International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
IJISAE
ISSN:2147-6799 www.ijisae.org Original Research Paper

Classification and Identification of Plant Leaf Disease Leveraging Advanced Machine Learning and Deep Learning Techniques

Harshita Bhati[1] [1], Monika Rathore*[2]

Submitted: 05/05/2024 **Revised**: 18/06/2024 **Accepted**: 25/06/2024

Abstract: In India and globally, agriculture is very important in human life, as it is essential for providing food and promoting economic growth. Yet, plant leaves and crops could be infected by several diseases that impair their growth and cause a notable decline in agricultural productivity. Therefore, recognition of these diseases should be made in early time to avoid further damage. The traditional methods to predict and classify plant leaf diseases have always been very tiresome and faulty. Manual detection may cause delays, leading to significant crop losses and low yields. Computer vision technology is a proven enabler, which helps farmers to reduce their damage and increase production. There are multiple ways to detect and classify infections in plants using their images. Although much progress has been made, researchers must continuously improve their work to accommodate new challenges and incorporate the latest advancements. In this paper, we are focusing on advanced machine learning technology which has helped with improving the classification. Our review has shown that machine learning along with transfer learning has proven to be an efficient solution. The paper then analyses the main issues that need to be examined to enable further growth and improvement, such as image dataset formation, big data auxiliary domain, and optimization.

Keywords: Advanced Machine learning, plant leaf disease identification, deep learning models

1. Introduction

Agriculture is an important source of revenue in India's economy. According to the Indian Council for Agriculture, total production in the country is estimated to be a record 329.68 million (approx.) tonnes in 2022-2023, and it will increase to 345 million tonnes by 2030. Hence it is very important to focus on reducing diseases in plants and crops in order to improve production. The conventional approach farmers follow is time-intensive and demands high proficiency and adequate monitoring to comprehensively assess the plant condition. Therefore, automating the diagnosis and detection process is necessary to improve overall efficiency and save time. Several researchers have developed models based on different approaches. The continuous use of deep learning procedures with image processing methods for recognizing plant disease has become a significant subject for review to give a programmed analysis.

The process involves examining and processing plant leaf diseased image data. Depending on the features of images, multiple classes were created using advanced machine learning. In the end, a classifier is applied to support the accurate identification of multiple illnesses. The ultimate goal of all studies using this approach is to provide technical assistance for preventing and managing agricultural plant diseases [9].

Agricultural disease recognition is more complex, especially if it is a leaf disease. From the 1980s till today, multiple technologies are used to tackle this issue. The various methods include [21][31], Support Vector Machine classifier [4][5], shallow neural network methods, and Bayesian classifier [6].

A lot of this work is still being done. On the other hand, traditional machine learning algorithms have several drawbacks when applied to the practical image recognition of plant leaf diseases:

- They need original high-quality images of the diseased part of the leaf. As a result, the image acquisition methods should be accurate.
- These methods have typical steps, such as image pre-processing, segmentation, feature extraction, and classification. To improve their effectiveness, they need to be studied more.
- Finally, if the training dataset is large, constructing a fine model could be typical, especially with the traditional machine learning methods.

2 Data Preprocessing

Data pre-processing: This step improves accuracy and is essential for preparing raw data for model construction and training [24]. Enhancing the quality of data makes it easier to extract insightful information. Matplotlib, Pandas, and NumPy are the most dominant libraries for processing the data. The dataset is improved and made compatible by

[1] *Research scholar, Rajasthan Technical University, Kota, Rajasthan, India*
Email: harshitabhati2@gmail.com
[2]*Associate Professor, International school of informatics & Management, Jaipur, Rajasthan, India*
Email: monikarathore@gmail.com

removing null values, outliers, etc., before running the actual technique used for classification. The data must be in a proper structure for machine learning to work.

- Extracting features: There are multiple ways to extract features from the data like color and shape. Researchers these days utilize texture features to detect plant disease. Images are first downsized and then features are collected from them by the convolutional layer [29]. ReLU is completed after convolution and precise pooling.

Classification: This determines whether the input image of a plant leaf is healthy or unhealthy. Convolutional and pooling layers are utilized for character extraction, while entirely associated layers are used for classification [29].

