# Automated Agricultural Plant Leaf Disease Detection using Butterfly Optimization Algorithm with Deep Learning

**D. Raja[1], Dr. M. Karthikeyan[2]**

[1]Research Scholar, Department of Computer and Information Science, Faculty of Science, Annamalai University, Tamil Nadu, India
Email: *agdrajaphd[at]gmail.com*

[2]Assistant Professor/Programmer, Department of Computer and Information Science, Faculty of Science, Annamalai University, Tamil Nadu, India
Email: *karthiaucse[at]gmail.com*

**Abstract:** ***Plant leaf disease automated classification and recognition can play a significant part in reducing these threats. Agricultural Plant Leaf Disease Detection (APLDD) model is Computer Vision (CV) based process developed for automatic classification and recognition of abnormalities or ailments in plant leaves. This approach can play a critical part in precision farming by allowing initial recognition and appropriate interference in preventing the disease's spread and reducing the loss of crops. By the influence of CV and Deep Learning (DL), the APLDD technique can remarkably assist scientists and agriculturalists in employing proactive measures in safeguarding crops, promote sustainable farming routines, and enhance yields. Therefore, this article presents an Automated Agricultural Plant Leaf Disease Detection using Butterfly Optimization Algorithm with Deep Learning (APLDD - BOADL) technique. The presented model integrates VGG16 for feature extraction, Butterfly Optimization Algorithm (BOA) for hyperparameter tuning, and Long Short - Term Memory (LSTM) for disease classification. The VGG16 architecture is exploited for the generation of high - level features from these images. For optimizing the results of the VGG16 model, Butterfly Optimization Algorithm (BOA) is applied for hyperparameter tuning process. Finally, the LSTM classification enables accurate disease identification and differentiation between healthy and diseased leaves. The results demonstrate the superior accuracy and robustness of our approach compared to traditional methods.***

**Keywords:** Agriculture; Plant disease detection; Computer vision; Deep learning; Butterfly optimization algorithm

## 1. Introduction

Agriculture is a significant source of economic development in India. The farmer chooses the necessary crop depending on the location's weather conditions, types of soil, and commercial value [1]. Due to weather changes, political uncertainty, and increasing populations, the agricultural sectors started to look for novel techniques for raising food production [2]. These researchers allow search for new high productivity inventions, which are accurate and efficient. With help of precise agriculture in Information Technology (IT), cultivators might gather data and information for making the proper decision on great agricultural production [3]. Precision agriculture (PA) is an advanced technology, which provides complicated methods to optimizer agricultural commodities. Economic growth in agriculture can be attained through exploitation of these complex technologies [4]. PA is utilized for several applications, like identification of plant diseases, weed detection, crop yield production, plant pest detection, and so on [5]. A farmer utilizes pesticides to avoid diseases, control pests, and raise crop yield. Crop diseases are causing difficulties for farmers because of lower output and financial losses and industrial cultivation [6]. Hence, disease detection and severity are focusing on the requirement that is described as appropriate.

Manual methods in traditional agricultural businesses are not cover wide field of crops and deliver earlier background data for decision making processes [7]. Therefore, the authors are not stopped looking for methods to advance automated practical approaches and efficient techniques for identifying plant diseases. The deep learning (DL) based techniques, specifically, have built several applications in plant disease detection [8]. To address the issuers related to conventional classification approaches and describe cutting edge techniques in this domain [9]. The DL technique is a new technique, which has exposed good promise and achieved in different areas where it is utilized. But it can be a category of Machine Learning (ML) technique, which attempts to model at a higher level of data abstraction using articulating structures of different transformations [10].

This article presents an Automated Agricultural Plant Leaf Disease Detection using Butterfly Optimization Algorithm with Deep Learning (APLDD - BOADL) technique. The presented model integrates VGG16 for feature extraction, Butterfly Optimization Algorithm (BOA) for hyperparameter tuning, and Long Short - Term Memory (LSTM) for disease classification. The VGG16 architecture is exploited for the generation of high - level features from these images. For optimizing the results of the VGG16 model, Butterfly Optimization Algorithm (BOA) is applied for hyperparameter tuning process. Finally, the LSTM classification enables accurate disease identification and differentiation between healthy and diseased leaves. The results demonstrate the superior accuracy and robustness of our approach compared to traditional methods.

