International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Markov Clustering in Leaf Disease Detection Based on Classification using Probabilistic Naïve Bayes Regression for Deep Learning Architecture

Mr. Ashwini Kumar[1], Dr. M Purushotham Reddy[2], Dr. C. Sushama[3], Suresh Limkar[4], Kuldeep Singh Kaswan[5], Dr. Mansing Rathod[6]

Submitted: 03/11/2022 **Accepted:** 30/01/2023

Abstract: Destructive insects and plant leaf diseases pose serious problems for the agricultural industry. A quicker and better accurate forecasting of crop leaf infections could aid in the creation of an early treatment method and significantly lower financial losses. Researchers have been able to significantly increase the performance and accuracy of object identification and recognition systems because to recent advances in deep learning. This study offers an innovative methodology in leaf disease detection relied upon clustering plus classification employing deep learning architectures. Here the input dataset has been taken as leaf disease as well as processed for noise removal, smoothening. Then the processed image has been clustered using markov clustering and classified using probabilistic naïve bayes regression based deep learning architecture. The peach plants' leaves were obtained for the trials in this study using a publicly accessible dataset called PlantVillage. For several datasets, There has been experimental assessment in perspective of accuracy, precision, recall, F-1 score, RMSE, and MAP. The recommended approach accomplished accuracy of 95%, precision of 85%, recall of 77%, F-1 score of 73%, RMSE of 65% and MAP of 55%.

Keywords: *leaf disease detection, clustering, classification, deep learning, accuracy.*

1. Introduction

It is getting harder and harder to identify plant diseases from their visible symptoms on plant leaves. Because of its intricacy, the sheer volume of produced crops, as well as the resulting phytopathological concerns, even experienced plant pathologists and agricultural specialists frequently struggle to correctly diagnose some diseases, which leads in inaccurate conclusions and unsatisfactory solutions [1]. Amateurs in the agricultural process can greatly benefit from an automated system created to aid in identification of plant diseases by appearance of plant and visual symptoms. Farmers will find this to be a valuable strategy that will warn them just in time to prevent the disease from spreading over a vast area. For disease detection, the image processing methods are suitable and efficient with the help of plant leaf images. Though continuously monitoring of health and disease detection of plant increase the quality and quantity of the yield, it is costly. Machine learning algorithms are experimented due to their better accuracy [2]. Additionally, the accuracy of Convolutional Neural Network (Alexnet) is assessed and contrasted. Artificial intelligence-based applications for learning have produced useful results. A system is trained using machine learning techniques so that it can learn autonomously and improve its performance over time [3]. It has frequently been noted that plant diseases are challenging to manage since their populations vary depending on the environment. Plants can get a variety of diseases, including bacterial, viral, and fungal ones. According to research, 85% of plants are harmed by organisms that resemble fungi. Farmers in poor nations utilise a traditional approach that takes more time and labour to complete [4].

[1] Assistant Professor, Department of Engineering & IT, Arka Jain University, Jamshedpur, Jharkhand, India.
Id-ashwini.kumar@arkajainuniversity.ac.in,
0000-0003-3558-8054
[2] Professor, Department of Information Technology, Institute of Aeronautical Engineering, Hyderabad, India.
purushotham.mps@gmail.com
0000-0002-0899-1609
[3] Associate professor, Department of CSE, Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India.
sushama.c@vidyanikethan.edu
0000-0003-4560-2964
[4] Associate Professor, Artificial Intelligence & Data Science, AISSMS Institute of Information Technology, Pune, India.
sureshlimkar@gmail.com
0000-0001-7837-9100
[5] Professor, Department of Computer Science and Engineering, Galgotias University, Greater Noida, Uttar Pradesh, India.
kuldeep.kaswan@galgotiasuniversity.edu.in
0000-0003-0876-0330
[6] Assistant Professor, Information Technology, K. J. Somaiya Institute of Engineering and Information Technology, Mumbai, India.
rathodm@somaiya.edu
0000-0002-5860-6617

The following is the significance of this study:

1. To suggest a unique strategy in leaf disease detection based on clustering and classification utilizing DL architectures.
2. Processed image has been clustered using markov clustering and classified using probabilistic naïve bayes regression based deep learning architecture.