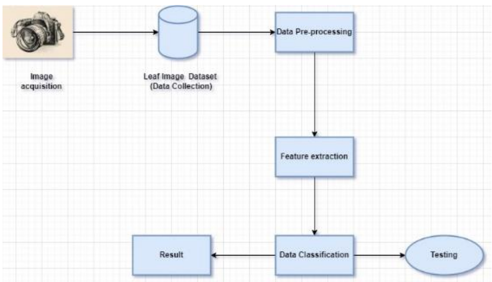

Fig. 1 Plant disease identification and classification process.

Recent improvements in machine learning approaches, like advanced deep learning, resulted in remarkable success in a multitude of applications and are now actively utilized to recognize plant leaf diseases through their images.

3. Overview of Advanced image recognition process of machine learning for plant leaf disease

3.1. Model - deep learning (DL)

The theory is based on research of ANN (Artificial Neural Network). This model is able to create and recognize categories utilizing different types of data like voice, text information etc., and can achieve excellent results. In some cases, this model was able to reach close to human-like results. Tagged data is used to train the model and can utilize advanced neural network with multiple layers topologies.

Presently, Convolutional Neural Network (ConvNet) [30] and Recurrent Neural Network [29], are the most utilized deep learning techniques.

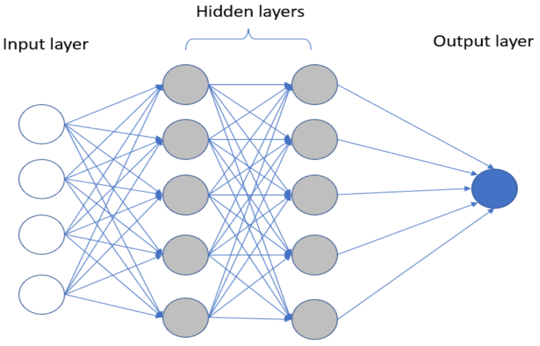

Fig. Deep Learning Model

3.2 Model - Transfer Learning (TL)

This model can utilize learning that is gained from one kind of task to another different task. This is done by reusing features from an already trained model without building a new one from scratch. Large datasets like ImageNet are used to train these models and the features (or weights) that are acquired from that previously trained model can be used with a mode that is customized to our needs. These customized models can better predict new and similar tasks/data. This approach has proven to significantly reduce the time it takes to train a new model and reduce anomalies.

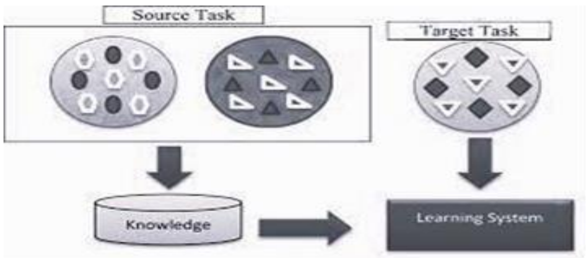

Fig. 3 Transfer learning model

3.3. Literature on Transfer Learning and Deep Learning

Conventional approaches are used for detecting plant disease with image recognition. A model is built using parameters from the features which mainly depend on existing knowledge. During the initial modeling stage, multilevel features can be fetched automatically from the given dataset. In this case, result precision is depended on the size of the data, meaning more data will lead to better accuracy.

Tirkey D et al. (2023) [1] In this paper, advance deep learning solutions are suggested for detecting and identifying soybean plant insects. The feasibility and accuracy of the proposed insect identification and detection accuracy method were investigated by trying out multiple transfer learning approaches. The proposed methodologies yield accuracy rates of 98.75%, 97%, and 97% when the author implements Yolov5 (Version5), CNN and

InceptionV3. As evident, YoloV5 showed better results.

Kirola M et al. (2022) [3] This paper has covered numerous machine learning and deep learning techniques including Convolution Neural Networks, Support Vector Machines (SVM), Random Forest (RF), LR and KNN. The objective of this study was to successful predict diseases in plant leaves. In the tested techniques, RF achieved the most accurate result of 97.12%. However, compared to the deep learning model, CNN performed the best (98.43%) among deep learning techniques.

Dai et al. (2023) [2] This paper introduces a deep learning model called PPLC-Net. Dilated convolution, along with features like multi-level attention and GAP layers make PPLC-Net a quite distinct choice. It also incorporates the most recent weather data to increase the data sample while also improving the feature extraction generalization. The test dataset validation results show that the proposed model is 99.702% accurate with 98.442% F1 score. This model has 15.486 million parameters and 5.338 billion FLOPs. These numbers indicate that the model fulfills accurate and fast recognition requirements.