## 2. Related Works

In [11], the authors have made an organized literature analysis on the applications of developed ML and DL technologies for classifying plant disease. All stated systems can be categorized by the particular processing methods such as image segmentation, and feature extraction, in addition to the standardized experimental setup metrics such as training database used or aggregate sum of testing, percentage of classification precision, quantity of diseases below considerations, and types of classifier employed. In [12], Efficient Net DL architecture was recommended in plant leaf disease classification and the implementation of such technique can be made compared to other advanced DL techniques. The Plant Village databases are used for training approaches. The other DL methods and Efficient Net model are trained by TL technique. In the TL method, each layer of technique is determined to be trainable. Panigrahi et al. [13] focussed on supervised ML methods such as KNN, DT, RF, SVM, and NB, for corn plant disease detection by method of plant images. The aforementioned classification techniques are studied and a comparison is done to select the great proper approach with high precision for predicting the plant disease. In [14], the authors have studied the current CNN network system related to plant leaf disease classification.

## 3. The Proposed Model

This article has introduced a new APLDD - BOADL technique for agricultural plant leaf disease detection and classification. It involves VGG16 feature extractor, BOA based hyperparameter optimization, and LSTM for disease classification. Fig.1 exhibits the overall process of APLDD - BOADL approach.

### 3.1. VGG16 Model

The VGG16 architecture is exploited for the generation of high-level features from these images. VGG-16 approach is implemented in extracting features [15]. A sequence of non-linearity and complex filters are employed to solve following formulas. CNN has level hierarchy. The $x_j$ is attained from the input signal $x$:

$$x_j = \rho W_j x_j - 1. \tag{1}$$

However, $W_j$ depicts the linear operator, and these conditions demonstrate non-linear conduct. Additionally, $W_j$ is frequently implemented in convolution in CNN, with an exponential sigmoid$[1 + \exp(-x)]^{-1}$ or $\rho$ as a rectifier max $(x; 0)$. $W_j$, the convolutional filter stack is considered.

