoing surveillance could aid in preventing the spread of disease. Identifying plant illnesses is the most significant research topics in agriculture [8]. The size, color and fit of disease symptoms differ like a violin. One of the main explanations for the reason of individuals color, whereas other diseases are distinct in color but similar in shape. Sometimes, farmers might become confused and incapable of making the right decisions about pesticides. These images may be sent to a targeted illness analysis system, which will diagnose the condition and offer details about it, along with possible treatments for pesticides. The core function of a system would be the automated identification of the sickness that has manifested. Anomalies in leaf development, color distortion, stunted growth and wilted and damaged units are common symptoms. Even though diseases and pests like rodents may have a significant negative influence on crops and spread to other plants, as well as human health. These need to be carefully examined and properly managed to protect the crops from significant losses. Figure 1 shows the sample image of the crop diseases. Infections may be discovered in a variety of plant components, including leaves, stems and natural products. Leaf offers some interesting differences between flowers and natural goods in any season.

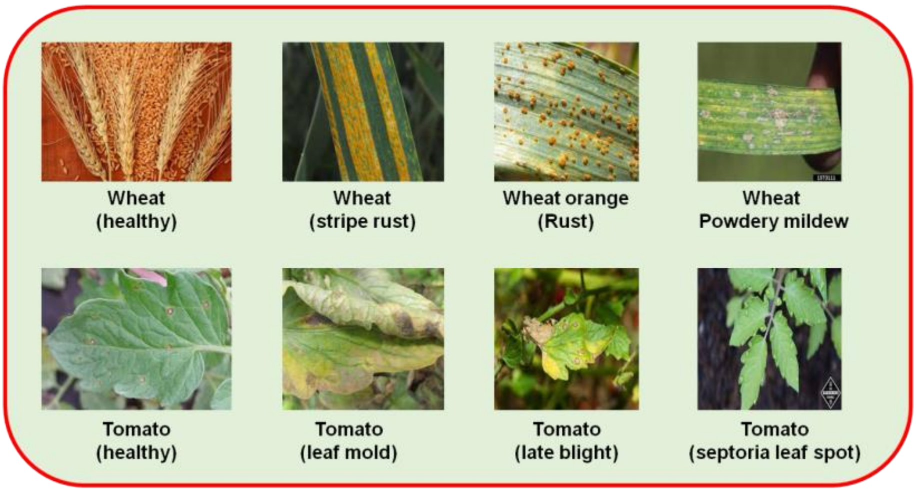


Figure 1: Sample images of crop diseases (Source: https://www.researchgate.net/figure/Sample-images-of-diseased-crops_fig1_335216374)

## 1.1 Objective of the study

The objective is to instruct ML models to identify the symptoms of plant diseases in their early stages so that preventative actions and prompt intervention may be taken. The technology uses sophisticated algorithms to offer farmers early notice of any crop diseases so they may take preventive action like crop rotation, targeted pesticide treatment, or other necessary interventions.

### 1.2 Significance of the study

To protect agricultural output and guarantee global food security, it is critical to use ML for early diagnosis of plant diseases. Visual inspection is a common approach used in traditional ways of diagnosing plant diseases, although it can be laborious and prone to human mistakes.

### 1.3 Research gap of the study

There is no automatic start-up of the cluster center, which enhances the suggested approach for reducing classification-related error, Suitable for several diseases of soybean plants. Combining cutting-edge image methods with AI algorithms is needed to identify plant leaf disease mechanically. Additional analysis employing various variations in the surrounding environment is required for the work. To identify plant diseases, a sophisticated system is needed.

### 1.4 Contributions of the study

A critical development in ensuring agricultural yield protection is the application of ML to the early identification of plant diseases. Through the use of complex algorithms, ML models are able to examine large datasets that contain a variety of factors, including weather patterns, plant traits and soil quality. The method reduces crop losses but also lessens the need for broad-spectrum treatments. By encouraging energy savings and reducing ecological impact, it contributes to responsible agricultural practices.

To contextualize and inform the study, a review of pertinent literature is included in section 2. A more detailed description of the process is given in section 3. In Section 4, a thorough examination, evaluation and discussion of the outcome are given. Section 5 provides detailed explanations that highlight the relevance of the conclusion.