Tiwari V. et al. (2021) [4] This paper is based on a DL technique for identifying and classifying diseases in plants through their images. Results of the experiments show that the suggested model can accurately classify different leaf diseases and was able to achieve 99.58% validation accuracy and 99.19 test accuracy on various plant leaf diseases with high accuracy on complex background images.

M. Chanda et al. (2019) [14], in this paper, plant leaf disease was accurately detected and identified with high accuracy, utilizing image processing. A classification method was introduced: first, backpropagation was use to calculate the weights between nodes in neural network, and then the weights are optimized the weights utilizing Particle Swarm technique (PSO). This was able to resolve overfitting and local optima problems in conventional neural network methods. The method was able to achieve an accuracy of 96.2%.

Mengistu AD et al. (2018) [18] In this research, a combination of image processing along with another decision tree (DT) model is used. Researchers have used BPNN (Backpropagation neural network) together with DT approaches for this use case. The data was split 70% and 30% for training and testing, respectively. Final accuracy was 94.5%.

Hari SS et al. (2019) [15] In this paper, to identify plant diseases, they used the CNN model and proposed a newly built architecture called plant disease detection neural network for better disease detection. The suggested plant disease detection neural network achieved an accuracy of 86.00% using techniques like AlexNet and also Google-Net.

Lamba M et al. (2021) [5] In this paper, the authors proposed DL and ML techniques and addressed use cases where two or more categories of plant diseases exist on dissimilar datasets. The proposed method achieved great results, with 99.4% accuracy: 99.9% for two classes and 99.2% for identification from multiple classes. When compared to methods mentioned and compared in the paper, such as SMO, LibSVM etc., the proposed framework outperformed in measures like recall, specificity, sensitivity, and Matthews correlation coefficient.

Brahimi et al. (2017) [9] opted for CNN to correctly categorize their images (14,828 approx.), which included infectious tomato leaves infected with 9 different disease types. This method attained 99.18%.

Liang et al. (2019) [12] developed rice blast disease detection using a Convolutional Neural Network (CNN) in rice plant leaves and compared the performance of this model with Haar-wavelet transforms and Local Binary Pattern Histogram. This method was able to attain an accuracy of 95% results.

Sladojevic et al. (2016) [19] suggested a (DNN) deep neural network model capable of categorizing various types of plant diseases, i.e., both healthy and unhealthy leaves, from a collection of images. The image recognition model achieved a maximum accuracy of 98% and a minimum accuracy of 91%.

Soni et al. (2016) [23] was able to identify diseases in crop leaves by utilizing a different kind of classifier called probabilistic neural networks. They applied a model to identify various plant leaf diseases in images gathered randomly from the Internet database. The experiment resulted in identification with high accuracy.

Sun et al. (2017) [21] created a CNN model specifically designed and customized for the identification of diseases in plant leaves, capable of identifying twenty-six distinct types of diseases across fourteen different plants. Running test images through this model demonstrated an enhanced accuracy of 99.56%, while the recall and precision count was 99.41%.

Fang et al. (2017) [22] enhanced the TrAdaBoost technique to create a system for transfer learning which is instance based. The system aimed to resolve the issue of lack of tagged data for machine learning in the field of plant leaf disease detection. Experimental findings indicate significant improvement in performance over techniques based on KNN and SVM.

Liu et al. (2018) [16] utilized a deep similarity network (DSN) for feature extraction of healthy maize plants, then they utilized a transfer learning technique for understanding unhealthy images of the maize plant features. According to the findings, this approach can accurately identify 10 different types of normal diseased maize with a precision of 90%.

Ding et al. (2018) [17] used AlexNet to construct an 8-layer CNN model, subsequently utilizing it to train a network. The comprehension precision for 12836 images extracted from Plant Village was able to exceed the accuracy to 95% with a learning rate of 0.001.

Zhang et al. (2019) [13] fine-tuned the Google Net model on the ImageNet dataset. The study focused on examining 1200 images captured using handheld mobile phones, and they compared its results with popular image classifying methods: SVM, KNN, and Back Propagation Neural Network, Support Vector Machine and, K-Nearest Neighbors. They identified cherry leaf diseases; the recommended technique had the highest accuracy of 99.6 percent.

4. Evaluation metrics

When implementing deep learning techniques to correctly recognize and categorize different diseases in plant leaves, we can overcome the limitations associated with artificial areas of disease spot features. This results in a more objective approach to plant disease feature extraction, enhancing research efficacy and expediting technology transformation.