2. Literature Review

We have just taken into account leaf-related disorders in this area. Artificial neural network (ANN) model for Phyllanthus Elegant Wall leaf diseases classification into two classes, such as either well or ill, was presented in work [5]. They used image processing techniques to change the colour scheme of the photos of herb plants. Based on the colour and size of the leaf, the photographs are categorised. In work [6], the linear SVM utilised for classifying leaf diseases was reported. The input photos of the grapes and the diseased regions are processed using preprocessing techniques. These are recognised with the aid of clustering algorithms, which also allow for the extraction of texture and colour data. Convolution neural network (CNN) technology was utilised in study [7] to identify leaf disease. They made utilisation images from a substantial dataset that contained both diseased as well as healthy plant leaves. They have worked with coloured, grayscaled, and segmented datasets, among others. This CNN model can quickly recognise the 26 illnesses that affect 14 crop species. Similar to this, work [8] used SVM to identify illnesses in soybean leaf. Based on the shape of the diseased plant, this method used the scale-invariant feature transform technique to automatically identify the disease. This offers the farmer online assistance with little effort. For automated leaf disease classification, author [9] employed a genetic algorithm. To categorise the diseases, the input image is first preprocessed and segmented using a genetic algorithm. Eleven statistical features and the Support Vector Machine classifier (SVM) were utilised in work [10] to identify the disorders. This has benefits that make the detection, identification, and classification process more effective. In terms of disease classification, it provides 93% accurate findings [11]. Monzurul Islam and colleagues combined image processing with machine learning to enable disease diagnosis from potato plant leaf photos. The disease classification accuracy utilising Color thresholder, GLCM, and multiclass SVM was 95% [12].

3. System Model

This passage discuss innovative methodology in leaf disease detection depending upon clustering and classification using deep learning architectures. Furthermore, the source dataset was chosen to represent leaf disease and was then treated to remove noise, smoothening. Then the processed image has been clustered using markov clustering and classified using probabilistic naïve bayes regression based deep learning architecture. Figure 1 depicts the planned architectural design.

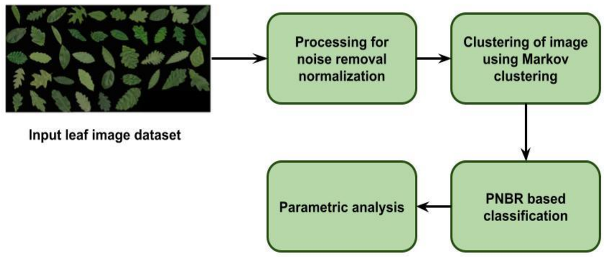

Fig. 1: general suggested structure

There are two ways to convert RGB to grayscale. These techniques have benefits and drawbacks. These are what they are: 1. Standard technique: It is a simple approach that takes the RGB average of the three colours into account. Grayscale = 0.3R + 0.59G + 0.11B is used in this method to calculate grayscale images.

markov clustering:

Let's now see how sii and sij changed as actions carried during during iteration t + 1 of the traditional Markov clustering process were carried out:

$$s_{ij}'' = \frac{b'}{b'+(v-1)a'} \text{ and } s_i'' = \frac{a'}{b'+(v-1)a'} \quad (1)$$

for any $i, j = 1,2, \dots, v$ with $i = j$;

Expansion raises matrix Σ'' to second power, giving $\Sigma^* = [s_i^*]$ with eq. (2)

$$s_i'' = \frac{b^2+(v-1)a^2}{|b^t+(v-1)a'|^2} \quad (2)$$

$s_i' = b'$ and $s_j' = a'$

$$\Leftrightarrow \frac{b^{2r}/a^{2v} + (v-1)}{2b'/a' + (v-2)} > \frac{b}{a}$$

$$s_i'' = \frac{b^2+(v-1)a^2}{|b^t+(v-1)a'|^2} \quad (3)$$

$$\Leftrightarrow \frac{p^2 + (v-1)}{2p^2 + (v-2)} > p$$

probabilistic naïve bayes regression:

By using the Parzen window approach, conditional density estimation in the PNN is made possible.