## 2. Related works

According to the author of, [9] offered regardless of the region, agricultural output is a crucial requirement for economic success. It is essential because it gives various individuals access to food, jobs and raw materials. The disparity in anticipated crop output throughout different regions of the world can be attributed to several factors. Among them include the excessive use of chemical fertilizers, the soil fertility, and amount of pollutants in the water supply, the uneven distribution of rainfall and other problems. A typical obstacle experienced by people worldwide, aside from these problems, was the loss of a significant portion of productivity illness. Author [10] investigated the potential of computer vision techniques for early and scalable plant disease diagnosis in the research. Enabling vision-based plant disease diagnosis faces significant obstacles the absence of suitable large-scale non-lab data sets. They introduce PlantDoc, a dataset for the visual diagnosis of plant diseases. The suggested two techniques: SCNN-RF Shallow CNN with Random Forest [11]. Three distinct datasets were used in the comparative tests using alternative deep learning models. The outcomes demonstrate that alternative retrained deep models are not as effective. The prompt diagnosis of plant pathogens can save losses of up to 50% of global crop production [12]. Plant disease detection and diagnosis using current molecular and imaging techniques has several drawbacks. They generated the pursuit of scientific methods applicable to general materials methods with little invasiveness that it referred to apply immediately outdoors, where crop illness verified diagnosis. The ML [13] techniques focus on the results of specific tasks and primarily apply to data; they can be used to identify illnesses. The paper presents a comparison of machine learning classification algorithms for plant disease detection lays out the processes of a general plant disease detecting system. Based on the results of the study, convolutional neural networks (CNN) technology identifies more illnesses in a variety of crops with a lot of precision. Study explores a possible solution to the issue by training segments of the image data. When compared to the framework for convolutional neural network (F-CNN) [14] model trained with whole photos, the selective convolutional neural network (S-CNN) model trained using segmented photographs achieves over twice the performance at precision despite several groups when evaluated on separate information that the models had barely observed. The presented an investigation that gives an improved trained efficacy, validity and application due to the findings of a mathematical model for the identification of diseases in plants that uses DB. The region proposal network (RPN) [15] is used to detect and find leaves in complex situations. Images that have been segmented using the Chan–Vese (CV) algorithm was based on the RPN algorithm's results that have the nature of the problems. The separated leaves were sent into a machine learning system that has been pre-conditioned with data on leaf diseases from a flat landscape background. Predictive modeling, data analytics, pathogen sensors and other enhanced detection technologies were some of the new tools that were required to address these significant difficulties and stop future epidemics [16]. Here, provide comprehensive research programs that referred to able to lessen the impact of upcoming plant disease global epidemics. A description of standard methods and procedures for the identification, measurement and categorization of illnesses is provided in the study [17]. The study focuses on essential weaknesses in current practices and improves them for disease early prediction. Author [18] described the outcomes that CNNs have obtained. The article uses deep CNN models in the field of machine vision to detect and identify leaf-borne illnesses in plants. CNN models need more computer power and a larger quantity of variables. How many parameters computational costs were decreased in the article by using convolution for ordinary convolution [19].

## 3. Materials and methods

The data set that was analyzed in the paper is presented here. The section discusses the application of ML to the problem of identifying damaged crops using photographs of their leaves. The photos of plant leaves are classified as unhealthy or healthy using ML programmed using transfer learning. Figure 2 represents the Framework for the proposed integrated spider monkey optimized efficiency gradient boosting (ISMO-EGB).