Authors in the literature

In their research, the authors measured their proposed methods using many measures, such as accuracy, recall, precision, F1-score, and Intersection over Union. Equations (1), (31), (32), (34), and (35) briefly define the evaluation of similar classification methods.

Accuracy: This metric explains how well a model performs in identifying a category for the input data (E.g., image). It could be important to use when all our categories have equal importance. It can be measured as below:

$$Accuracy = \frac{True\ Negative\ (TN) + True\ Positive (TP)}{(TN + TP + False\ Positive\ (FP) + False\ Negative\ (FP))}$$

Recall: calculated as the ratio of samples that are correctly identified to the no. of actual correct samples. It measures the capability of the model to detect positive sample datasets.

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Negative (FP)}$$

Precision: is a measure to know how correctly can a model classify the data. This can be calculated by taking a ratio of correctly idenfified output to the total output (both True and False positive).

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Postive}$$

$\mathbf{F1 - score}$: This metric is an alternative machine learning measure that can assess the prediction skills of a model. It incorporates both the precision and recall scores of a model.

$$\boldsymbol{F1\ Score} = \frac{\mathbf{2}}{\frac{\mathbf{1}}{\boldsymbol{Precison}} + \frac{\mathbf{1}}{\boldsymbol{Recall}}}$$

$$= \frac{\mathbf{2} \times \boldsymbol{Precison} \times \boldsymbol{Recall}}{\boldsymbol{Precison} + \boldsymbol{Recall}}$$

Intersection over Union (IoU)

IoU = popular metric for measuring localization error and assessing localization accuracy in object detection models.

$$IoU = \frac{True\ Positive}{(False\ Negative + False\ Positive + True\ Positive)} \times 2$$

5. Results and Discussions

Advanced agricultural plant leaf disease image recognition techniques are reviewed in the above sections with a focus on deep learning (DL) and transfer learning (TL). Table 1 outlines the most relevant elements of these two techniques to make a more natural comparison.

Table 1: Comparison between deep learning and Machine learning

Factors	**Deep Learning**	**Machine learning**
Learning technique	Mimics human brains, Complex neural networks	Dependent on algorithms and statistical models
Model construction	End to end learning and building	Feature engineering required
Model Complexity	High-complexity	Simple
Modeling time	Longer	Shorter

Training sample size	Large	Small to medium
Data distribution	Same	Different
Generalization	Weak	strong

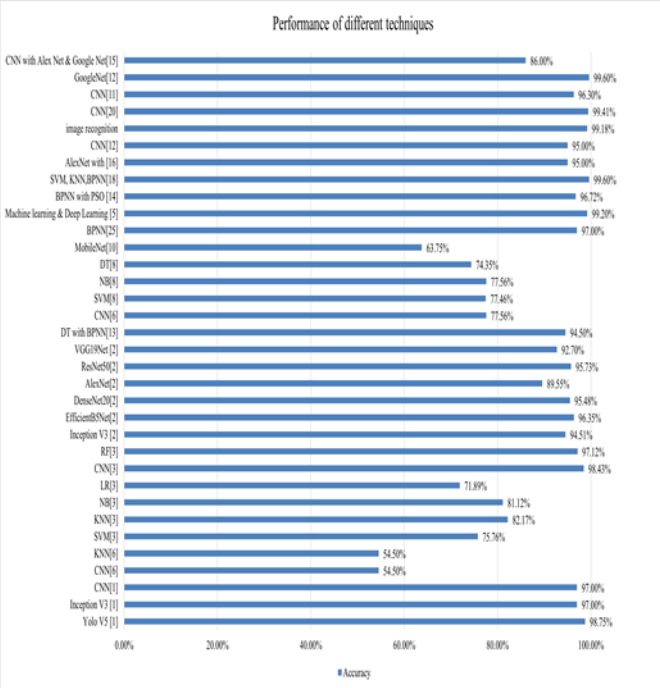

Fig 4: Performances of the different techniques for detecting diseases in plant leaves

The reviewed papers' analysis showed that detecting and classifying plant diseases are crucial problems to be solved that require further investigation. Pre-trained techniques provided higher accuracy compared to using just machine learning techniques. However, machine learning and deep learning together could result in even better accuracy.

6. Conclusion

The latest innovative technologies for the plant leaf disease detection with image recognition process were examined in this paper, focusing mostly on transfer learning and deep learning.