$$P(x \mid C_k) = \sum_{n=L_i} P(s, x) \quad (4)$$

The kernel density estimator (KDE), which underlies the functioning of the PNN classifier, often takes the following by eq (5)

$$\hat{f}(\mathbf{x}) = \frac{1}{Ph} \sum_{j=1}^{P} K\left(\frac{\mathbf{x}-\mathbf{x}^{(p)}}{h}\right) \quad (5)$$

$$K(\mathrm{x}) = K(x_1) \cdot K(x_2) \cdot \ldots \cdot \mathcal{K}(x_N).$$

$$K(x_i) = \frac{2}{\pi(x_i^2+1)^2} \Bigg) \quad (6)$$

We employ the second strategy, thus the output of the summation neuron is described by the following equation (7)

$$f_j(\mathbf{x}) = \frac{1}{P_j \det(\mathbf{h})} \sum_{j=1}^{P_j} \frac{1}{s_j^N} K(x' \frac{\left(\mathbf{x}_j^{(p)}\right)^T \mathbf{h}^{-1}}{s_p} \quad (7)$$

If we use equations (7) and (8) to represent the pattern neuron activation function, KDE for the PNN's jth class equals eq (8)

$$\frac{\partial \hat{f}_j\left(\mathbf{x}_j^{(p)}\right)}{\partial x_{j,i}^{(r)}} = \frac{1}{P_j \det(\mathbf{h})} \frac{1}{s_r^N} \frac{\partial}{\partial x_{j,i}^{(r)}} K\left(\frac{\left(\mathbf{x}_j^{(p)} - \mathbf{x}_j^{(r)}\right)^T \mathbf{h}^{-1}}{s_r}\right) \quad (8)$$

When KDE is chosen based on product kernel (7), it is advised to calculate h using the plug-in approach. The following eq is used to calculate each coefficient hi (10)

$$h = \left[\frac{R(\mathcal{K})}{U(\mathcal{K})^2} \frac{8\sqrt{\pi}\tilde{\sigma}^9}{3P}\right]^{\frac{1}{8}}$$
$$R(\mathcal{K}) = \int_{R^N} \mathcal{K}(x^2)dx$$
$$U(\mathcal{K}) = \int_{R^N} x^2 \mathcal{K}(x)dx, \quad (10)$$

Following the plug-in technique computation of the temporary values of the smoothing parameter vector, the KDE quantities are calculated in accordance (11)

$$s_p = \left(\frac{\hat{f}(\mathbf{x}^{(p)})}{\tilde{s}}\right)^{-c} \quad (11)$$

$$\frac{\partial \hat{f}_j\left(\mathbf{x}_j^{(p)}\right)}{\partial x_{j,i}^{(r)}} = \frac{1}{P_j \det(\mathbf{h})} \frac{1}{s_r^N} \mathcal{K}\left(\frac{x_{j,1}^{(p)} - x_{j,1}^{(r)}}{h_1 s_r}\right) \ldots \quad (12)$$

The elements of the rth column in (14), computed for a certain input pattern p, indicate the sensitivities of KDE in jth class with respect to each rth pattern neuron. Since eq. (14) must be used to derive the following gradient because each item of Sj's denominator is a vector (15)

$$\nabla \hat{f}_j^{(p,r)} = \frac{\partial \hat{f}_j\left(\mathbf{x}_j^{(p)}\right)}{\partial \mathbf{x}_j^{(r)}} = \left[\frac{\partial \hat{f}_j\left(\mathbf{x}_j^{(p)}\right)}{\partial x_{j,1}^{(r)}}, \ldots, \frac{\partial \hat{f}_j\left(\mathbf{x}_j^{(p)}\right)}{\partial x_{j,N}^{(r)}}\right] \quad (14)$$

Where

$$\frac{\partial \hat{f}_j\left(\mathbf{x}_j^{(p)}\right)}{\partial x_{j,i}^{(r)}} = \frac{1}{P_j \det(\mathbf{h})} \frac{1}{s_r^N} \frac{\partial}{\partial x_{j,i}^{(r)}} K\left(\frac{\left(\mathbf{x}_j^{(p)} - \mathbf{x}_j^{(r)}\right)^T \mathbf{h}^{-1}}{s_r}\right) \quad (15)$$

$$\frac{\partial}{\partial x_{j,i}^{(r)}} \mathcal{N}^*\left(x_{j,i}^{(p)}\right) = \frac{8\left(x_{j,i}^{(p)} - x_{j,i}^{(r)}\right)}{\pi h_i^2 s_r^2 \left(\left(\frac{x_{j,i}^{(p)} - x_{j,i}^{(r)}}{h_i s_r}\right)^2 + 1\right)^3} \quad (16)$$

$$\frac{\partial}{\partial x_{j,i}^{(r)}} \mathcal{K}^*\left(\frac{x_{j,i}^{(p)} - x_{j,i}^{(r)}}{h_i s_r}\right) \ldots \mathcal{K}\left(\frac{x_{j,N}^{(p)} - x_{j,N}^{(r)}}{h_N s_r}\right) \quad (17)$$

Since the PNN is a parallel Bayesian classifier, it is not trained iteratively. As a result, PNNs train orders of magnitude quicker than other multilayer feedforward neural network training paradigms. As a result, it is simpler to construct and contains fewer parameters that are crucial for maximising network performance.