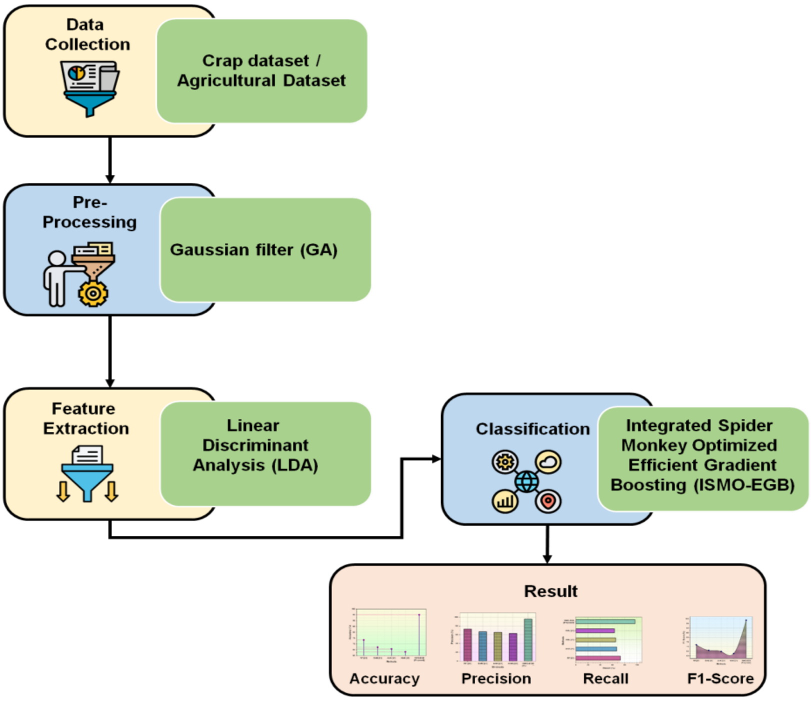


Figure 2: Framework for proposed ISMO-EGB (Source: Author)

3.1 Data collection:  Illness categorization on leaves

Plant Village, a publicly accessible dataset of 54,205 pictures of various branches, both excellent and diseased, associated with 12 distinct deep learning models was trained using a dataset consisting of several types of plants. The purpose of making the images suitable for the models' baseline values was that they were shrunk while adjusted by dividing the pixel values by 250 in Figure 3. To prevent overfitting, the dataset was split 70%, 20% and 10% into training, validation and testing datasets [20].

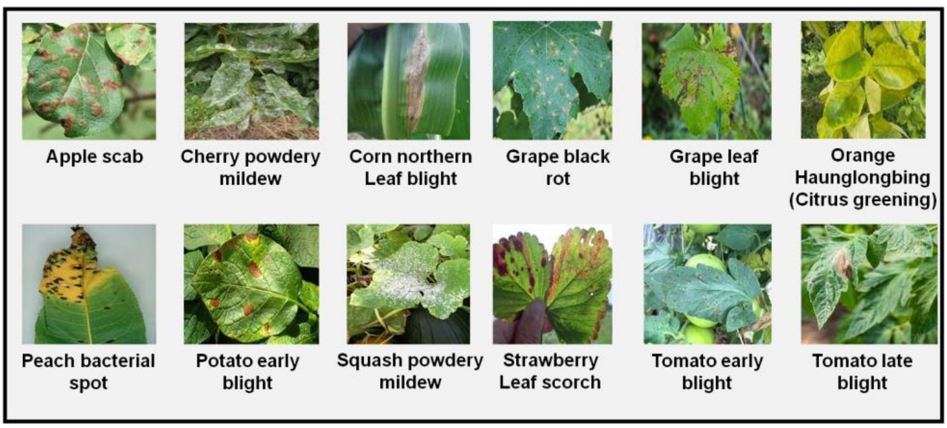


Figure 3: A few of the Plant Villages dataset's plant diseases (Source: https://www.mdpi.com/2223-7747/9/10/1319)

3.2 Data preprocessing using Gaussian filter (GA)

A 2D convolution operator called the Gaussian filter is used to reduce noise and smooth out pictures. Pictures must be pre-processed to get rid of noise. The peak signal to noise ratio (PSNR) value can be used to determine whether noise is present in magnetic resonant imaging (MRI) pictures. An image's PSNR value indicates whether there is more noise or less noise. If it is higher, there is more distortion. Since noise is present in the majority of the photos utilized in the proposed study, filtering is required and the Gaussian filtering approach is used. Convolution of a picture $\omega(w,z)$ and $\omega(w,z) * e(w,z)$ is given in Equation (1)

$$\omega(w,z) * e(w,z) = \sum_{t=-b}^{b} \sum_{t=-a}^{a} \omega(t,s) e(w=t, z-s) \qquad (1)$$

Where a and b are provided by $b = \frac{n-1}{2}$ and $a = \frac{m-1}{2}$. An illustrated depiction $e(w,z)$ as shown in Equation (2)

$$e(w,z) = \frac{1}{2\pi\sigma} f^{-\frac{w^2+z^2}{2\sigma^2}} \qquad (2)$$

The two parameters of the Gaussian filter are the normal variation σ and its size. Gaussian smoothing filters are efficient low-pass filters that applied by engineers in real-world vision applications. They are effective in frequency and spatial domains. The 2D distribution is employed as a point-spread function convolution. The Gaussian Smoothing works. Since the picture is stored as discrete pixels, a discrete approximation of the Gaussian function must be created before the conversion function performed.