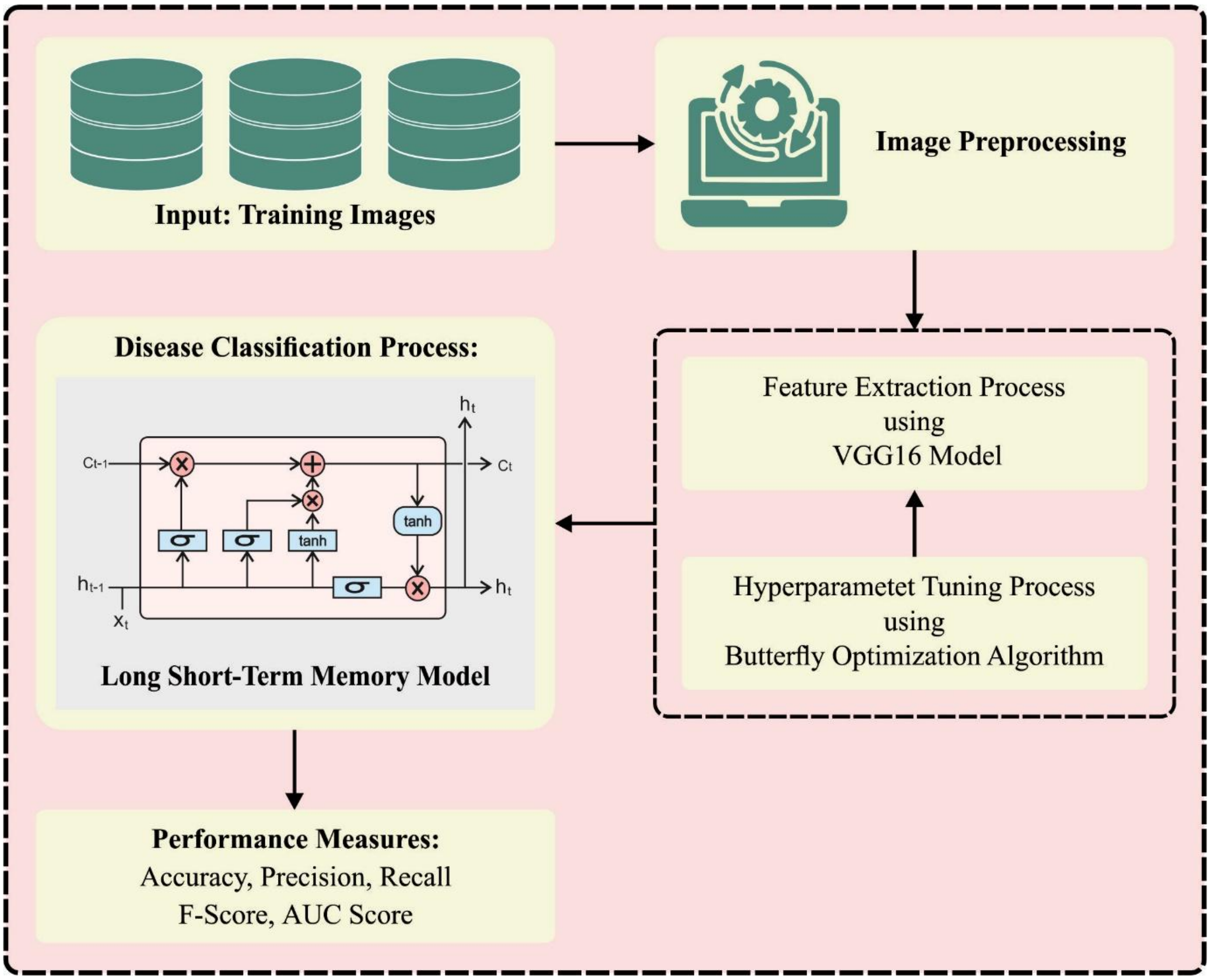



**Figure 1:** Overall process of APLDD - BOADL system

Hence, the convolution in the overall layers is resolved by the comprehensive convolution layers.

$$x_j(u, k_j) = \rho\left(\sum_k \left(W_{j,k_j}(.,k) * x_{j-1}(.,k)\right)(u)\right) \quad (2)$$

Additionally, a distinct convolution model known as * is enforced:

$$(g * f)(x) = \sum_{u=-\infty}^{\infty} g(u) f(x-u). \quad (3)$$

The optimization of CNN is non-convolutional issue. Weight $W_j$ is regularly trained by stochastic gradient descent, with the gradient largely specified by the back propagation. The data reliability is of greatest significance in DL technique and majorly relies on a tremendous data number that has to be trained. For training the basic patterns of the data, an enormous data number is required. TL is employed to address the training dataset issue that is dispersed consistently independent of the testing dataset dispersion, which inspires in applying TL to address the incompetently trained dataset issues. The VGG feature sequence network is the $3 \times 3$ complex layers that are on top of one another, and the depth gets mostly higher.

### 3.2. BOA based Parameter Tuning

For optimizing the results of the VGG16 model, the BOA is applied for hyperparameter tuning process. BOA duplicates the courting and foraging conduct of the butterflies. In BOA, every butterfly has a discrete perception and smelling capability, and the robustness of these capabilities differs among individuals [16]. Eq. (4) portrays how stern other butterflies are capable of smelling an individual amidst them and is represented as follows:

$$f(x) = cI^a \quad (4)$$

Also, $f(x)$, $c$, $I$, and $a$ depict the odor concentration function, sensory shape coefficient, stimulus intensity, i.e., the fitness value, and the intensity coefficient, and the values lie between the range [0,1].

The variable $c$ that portrays the sensory shape coefficient can supposedly adopt any value inside the limit $[0, \infty)$, and its computation is represented in Eq. (5).

$$c_{t+1} = c_t + [0.025/(c_t \cdot T_{\max})] \quad (5)$$

Also, 0.01 is the original value of $c$, and $T_{\max}$ specifies the algorithm's greatest iteration number. The BOA model determines the model's local and global search according to $p$, the switching probability and the location upgrade formula are depicted in Eq. (6):

$$x_{i\,t+1} = \begin{cases} x_i^t + (r^2 \cdot g^* - x_i^t) \cdot f_{i'} p < rand \\ x_i^t + (r^2 \cdot x_j^t - x_k^t) \cdot f, p \geq rand \end{cases} \quad (6)$$

Also, $g^*$, $x_j^t$,$x_k^t$, $f_i$, and $r$illustrates the present iteration's optimum location of the overall butterflies, $j$-th and k-th butterflies' spatial locations in the $t$-th iteration,the $i$-th butterfly's fitness value, and the random number value, subsequently.

### 3.3. LSTM Classification

At last, the LSTM model is used for classification process. LSTM is a kind of Recurrent Neural Network (RNN) construction that is particularly built for modelling and handlining serial data [17]. LSTM is specifically efficient in learning and capturing long - term reliabilities and patterning in time sequence data, enabling it most appropriate for time sequence vibration data. The LSTM unit's mathematical representation of formulas is as follows:

$$z = \tanh\left(w\left[x^t, h^{t-1}\right] + b\right) \quad (7)$$

$$z^i = \sigma\left(w^i\left[x^t, h^{t-1}\right] + b^i\right) \quad (8)$$

$$z^f = \sigma\left(w^f\left[x^t, h^{t-1}\right] + b^f\right) \quad (9)$$

$$z^0 = \sigma\left(w^0\left[x^t, h^{t-1}\right] + b^0\right) \quad (10)$$

The cell state $c^t$ is provided by the addition of the prior cell state, the Hadamard product () of forget gate, and the Hadamard product of input gate and cell update $z$.The formula of cell state $c^t$ is as follows:

$$c^t = z^f * c^{t-1} + z^i * z; \quad (11)$$

Likewise, $h^t$, the new hidden state is represented by:

$$h^t = z^0 * \tanh(c^t); \quad (12)$$

And $y^t$, the current output is depicted by:

$$y^t = \sigma(w' h^t). \quad (13)$$

The construction of the LSTM allows networking to efficiently retain and capture long-term reliabilities in serial data. Through restrictively forgetting and storing data at every time step, the networking can be learning to leverage and detect appropriate patterns in the information, though they happen over longer time-span.

## 4. Performance Validation

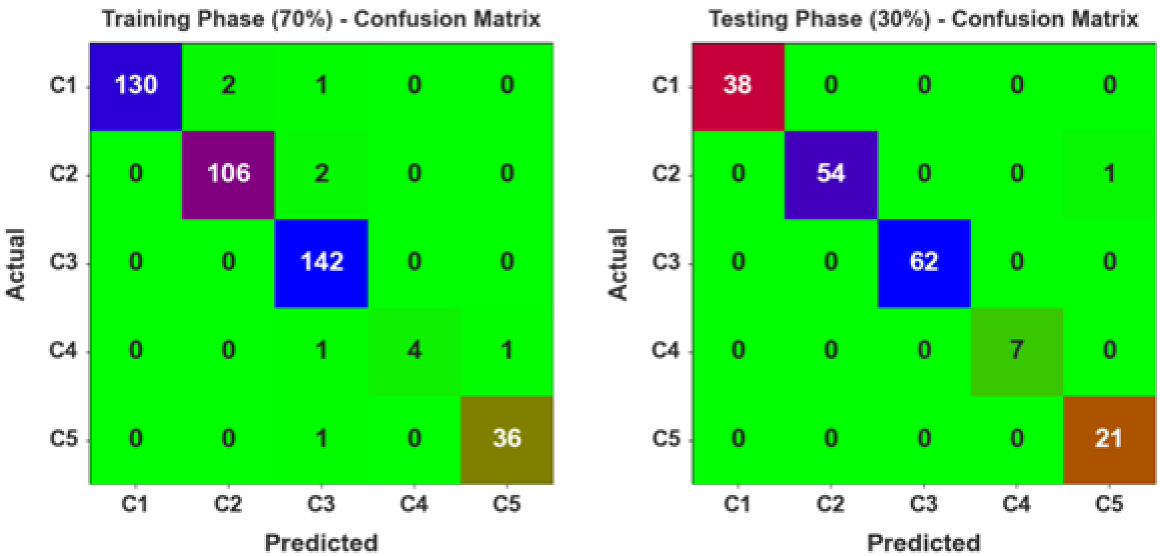
In this section, the outcomes of the APLDD - BOADL method are studied on the plant disease dataset [18], comprising 609 samples as demonstrated in Table 1. Fig.2 demonstrates the sample images.

**Table 1:** Description of database

| Label | Class Name | No. of Images |
|---|---|---|
| C1 | Black spot | 171 |
| C2 | Canker | 163 |
| C3 | Greening | 204 |
| C4 | Melanose | 13 |
| C5 | Healthy | 58 |
| **Total images** | | **609** |

**Figure 2:** Sample Images

Fig.3depicts the classifier results of the APLDD – BOADL method under test dataset. Figs.3a - 3b demonstrates the confusion matrix offered by the APLDD – BOADL technique at 70: 30 of TR set/TS set. The figure showed that the APLDD – BOADL approach has detected and classified all 5 class labels accurately. Likewise, Fig.3c reveals the PR examination of the APLDD – BOADL method. The figures described that the APLDD – BOADL approach has attained high PR performance under 5 classes. At last, Fig.3d demonstrates the ROC inspection of the APLDD - BOADL model. The figure illustrated that the APLDD – BOADL method has resulted in promising outcomes with high ROC values under 5 class labels.