4. Experimental Analysis

All of the models compiled for this investigation included GPU support. The 64-bit Debian GNU/Linux 9.11 operating system, an Intel (R) Xeon (R) Gold CPU clocked at 2.20GHz, 16 GB of RAM, and an NVIDIA Tesla K80 with 12 GB of memory were used for all experimental experiments, which were performed in a cloud setting provided by Google. All scripts are implemented with the aid of the Python-based open source Keras 2.3.1 framework for deep neural networks.

Dataset description: The dataset employed in this study, known as PlantVillage, contains 38 classes and 54,305 photos of a total of 14 different plant species, 12 are in good health, whereas 26 have diseases (Hughes and Salathe, 2015). Dataset contains coloured images of various sizes. One more class in the collection identifies 1143 backdrop photos. Consequently, there are 55,448 photos in the collection as a whole.

The ImageNet collection has 1000 class categories and about 1.2 million images. The DL methods employed in this study's training process were sped up by using pre-trained weights from the most recent CNN models on ImageNet. Every simulation has 1000 outputs had last FC layers of them adjusted to 39 outputs to better fit the issue. Pre-trained models had all of their layers set to trainable. In the last layer, categorical cross entropy was chosen as loss function as well as Softmax as activation function. The early halting approach was applied during training, with patience set to 5 and the lowest variation in loss set to 1e-3. The same optimization methodology utilised for the training of the ImageNet dataset was applied to pre-trained models. As a result, all other models employ the Adam optimization approach, whereas the VGG16 model uses the SGD optimization method. Additionally, the SGD method's learning rate was set to 0.01 while the Adam method's was 0.001. For all models, the validation step was set to 1.

Table-1 Assessment of the suggested and current methods predicated upon a variety of leaf image datasets

Parameters	CNN	SVM	MCLDD_NBR_DL
Accuracy	89	92	95
Precision	81	83	85
Recall	72	75	77
F1_Score	65	69	73
RMSE	61	63	65
MAP	51	53	55