3.3 Feature extraction using linear discriminant analysis (LDA)

A machine learning algorithms to process and analyze data more rapidly provide precise predictions or classifications, feature extraction seeks to draw attention to the essential characteristics as well as patterns with the data. On everyday use of LDA is to identify the linear characteristics that optimize among-class separation of data while reducing with the class dispersion. Consider a training data set containing b samples {s^1 .... t^B }, where s^g represents a column vector of distance for each example. The training examples are instances of the R classes. Let $b = |X_e|$be the collection of the cases of course and let a be the total number of examples in class e = 1...E. In LDA, the scatter matrices among types are computed as follows in Equations (3 and 4):

$$e_y = \frac{\sum_f \sum_{g \in y_e} (s^h - a_e)(t^h - a_e)^t}{m} \qquad (3)$$

$$e_o = \frac{\sum_e b_e (a_e - b)(a_e - a)^S}{m} \qquad (4)$$

Where $t^h$is is the average of the eth class,$n_e = \frac{1}{n_e} \sum_{h \epsilon C_e} t^h$and $b = \frac{1}{N} \sum_h t^h$is the standard of the data sample. We look for a straightforward adjustment $t \rightarrow Y^S t$ that improves the virtual among-class conflict to the inside-group conflict, where Y is a $r \times r'$matrix and r is the desired size. Demonstrating that $F_o y = \lambda F_y y$ is the result of the correspondence between the generalized eigenvectors and the r largest eigenvalues is represented by the columns of the ideal Y. This conclusion leads to the simultaneous diagonalization by Y of the scatter matrices

$Y^S F_o Y$and $Y^S F_y Y$. As an alternative, LDA rectifies data and across classes.

### 3.4 Integrated spider monkey optimized efficient gradient boosting (ISMO-EGB)

The integrated spider monkey optimized (ISMO) algorithm and the efficient gradient boosting (EGB) are two potent optimization approaches that have been combined to create the integrated spider monkey optimized efficient gradient boosting (ISMO-EGB). This hybrid approach leverages the strong learning capabilities of efficient gradient boosting and the collective intelligence of a spider monkey-optimized algorithm to tackle the challenges of sequential data analysis and complex optimization issues.

#### 3.4.1 Efficient Gradient boosting (GB)

Boosting algorithms combine weak learners, that is, learners who perform better than random strong learners. Gradient boosting is a technique for regression that resembles boosting. If there is improper regularization of the iterative phase, this method experience over-fitting. In some loss functions, the next repetition stops immediately once the pseudo-residuals become zero. Several hyper-parameters of normalization are taken into consideration to regulate the gradient-enhancing additive process. It makes sense to use shrinking to reduce the size of each slope descent step while doing regularizing gradient boosting. It is possible to attain further regularization by restricting the complexity of the trained models. In the case of decision trees, a restriction on the minimum number of occurrences is required to divide a node or the depth of the plants. In gradient boosting, as opposed to Random forests with the standard hyper-parameter values, it restricts the trees expressive potential. Lastly, a different family of hyper-parameters that are included in the various gradient boosting variants is the one that randomizes the base learners, which can enhance the ensemble's generalization even more haphazard sub-sampling without substitution in Equation (5).

$$E_n(W) = E_{n-1}(W) + \rho_n g_n(W) \tag{5}$$

$E^*(W)$ by minimizing the probability that a loss function will be applied. Using a weighted sum of parts, gradient boost creates an additive estimate in Equations (6-8).

$$E_0(W) = \underset{\alpha}{\arg\min} \sum_{j=1}^{M} K(z_j, \alpha) \tag{6}$$

$$\rho_n g_n(W) = \underset{\rho, g}{\arg\min} \sum_{j=1}^{M} K\left(z_j, E_{n-1}(W_j) + \rho g(W_j)\right) \tag{7}$$

$$q_{nj} = \left[\frac{\partial K(w_j, E(W))}{\partial E(W)}\right]_{E(W)=E_{n-1}(W)} \tag{8}$$

Regularizing gradient boosting entails using a contraction to lessen the size of each gradient improvement step $E_n(W) = E_{n-1}(W) + u\rho_n g_n(W)$. Regularization can be accomplished by putting constraints on the intricacy of the learned system.