- Growing usage of intelligent mobile terminals suggests that lightweight model creation should be a priority for future studies. Several studies, including Inception V3[1], YOLO [1], and Google Net [2], have been conducted to address this issue. These lightweight models make them appropriate for mobile users or edge computing in real applications, although additional research is needed.
- All current techniques only use homogeneous datasets (images), but none have utilized heterogeneous datasets (images, text, video, audio, etc.). This could feed the learning process with additional features like expert opinion through text for better accuracy. That way, we could ensure to utilize the information and knowledge from different types of data like images, text, video etc. can be helped with learning and modeling in the target domain.

Acknowledgments

The authors hereby declare that they have no competing interest with regard to the submitted article.

This research work did not receive any funding.

References

[1] Tirkey D, Singh KK, Tripathi S. Performance analysis of AI-based solutions for crop disease identification detection and classification. Smart Agric Technol. 2023. https://doi.org/10.1016/j.atech.2023.100238.

[2] Dai G, Fan J, Tian Z, Wang C. PPLC-Net: neural network-based plant disease identification model supported by weather data augmentation and multi-level attention mechanism. *J King Saud Univ Comput Inf Sci. 2023;35(5):101555. https://doi.org/10.1016/j.jksuci.2023.101555.

[3] Kirola M, Joshi K, Chaudhary S, Singh N, Anandaram H, Gupta A. Plants diseases prediction framework: an image-based system using deep learning. *Proc IEEE World Conf Appl Intell Comput. 2022. https://doi.org/10.1109/AIC55036.2022.9848899.

[4] Tiwari V, Joshi RC, Dutta MK. Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images. Ecol Inform. 2021;63: 101289. https://doi.org/10.1016/j.ecoinf.2021.101289.

[5] Lamba M, Gigras Y, Dhull A. Classification of plant diseases using machine and deep learning. Open Comput Sci. 2021;11(1):491–508. https://doi.org/10.1515/comp-2020-0122.

[6] A. Abbas, S. Jain, M. Gour, and S. Vankudothu. Tomato plant disease detection using transfer learning with C-GAN synthetic images. 2021

[7] S. Mathulaprangsan, K. Lanthong, D. Jetpipattanapong, S. Sateanpattanakul and S. Patarapuwadol, "Rice Diseases Recognition Using Effective Deep Learning Models," 2020 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT &

NCON), Pattaya, Thailand, 2020, pp. 386-389, doi:10.1109/ECTIDAMTNCON48261.2020.9090709.

[8] Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017). Deep Learning for Tomato Diseases: Classification and Symptoms Visualization. *Applied Artificial Intelligence*, *31*(4), 299–315. https://doi.org/10.1080/08839514.2017.1315516.

[9] Wan H, Lu Z, Qi W, Chen Y. Plant disease classification using deep learning methods. In: 4th International Conference on Machine Learning and Soft Computing. 2020. p.5-9.

[10] Agarwal M, Singh A, Arjaria S, Sinha A, Gupta S. ToLeD: tomato leaf disease detection using convolution neural network. Procedia Comput Sci. 2020;167(2019):293–301. https://doi.org/10.1016/j.procs.2020.03.225.

[11] Wang Y, Zhang H, Liu Q, Zhang Y. Image classification of tomato leaf diseases based on transfer learning. *Journal of China Agricultural University. 2019;24(6):124-130.

[12] Liang W, Zhang H, Zhang G, Cao H. Rice blast disease recognition using a deep convolutional neural network. *Scientific Reports. 2019;9(1):2869.

[13] Zhang K, Zhang L, Wu Q. Identification of cherry leaf disease infected by pedosphere pannosa via convolutional neural network. International Journal of Agricultural and Environmental Information Systems. 2019;10(2):98-110.

[14] Chanda M, Biswas M. Plant disease identification and classification using Back-Propagation Neural Network with Particle Swarm Optimization. 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI). Tirunelveli, India, 2019. https://doi.org/10.1109/ICOEI.2019.8862552.

[15] Hari SS, Sivakumar MM, Renuga DP, Karthikeyan S, Suriya S. Plant disease detection by leaf image classification using convolutional neural network. Comput Sci Commun Inf. 2019. https://doi.org/10.1007/978-981-15-1384-8_16

[16] Liu Y, Lei B, Cao Y, Tang J, Hu L. Maize diseases identification based on deep convolutional neural network. *Chinese Agricultural Science Bulletin. 2018; 34(36):165-170.