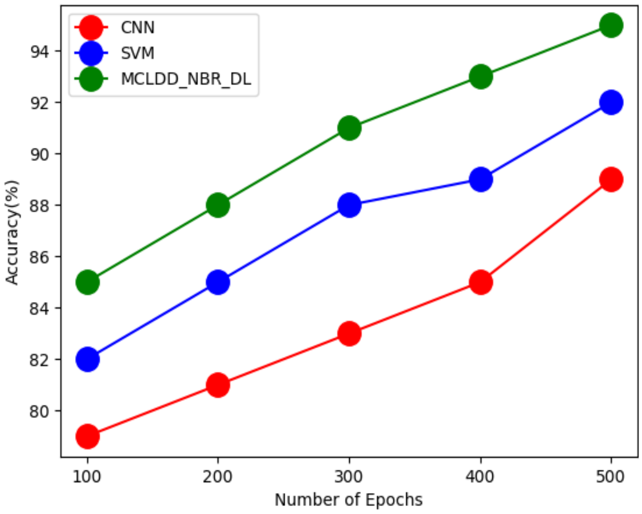

Fig. 2 Comparison of accuracy

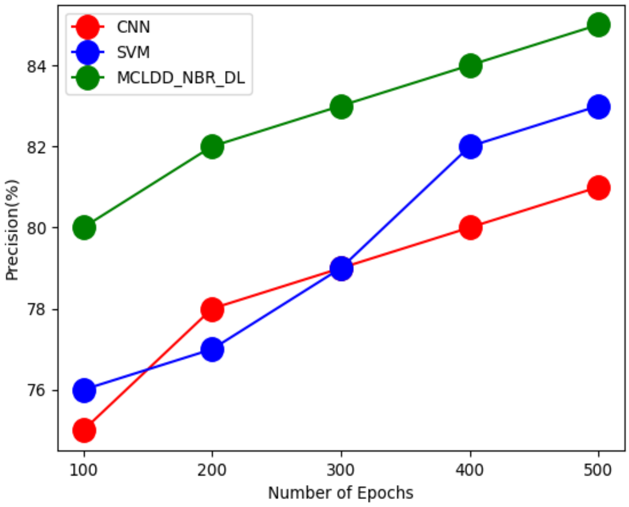

Fig. 3 Comparison of precision

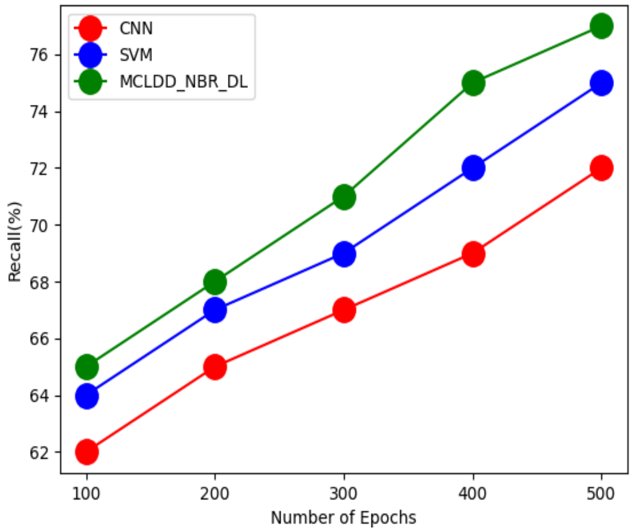

Fig. 4 Comparison of recall

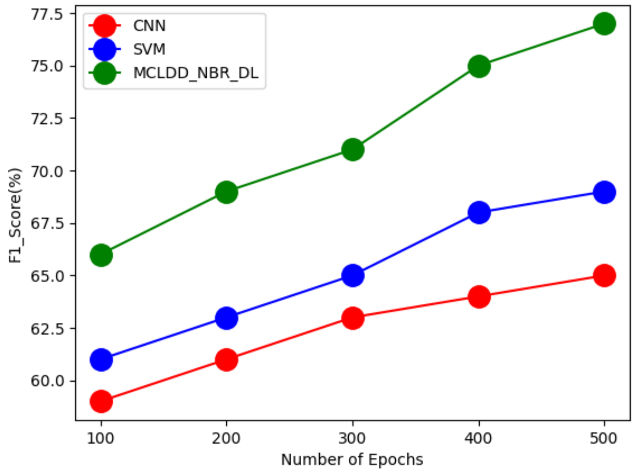

Fig. 5 Comparison of F-1 score

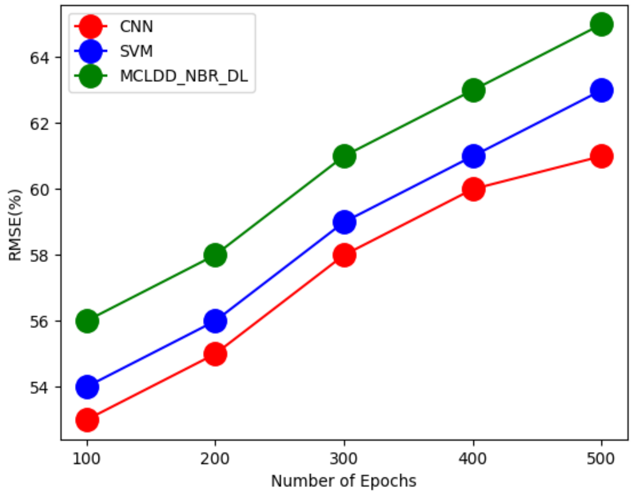

Fig. 6 Comparison of Recall

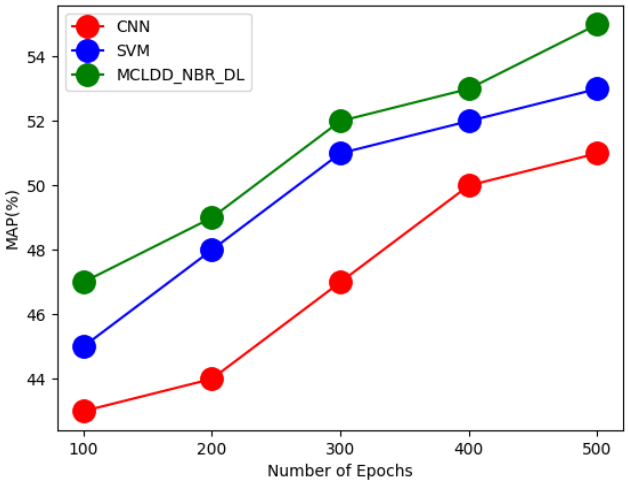

Fig. 7 Comparison of MAP

Table-1 and figure 2-7 compares and contrasts the suggested and current methods for various leaf image dataset. Here parameters analysed are accuracy, precision recall, F-1 score, RMSE and MAP. The recommended approach achieved accuracy of 95%, precision of 85%, recall of 77%, F-1 score of 73%, RMSE of 65% and MAP of 55%; while existing CNN attained accuracy of 89%, precision of 81%, recall of 72%, F-1 score of 65%, RMSE of 61% and MAP of 51%, SVM accomplished accuracy of 92%, precision of 83%, recall of 75%, F-1 score of 69%, RMSE of 63% and MAP of 53%.

5. Conclusion

This study offers an innovative method in leaf disease detection by classification utilizing DL architectures. Processed image has been clustered using markov clustering and classified using probabilistic naïve bayes regression based deep learning architecture. Multiple phases of categorization are carried out to rule out alternatives at each level, improving forecast accuracy. It takes constant examination of the plants to spot infections in the leaves. This ongoing inspection of the plants requires a lot of human labour and takes a lot of time. To put it simply, monitoring the plants requires some form of planned method. An analysis of the experimental data has been done in perspective of accuracy, precision, recall, F-1 score, MAP, RMSE. Accuracy of 95%, precision of 85%, recall of 77%, F-1 score of 73%, RMSE of 65% and MAP of 55% accomplished using the suggested approach. These models can be applied in the future utilising a variety of high-dimensional data sets and different categorization techniques.

Reference:

[1] Szilágyi, L., & Szilágyi, S. M. (2016). A modified two-stage Markov clustering algorithm for large and sparse networks. *computer methods and programs in biomedicine*, *135*, 15-26.

[2] Kusy, M., & Kowalski, P. A. (2018). Weighted probabilistic neural network. *Information Sciences*, *430*, 65-76.