#### 3.4.2 Integrated Spider Monkey Optimization Algorithm (ISMO)

Every solution at the local leader stage has the opportunity to upgrade its position by using optimization. The remaining selection processes resemble those of the standard Integrated Spider Monkey Optimization Algorithm (ISMO). The algorithm known as the ISMO generates the random population of solutions with D-dimension. These function values are used to compute each value. The answer is to be updated using the conventional $W < W_i$, the new trail $(T_{mtrial})$ mentioned as $T_{ni.}$ NMM changes control the new approach in addition to $T_{ni.}$ is replace into $T_{mtrial}$ if $E(T_{m_i})$. Apply reflection transformation in Equation (9).

$$Tn_{new,j} = Tn_{Centroid\ gr'}\ j + \beta^{*(Tn_{Centroid\ gr'}j - KN_{gr,}{}^{j})} \tag{9}$$

Thus, $j^{th}$ with the updated reflected location, the measure is $Tn_{new,j}$ the least favorable location globally in $l^{th}$ length is $KN_{l,j}$ the center of the whole population is Centroid $_{gr}$as shown in Equation (10).

$$Centroid_{gr} = \frac{1}{GrT}\sum_{i=(GT\_gr_{wi})}^{(GT\_gr_{wi})} i \tag{10}$$

The group's magnitude $gr^{th}$GrT, the result of the operation$Tn_{new,j}$, is computed using two options. In Equation (11), apply the expansive modification.

$$Tn_{new\ es,j} = Tn_{Centroid\ g'}\ j + \alpha^{*(Tn_{new\ ex,j}} - Tn_{Centroid\ g'}\ j) \tag{11}$$

The reflecting $Tn_{new\ ex,j}$ the expanding experiment that's $Tn_{new\ ex,j}$ it's a brilliant idea to replace $Tn_{i,j}$ with $Tn_{new,j}$ If the function value is unknown, the NMM searches for a new record. $Tn_{new,j}$ is more than $Tn_{trial,j}$, as shown by Equation (12) using reduction conversion.

$$Tn_{new\ es,j} = Tn_{Centroid\ gr'}\ j + \gamma * (Tn_{Centroid\ gr'}\ j - KN_{\ gr,s}) \tag{12}$$

During the phase of contraction in the $j^{th}$ quantity $Tn_{new\ es,j}$ refers to the updated role (Tn). The present location. $Tn_{i,j}$ is used when the contraction function value falls short of optimal $Tn_{trial,j}$. Our updated technique never made use of the shrink transformation function.

## 4. Result

MATLAB software is used in this research to stimulate results. To evaluate the efficiency of our proposed method by comparing it with traditional methods such as random forest (RF) [21], Support vector machine (SVM) [21], artificial neural network (ANN) [21] and K-nearest neighbour (KNN) [21] with regard to f1-score, accuracy, precision and recall. As a result, the outcomes of our suggested technique are higher than other existing approaches.

### 4.1 Accuracy

One performance metric that is used in classification tasks is accuracy. Comparing the actual labels in the input data with the expected labels is one technique to evaluate the measurement's effectiveness. An accurate forecast is one when the actual and expected titles coincide. Measures of accuracy from early plant illness detection Figure 4 and Table 1 demonstrate the comparison of accuracy between the existing and proposed methods as shown in Equation (13). ISMO-EGB has an accuracy value of 95.05%. This proves that our method performs better than other predictable methods.

$$Accuracy = \frac{Amountofperfectlycategorizeddatainstances}{Totaldatainstances} \times 100 \tag{13}$$

Table 1: Comparison of Accuracy among suggested and existing methods

| Methods | Accuracy (%) |
|---|---|
| RF [21] | 73.38 |
| SVM [21] | 67.23 |
| ANN [21] | 65.68 |
| KNN [21] | 63.2 |
| ISMO-EGB [Proposed] | 95.05 |

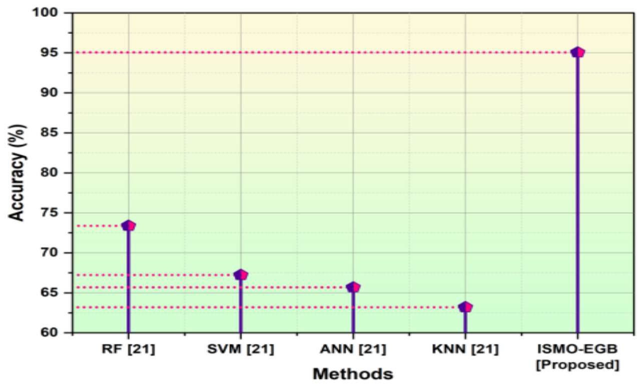


Figure 4: Comparison of Accuracy

4.2 Precision

Based on the number of accurate forecasts it generates, a model calculates the percentage of genuine optimistic predictions or right optimistic predictions. The accuracy is referred to calculate based on the total number of accurate projections as shown in Equaiton (13). Figure 5 and Table 2 present a comparison of accuracy between the conventional and suggested procedures. The value for ISMO-EGB is higher by 95.65%. This value proved that our proposed method performs better than other methods.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \times 100 \quad (14)$$