(a) (b)

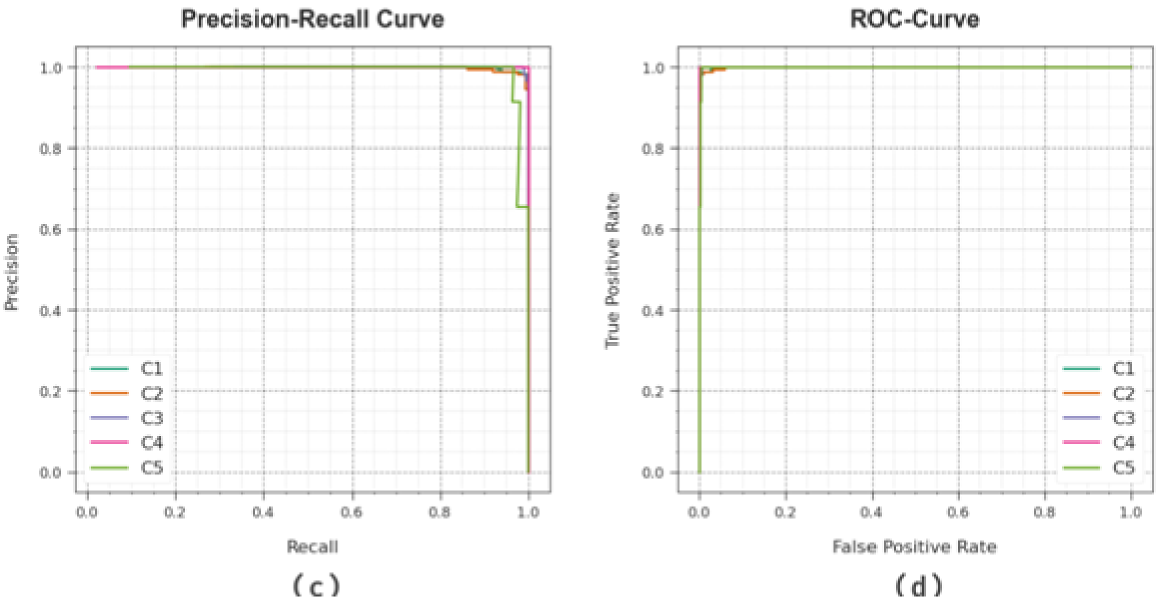



**Figure 3:** Performance of (a - b) Confusion matrices, (c) PR_curve, and (d) ROC_curve

The detection results of the APLDD-BOADL technique are studied in Table 2 and Fig. 4. The table values inferred the enhanced results of the APLDD-BOADL technique on all classes. On 70% of TR set, the APLDD-BOADL technique provides average $accu_y$ of 99.25%, $prec_n$ of 98.41%, $reca_l$ of 91.97%, $F_{score}$ of 94.51%, and $AUC_{score}$ of 95.72%.Also, on 30% of TS set, the APLDD-BOADL method provides average $accu_y$ of 99.78%, $prec_n$ of 99.09%, $reca_l$ of 99.64%, $F_{score}$ of 99.35%, and $AUC_{score}$ of 99.76%.

**Table 2:** Detection outcome of APLDD - BOADL algorithm on 70% of TR set/30% of TS set

| Class | $Accu_y$ | $Prec_n$ | $Reca_l$ | $F_{score}$ | $AUC_{score}$ |
|---|---|---|---|---|---|
| **Training Phase (70%)** | | | | | |
| Black spot (C1) | 99.30 | 100.00 | 97.74 | 98.86 | 98.87 |
| Canker (C2) | 99.06 | 98.15 | 98.15 | 98.15 | 98.76 |
| Greening (C3) | 98.83 | 96.60 | 100.00 | 98.27 | 99.12 |
| Melanose (C4) | 99.53 | 100.00 | 66.67 | 80.00 | 83.33 |
| Healthy (C5) | 99.53 | 97.30 | 97.30 | 97.30 | 98.52 |
| **Average** | **99.25** | **98.41** | **91.97** | **94.51** | **95.72** |
| **Testing Phase (30%)** | | | | | |
| Black spot (C1) | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| Canker (C2) | 99.45 | 100.00 | 98.18 | 99.08 | 99.09 |
| Greening (C3) | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| Melanose (C4) | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| Healthy (C5) | 99.45 | 95.45 | 100.00 | 97.67 | 99.69 |
| **Average** | **99.78** | **99.09** | **99.64** | **99.35** | **99.76** |