[3] Tiwari, V., Joshi, R. C., & Dutta, M. K. (2021). Dense convolutional neural networks based multiclass plant disease detection and classification using leaf images. *Ecological Informatics*, *63*, 101289.

[4] Goyal, L., Sharma, C. M., Singh, A., & Singh, P. K. (2021). Leaf and spike wheat disease detection & classification using an improved deep convolutional architecture. *Informatics in Medicine Unlocked*, *25*, 100642.'

[5] Appalanaidu, M. V., & Kumaravelan, G. (2021). Plant leaf disease detection and classification using machine learning approaches: a review. *Innovations in Computer Science and Engineering*, 515-525.

[6] Sambasivam, G., & Opiyo, G. D. (2021). A predictive machine learning application in agriculture: Cassava disease detection and classification with imbalanced dataset using convolutional neural networks. *Egyptian Informatics Journal*, *22*(1), 27-34.

[7] Wang, Y., Wang, H., & Peng, Z. (2021). Rice diseases detection and classification using attention based neural network and bayesian optimization. *Expert Systems with Applications*, *178*, 114770.

[8] Vasavi, P., Punitha, A., & Rao, T. V. N. (2022). Crop leaf disease detection and classification using machine learning and deep learning algorithms by visual symptoms: A review. *International Journal of Electrical and Computer Engineering*, *12*(2), 2079.

[9] Syed-Ab-Rahman, S. F., Hesamian, M. H., & Prasad, M. (2022). Citrus disease detection and classification using end-to-end anchor-based deep learning model. *Applied Intelligence*, *52*(1), 927-938.

[10] Ashwinkumar, S., Rajagopal, S., Manimaran, V., & Jegajothi, B. (2022). Automated plant leaf disease detection and classification using optimal MobileNet based convolutional neural networks. *Materials Today: Proceedings*, *51*, 480-487.

[11] Singh, A., & Kaur, H. (2021). Potato Plant Leaves Disease Detection and Classification using Machine Learning Methodologies. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1022, No. 1, p. 012121). IOP Publishing.

[12] Prabu, M., & Chelliah, B. J. (2022). Mango leaf disease identification and classification using a CNN architecture optimized by crossover-based levy flight distribution algorithm. *Neural Computing and Applications*, *34*(9), 7311-7324.