[17] Ding R, Zhou P. Identification of typical crop leaf diseases based on convolutional neural network. Packaging Journal. 2018;10(06):79-85.

[18] Mengistu AD, Mengistu SG, Melesew D. An automatic coffee plant diseases identification using hybrid approaches of image processing and decision tree. Indones J Electr Eng Comput Sci. 2018;9(3):806–811. https://doi.org/10.11591/ijeecs.v9.i3.pp806-811.

[19] Brahimi M, Boukhalfa K, Moussaoui A. Deep learning for tomato diseases: classification and symptoms visualization. *Applied Artificial Intelligence. 2017;31(4):299-315.

[20] [20] Sladojevic S, Arsenovic M, Anderla A, Culibrk D, Stefanovic D. Deep neural networks-based recognition of plant diseases by leaf image classification. Computational Intelligence and Neuroscience. 2016; 2016:1-11.

[21] Sun J, Tan W, Mao H, Wu X, Chen Y, Wang L. Recognition of multiple plant leaf diseases based on improved convolutional neural network. Transactions of the Chinese Society of Agricultural Engineering. 2017;33(19):209-215.

[22] Fang S, Yuan Y, Chen L, Zhang J, Li M, Song S. Crop disease image recognition based on transfer learning. In: International Conference on Image and Graphics. 2017. p.54-554.

[23] Soni P, Chahar R. A segmentation improved robust PNN model for disease identification in different leaf images. In: 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control, and Energy Systems. 2016. p.1-5.

[24] Pujari J, Yakkundimath R, Byadgi AS. SVM and ANN-based classification of plant diseases using feature reduction technique. International Journal of Interactive Multimedia and Artificial Intelligence. 2016;3(7):6-14.

[25] Ramakrishnan M, Sahaya ANA. "Groundnut leaf disease detection and classification by using back probagation algorithm", 2015 Int. Conf Commun Signal Process ICCSP. 2015; 2015:964–8. https://doi.org/10.1109/ICCSP.2015.7322641.

[26] Hari SS, Sivakumar MM, Renuga DP, Karthikeyan S, Suriya S. Plant disease detection by leaf image classification using convolutional neural network. Comput Sci Commun Inf. 2019. https://doi.org/10.1007/978-981-15-1384-8_16

[27] Zhou R, Kaneko S, Tanaka F, Kayamori M, Shimizu M. Early detection and continuous quantization of plant disease using template matching and support vector machine algorithms. In: The First International Symposium on Computing and Networking - Across Practical Development and Theoretical Research. 2013. p.300-304.

[28] Yu H, Chen G, Liu D. A simplified Bayesian network model applied in crop or animal disease diagnosis. In:

Computer and Computing Technologies in Agriculture II, Volume 2. IFIP Advances in Information and Communication Technology. 2008. p.1001-1009.

[29] Hihi SE, Bengio Y. Hierarchical recurrent neural networks for long-term dependencies. In: Advances in Neural Information Processing Systems. 1995. p.493-499.

[30] LeCun Y, Boser BE, Denker JS, Henderson D, Howard RE, Hubbard WE, Jackel LD. Back-propagation applied to handwritten zip code recognition. Neural Computation. 1989;1(4):541-551.

[31] Kempegowda B, Rao M. Tomato plant leaves disease classification using KNN and PNN. International Journal of Computer Vision and Image Processing. 2019;9(1):51-63.

[32] Krishnan VG, Deepa J, Rao PV, Divya V, Kaviarasan S. An automated segmentation and classification model for banana leaf disease detection. J Appl Biol Biotechnol. 2022;10(1):213–20. https://doi.org/10.7324/JABB.2021.100126.

[33] S Mathulaprangsan K Lanthong S Patarapuwadol. 2020. Rice Diseases Recognition Using Effective Deep Learning Models. Telecommun. Eng Media Technol with ECTI North Sect Conf Electr Electron Jt Int Conf Digit Arts Comput. https://doi.org/10.1109/ECTIDAMTNCON48261.2020.9090709

[34] Heltin Genitha C, Dhinesh E, Jagan A. Detection of leaf disease using principal component analysis and linear support vector machine. Adv Comput ICoAC Proc Int Conf. 2019. https://doi.org/10.1109/ICoAC48765.2019.246866.

[35] P. S. Gupta, P. Hans, and S. Chand. 2020. Classification Of Plant Leaf Diseases Using Machine Learning and Image Preprocessing Techniques.