Table 2: Comparison of precision among suggested and existing methods

| Methods | Precision (%) |
|---|---|
| RF [21] | 72.88 |
| SVM [21] | 66.99 |
| ANN [21] | 65.6 |
| KNN [21] | 63.18 |
| ISMO-EGB [Proposed] | 95.65 |

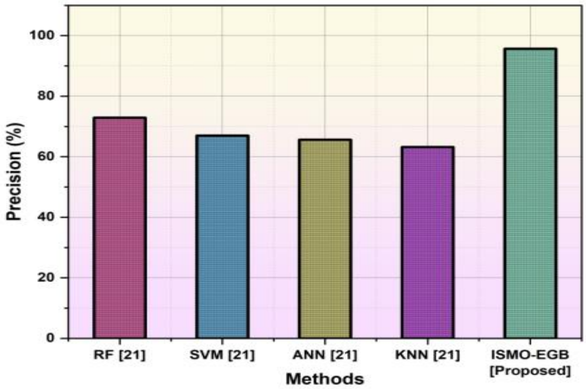


Figure 5: Comparison of precision

4.3 Recall

One crucial performance metric is recall. Recall is the product of actual positive rate and sensitivity. To calculate the memory score, divide the total number of projected positive outcomes and predicted favourable outcomes by the entire number of correctly anticipated positive results and incorrectly assessed painful consequences. To minimize false negatives and maximize true positives when detecting plant sickness in crops, recall is essential. Figure 6 and Table 3 compare the recall of the recommended approach with the old system. With a recall score of 97%, our suggested method demonstrated that the proposed model outperforms alternative techniques as shown in Equation (15).

$$\text{Recall} = \frac{\text{TP}}{\text{TP+False Negative (FN)}} \times 100 \quad (15)$$

Table 3: Comparison of recall

| Methods | Recall (%) |
|---|---|
| RF [21] | 72.9 |
| SVM [21] | 66.97 |
| ANN [21] | 65.3 |
| KNN [21] | 63.15 |
| ISMO-EGB [Proposed] | 97 |

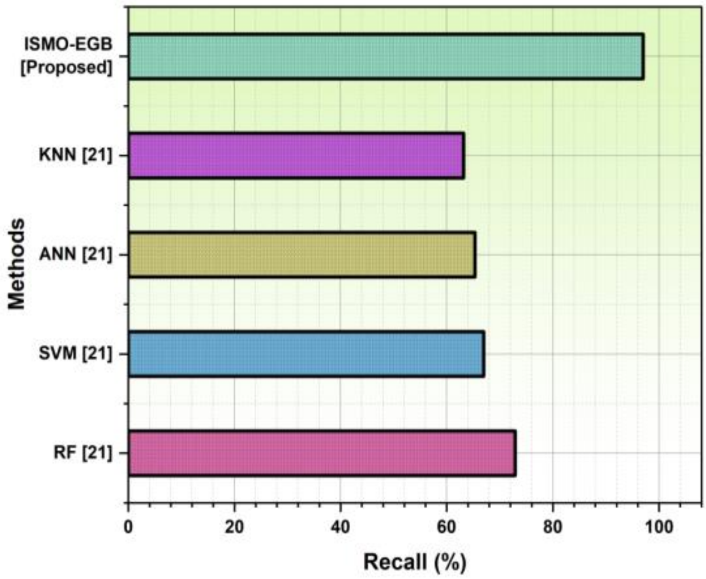


Fig.6. Comparison of recall among suggested and existing methods

4.4 F1- Score

When completing classification jobs, the F1 measure comes in handy. When the F1 value is more excellent, it indicates that the model balances recall and accuracy fairly. Figure 6 compares the f1-scores using the current and suggested methods. The ISMO-EGB score is 90.9% greater than other traditional approaches, demonstrating the improved performance of our proposed method. The results of the F1-measure are shown in Figure 7 and Table 4.

$$\mathrm{F1-score} = \frac{(\mathrm{recall})\times(\mathrm{precision})\times 2}{\mathrm{recall}+\mathrm{precision}} \quad (16)$$