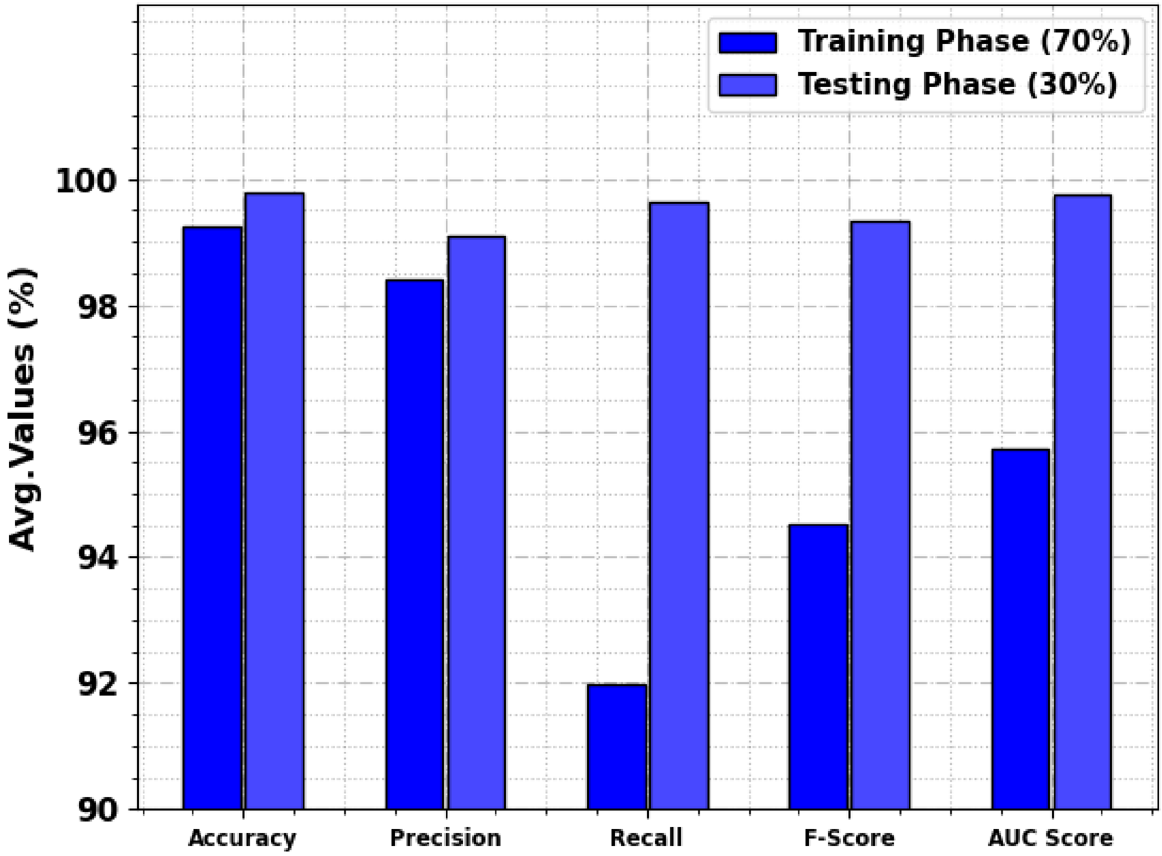



**Figure 4:** Average outcome of APLDD - BOADL algorithm on 70% of TR set/30% of TS set

Fig. 5 shows the training accuracy $TR\_accu_y$ and $VL\_accu_y$ of the APLDD-BOADL method. The $TL\_accu_y$ is defined by the evaluation of the APLDD-BOADL system on TR dataset while the $VL\_accu_y$ is calculated by assessing the performance on a separate testing dataset. The outcomes show that $TR\_accu_y$ and $VL\_accu_y$ increase with an upsurge in epochs. Consequently, the performance of APLDD-BOADL method gets improved on the TR and TS datasets with a maximum number of epochs.


**Figure 5:** $Accu_y$ curve of the APLDD-BOADL algorithm

Fig. 6, shows the $TR\_loss$ and $VR\_loss$ outcomes of the APLDD-BOADL method. The $TR\_loss$ describes the error amongst the original values and predictive performance on the TR data. The $VR\_loss$ represent the measure of the performance of the APLDD-BOADL method on individual validation data. The results show that the $TR\_loss$ and $VR\_loss$ tends to decrease with rising epochs. It depicted the improved performance of the APLDD-BOADL method and its ability to produce accurate classification. The reduced value of $TR\_loss$ and $VR\_loss$ illustrates the improved performance of the APLDD-BOADL method on capturing relationships and patterns.