Table 4: Comparison of F score

| Methods | F1-Score (%) |
|---|---|
| RF [21] | 71.98 |
| SVM [21] | 66.02 |
| ANN [21] | 64.76 |
| KNN [21] | 62.55 |
| ISMO-EGB [Proposed] | 98.52 |

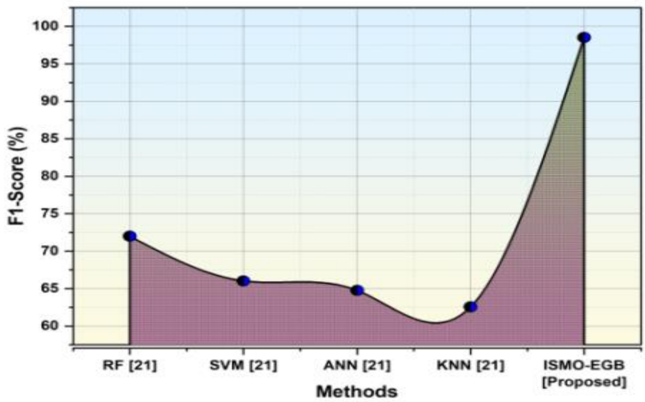


Figure 7: Comparison of F score among suggested and existing methods

4.5 Discussion

An artificial neural network (ANN) consists of an input layer, an undetected layer and a result layer. [21]. Neurons make up each layer and each layer carries weights that are sent to the neurons in the layer below. It functions similarly to the neural cells seen in the human brain. The ANN type requires instruction. One of the popular algorithms that excel in classes and regression is the SVM [21]. It works well when applied to solve linear and non-linear issues. One kind of group learning technique is random forest. It referred to regression and classification SVM. Decision trees are the fundamental building component used by Random Forest. K-nearest neighbor (KNN) is regarded as a statistical method that makes no basic presumptions about the data's dispersion. The evaluation metrics show that our proposed Integrated ISMO-EGB model outperforms modern techniques for accurate picture recognition as well as identification.

## 5. Conclusion

Classifying plant leaf pictures as healthy or unhealthy was the precise aim of this paper. This work used the machine learning ISMO-EGB architecture to do this. In order to increase model accuracy without lowering training efficiency, the model was adjusted utilizing strategies including differential learning rates. The accuracy, precision, recall and f1-score of the trained model were 95.05%, 95.65%, 97.00% and 98.52%. The experimental findings demonstrated that, despite the intricate inter and interclass variances, the model's capacity to classify several categories of plant leaf pictures as either healthy or sick appears optimistic. In the future, crop protection tactics could become even more precise and effective the combination of machine learning and other cutting-edge technologies like quantum computing for intricate models and cryptocurrency for traceability.

## References


1. Ead, Waleed M.and Mohamed M. Abbassy. "IOT based on plant diseases detection and classification." In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 2030-2033. IEEE, 2021. DOI: https://doi.org/10.1109/ICACCS51430.2021.9441954
2. Nawaz, Muhammad Amir, Rana Mudassar Rasool, Maryam Kausar, Amir Usman, Tanvir Fatima Naik Bukht, Rizwan Ahmadand Ahmad Jaleel. "Plant disease detection using internet of thing (IoT)." International Journal of Advanced Computer Science and Applications 11, no. 1 (2020).
3. Mugithe, Pradeep Kumar, Rohit Varma Mudunuri, B. Rajasekarand S. Karthikeyan. "Image processing technique for automatic detection of plant diseases and alerting system in agricultural farms." In 2020 International Conference on Communication and Signal Processing (ICCSP), pp. 1603-1607. IEEE, 2020. https://doi.org/10.1109/ICCSP48568.2020.9182065
4. Panchal, Poojan, Vignesh Charan Ramanand Shamla Mantri. "Plant diseases detection and classification using machine learning models." In 2019 4th international conference on computational systems and information Technology for Sustainable Solution (CSITSS), pp. 1-6. IEEE, 2019. https://doi.org/10.1109/CSITSS47250.2019.9031029
5. Abbas, Amreen, Sweta Jain, Mahesh Gourand Swetha Vankudothu. "Tomato plant disease detection using transfer learning with C-GAN synthetic images." Computers and Electronics in Agriculture 187 (2021): 106279. https://doi.org/10.1016/j.compag.2021.106279
6. Poornappriya, T. S.and R. Gopinath. "Rice plant disease identification using artificial intelligence approaches." International Journal of Electrical Engineering and Technology 11, no. 10 (2022): 392-402. https://doi.org/10.34218/IJEET.11.10.2020.050
7. Tulshan, Amrita S.and Nataasha Raul. "Plant leaf disease detection using machine learning." In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-6. IEEE, 2019. https://doi.org/10.1109/ICCCNT45670.2019.8944556
8. Sujatha, Radhakrishnan, Jyotir Moy Chatterjee, N. Z. Jhanjhiand Sarfraz Nawaz Brohi. "Performance of deep learning vs machine learning in plant leaf disease detection." Microprocessors and Microsystems 80 (2021): 103615. https://doi.org/10.1016/j.micpro.2020.103615
9. Singh, Vijai, Namita Sharmaand Shikha Singh. "A review of imaging techniques for plant disease detection." Artificial Intelligence in Agriculture 4 (2020): 229-242. https://doi.org/10.1016/j.aiia.2020.10.002
10. Singh, Davinder, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawatand Nipun Batra. "PlantDoc: A dataset for visual plant disease detection." In Proceedings of the 7th ACM IKDD CoDS and 25th COMAD, pp. 249-253. 2020. https://doi.org/10.1145/3371158.3371196
11. Li, Yang, Jing Nieand Xuewei Chao. "Do we really need deep CNN for plant diseases identification?." Computers and Electronics in Agriculture 178 (2020): 105803. https://doi.org/10.1016/j.compag.2020.105803
12. Farber, Charles, Mark Mahnke, Lee Sanchezand Dmitry Kurouski. "Advanced spectroscopic techniques for plant disease diagnostics. A review." TrAC Trends in Analytical Chemistry 118 (2019): 43-49. https://doi.org/10.1016/j.trac.2019.05.022
13. Shruthi, U., V. Nagaveniand B. K. Raghavendra. "A review on machine learning classification techniques for plant disease detection." In 2019 5th International conference on advanced computing & communication systems (ICACCS), pp. 281-284. IEEE, 2019. https://doi.org/10.1109/ICACCS.2019.8728415
14. Sharma, Parul, Yash Paul Singh Berwaland Wiqas Ghai. "Performance analysis of deep

learning CNN models for disease detection in plants using image segmentation." Information Processing in Agriculture 7, no. 4 (2020): 566-574. https://doi.org/10.1016/j.inpa.2019.11.001
15. Guo, Yan, Jin Zhang, Chengxin Yin, Xiaonan Hu, Yu Zou, Zhipeng Xueand Wei Wang. "Plant disease identification based on deep learning algorithm in smart farming." Discrete Dynamics in Nature and Society 2020 (2020): 1-11. https://doi.org/10.1155/2020/2479172
16. Ristaino, Jean B., Pamela K. Anderson, Daniel P. Bebber, Kate A. Brauman, Nik J. Cunniffe, Nina V. Fedoroff, Cambria Finegold et al. "The persistent threat of emerging plant disease pandemics to global food security." Proceedings of the National Academy of Sciences 118, no. 23 (2021): e2022239118. https://doi.org/10.1073/pnas.2022239118
17. Sinha, Adityaand Rajveer Singh Shekhawat. "Review of image processing approaches for detecting plant diseases." IET Image Processing 14, no. 8 (2020): 1427-1439. https://doi.org/10.1049/iet-ipr.2018.6210
18. Hassan, Sk Mahmudul, Arnab Kumar Maji, Michał Jasiński, Zbigniew Leonowiczand Elżbieta Jasińska. "Identification of plant-leaf diseases using CNN and transfer-learning approach." Electronics 10, no. 12 (2021): 1388. https://doi.org/10.3390/electronics10121388
19. Chandol, Mohan Kumar, M. Elangovan, U. Muthusamyand K. Sankar. "Enhancement of agriculture based crop yield prediction using R tool and machine learning." Turkish Online Journal of Qualitative Inquiry (TOJQI) 12, no. 7 (2021): 5155-5165.
20. Saleem, Muhammad Hammad, Johan Potgieterand Khalid Mahmood Arif. "Plant disease classification: A comparative evaluation of convolutional neural networks and deep learning optimizers." Plants 9, no. 10 (2020): 1319. https://doi.org/10.3390/plants9101319
21. Ganatra, Nilayand Atul Patel. "A multiclass plant leaf disease detection using image processing and machine learning techniques." International Journal on Emerging Technologies 11, no. 2 (2020): 1082-1086.