**Figure 6:** Loss curve of the APLDD - BOADL algorithm

**Table 3:** Comparative outcome of APLDD - BOADL technique with other approaches

| Methods | Precision | F - Score | Accuracy |
|---|---|---|---|
| RF | 78.00 | 76.70 | 76.80 |
| SGD | 86.90 | 86.50 | 86.50 |
| SVM | 87.30 | 87.10 | 87.00 |
| VGG19 | 87.70 | 87.40 | 87.40 |
| InceptionV3 | 89.20 | 89.00 | 89.00 |
| VGG16 | 89.60 | 89.50 | 89.50 |
| SOSDL - APLDD | 99.12 | 99.26 | 99.56 |
| APLDD - BOADL | 99.25 | 99.35 | 99.78 |

The comprehensive comparison study of the APLDD-BOADL method is reported in Table 3 and Fig. 7 [7]. The results highlighted that the APLDD-BOADL technique accomplishes enhanced performance over recent models. Based on $prec_n$, the APLDD-BOADL system gains maximum $prec_n$ of 99.25% while the RF, SGD, SVM, VGG19, Inceptionv3, VGG-16, and SOSDL-APLDD techniques attain minimum $prec_n$ values.Besides, based on $F_{score}$, the APLDD-BOADL method gains maximum $F_{score}$ of 99.35% whereas the RF, SGD, SVM, VGG19, Inceptionv3, VGG-16, and SOSDL-APLDD systems attain reduced $F_{score}$ values. At last, based on $accu_y$, the APLDD-BOADL method gains increasing $accu_y$ of 99.78% whereas the RF, SGD, SVM, VGG19, Inceptionv3, VGG16, and SOSDL-APLDD approaches attain decreased $accu_y$ values.

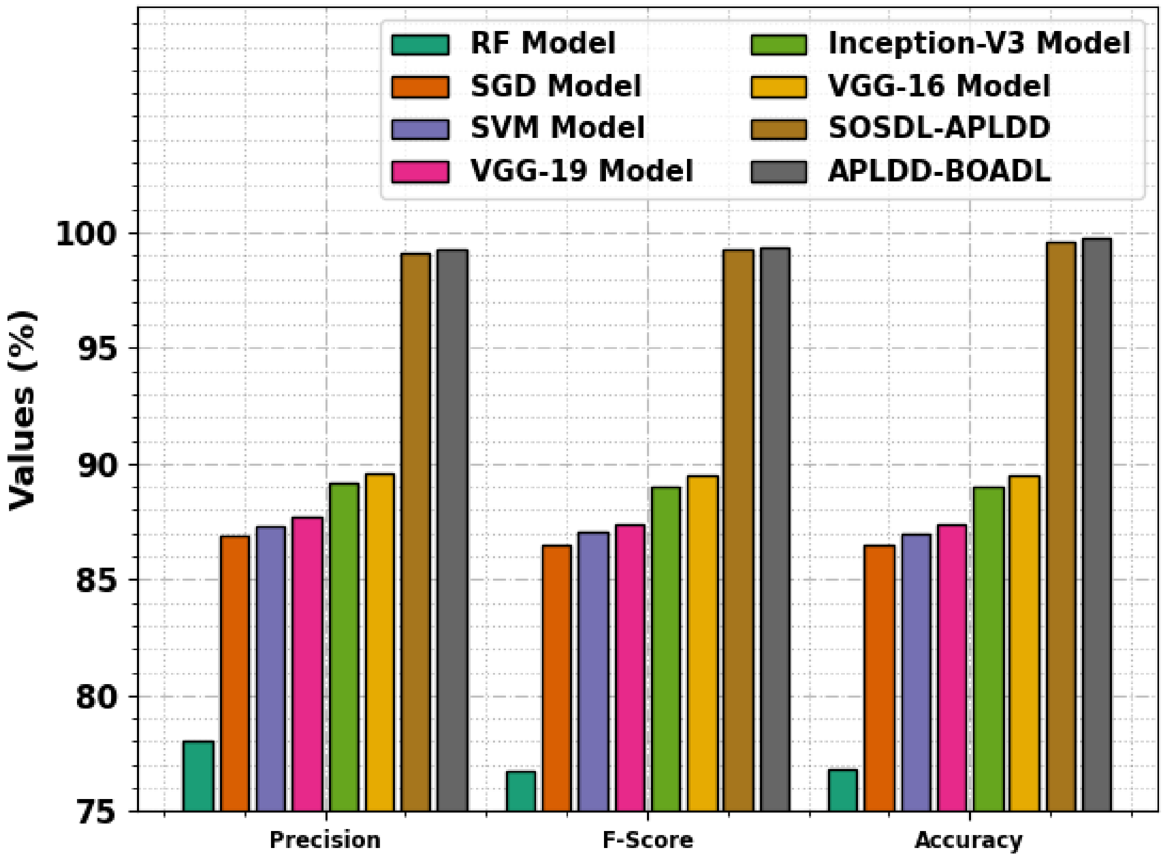



**Figure 7:** Comparative outcome of APLDD - BOADL technique with other approaches

## 5. Conclusion

This article has introduced a new APLDD - BOADL method for agricultural plant leaf disease detectionand classification. It involves VGG16 feature extractor, BOA based hyperparameter optimization, and LSTM for disease classification. The VGG16 architecture is exploited for the generation of high - level features from these images. For optimizing the results of the VGG16 model, the BOA is employed for hyperparameter tuning process. Finally, the LSTM classification enables accurate disease identification and differentiation between healthy and diseased leaves. The results demonstrate the superior accuracy and robustness of our approach compared to traditional methods.

## References

[1] Kumar, S., Prasad, K. M. V. V., Srilekha, A., Suman, T., Rao, B. P. and Krishna, J. N. V., 2020, October. Leaf disease detection and classification based on machine learning. In *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)* (pp.361 - 365). IEEE.

[2] Shruthi, U., Nagaveni, V. and Raghavendra, B. K., 2019, March. 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.

[3] Sharma, P., Hans, P. and Gupta, S. C., 2020, January. Classification of plant leaf diseases using machine learning and image preprocessing techniques. In *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp.480 - 484). IEEE.

[4] Annabel, L. S. P., Annapoorani, T. and Deepalakshmi, P., 2019, April. Machine Learning for Plant Leaf Disease Detection and Classification–A Review. In *2019 International Conference on Communication and Signal Processing (ICCSP)* (pp.0538 - 0542). IEEE.

[5] Swain, S., Nayak, S. K. and Barik, S. S., 2020. A review on plant leaf diseases detection and classification based on machine learning models. *Muktshabd*, *9* (6), pp.5195 - 5205.

[6] Ferentinos, K. P., 2018. Deep learning models for plant disease detection and diagnosis. *Computers and electronics in agriculture*, *145*, pp.311 - 318.

[7] Sujatha, R., Chatterjee, J. M., Jhanjhi, N. Z. and Brohi, S. N., 2021. Performance of deep learning vs machine learning in plant leaf disease detection. *Microprocessors and Microsystems*, *80*, p.103615.

[8] Hossain, E., Hossain, M. F. and Rahaman, M. A., 2019, February. A color and texture based approach for the detection and classification of plant leaf disease using KNN classifier. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp.1 - 6). IEEE.

[9] Trivedi, J., Shamnani, Y. and Gajjar, R., 2020, February. Plant leaf disease detection using machine learning. In *International conference on emerging technology trends in electronics communication and networking* (pp.267 - 276). Springer, Singapore.

[10] Ganatra, N. and Patel, A., 2020. A multiclass plant leaf disease detection using image processing and machine learning techniques. *International Journal on Emerging Technologies*, *11* (2), pp.1082 - 1086.

[11] Applalanaidu, M. V. and Kumaravelan, G., 2021, February. A Review of Machine Learning Approaches in Plant Leaf Disease Detection and Classification. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp.716 - 724). IEEE.

[12] Atila, Ü., Uçar, M., Akyol, K. and Uçar, E., 2021. Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, *61*, p.101182.

[13] Panigrahi, K. P., Das, H., Sahoo, A. K. and Moharana, S. C., 2020. Maize leaf disease detection and classification using machine learning algorithms. In *Progress in Computing, Analytics and Networking* (pp.659 - 669). Springer, Singapore.

[14] Lu, J., Tan, L. and Jiang, H., 2021. Review on convolutional neural network (CNN) applied to plant leaf disease classification. *Agriculture*, *11* (8), p.707.

[15] Deepa, N. and Chokkalingam, S. P., 2022. Optimization of VGG16 utilizing the Arithmetic Optimization Algorithm for early detection of Alzheimer's disease. *Biomedical Signal Processing and Control*, *74*, p.103455.

[16] Yue, Y., Cao, L., Chen, H., Chen, Y. and Su, Z., 2023. Towards an Optimal KELM Using the PSO - BOA Optimization Strategy with Applications in Data Classification. *Biomimetics*, *8* (3), p.306.

[17] Ahsan, M. and Salah, M. M., 2023. Efficient DCNN - LSTM Model for Fault Diagnosis of Raw Vibration Signals: Applications to Variable Speed Rotating Machines and Diverse Fault Depths Datasets. *Symmetry*, *15* (7), p.1413.

[18] Rauf, H. T., Saleem, B. A., Lali, M. I. U., Khan, M. A., Sharif, M. and Bukhari, S. A. C., 2019. A citrus fruits and leaves dataset for detection and classification of citrus diseases through machine learning. *Data in brief*, *26*, p.104340.