2024 OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 4.0 | 979-8-3503-7378-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/OTCON60325.2024.10688104

Detection and categorization of diseases affecting plant leaves with the use of machine learning

1st Ashish Nagila
Department of CSE
IFTM University,
Moradabad, Uttar Pradesh, India
ashishnagila01@gmail.com

2nd Abhishek K Mishra
Department of CSE
IFTM University,
Moradabad, Uttar Pradesh, India
abhimishra2@gmail.com

***Abstract*- The use of sensors and various technologies based on machine learning is called "smart agriculture" and has now become an agricultural trend. According to a recent study, foliar disease caused major loss to 56% of the farm. It is important to increase agricultural productivity and monitor the spread of diseases. Early detection and prevention of the disease is important to prevent its spread. Therefore, we can use some research to solve this problem. In this study, we compare two different models: Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). The two models discussed and analyzed in this study can describe eight different diseases. Using the soyabean leaf disease dataset as its training set, the SVM model achieved a 80% accuracy rate, surpassing that of the KNN model, which achieved 64% accuracy rates, respectively.**

***Keywords*- SVM, KNN, Machine Learning**

I. INTRODUCTION

Feeding India's massive population is no easy feat, given that the country is the second biggest in the world. A food crisis and a dramatic increase in food prices are also happening at the same time. The primary reason for the scarcity is the proliferation of crop diseases, which disturbs farming practices, wears down soil fertility, and ultimately leads to crop failure. The root causes of leaf diseases are bacteria and fungi. In order to stop the spread of a disease and stop it from getting worse, early detection is crucial. More time to treat the disease and avoid crop loss is available if it is detected earlier.

The plant's longevity might be compromised by the infections in its leaves, shortening its life expectancy to just two or three years. Inedible seeds can be produced when a plant becomes sick and its reproduction rate is impaired. Because of these seeds, the soil becomes unsuitable for plant growth. Fresh plant sown in the field have also been affected by disease, which has been inherited through generations in the soil and has caused crop failure. In the past, we would have to wait until the entire crop was sick before we could determine which disease had infected the plants. We need to implement cutting-edge tech like AI and ML to stop crop failure. We will provide a dataset for training and testing the algorithms, and we will write the code using SVM, KNN, and CNN approaches.
Because there are so many labels that need to be sorted, this is an example of a multi-class- classification problem.

II. FOLLOWING DISEASES ARE BEING CATEGORIZED

A. Bacterial blight

This prevalent disease disproportionately affects soybeans when the weather is cool and wet. The infection can be transmitted through seeds, and the disease is typically found at low levels.

B. Brown Spot

Inconsistent watering and fungal or bacterial infections on the leaves generate this disease. The disease is identified by the appearance of multiple large spots.

C. Downy mildew

This fungal-like organism causes downy mildew, a leaf disease. Spores in the air allow it to travel from one plant to another. Prolonged leaf moisture aids the infection, hence it is wet-weather sickness. When a leaf is free of disease, it can be classified as belonging to the "Healthy" class, which includes a set of healthy leaves.

D. Soybean Mosaic Virus

This disease emerges in winter and disappears in summer. It is likely that all herbaceous perennials are susceptible to this fungus. Both aphids and seeds can transmit the virus. If the virus is viable and can spread, then conditions that encourage the growth of aphids can also make this disease more prevalent.

III. RELATED WORK

This study presents the authors' findings from the literature review regarding the detection and classification of leaf diseases using various models and methods.
Some deep learning-based solutions have been proposed for the detection and identification of soybean crop pests, as described in [1]. We examined the results of the transfer learning (TL) model to evaluate the efficiency and reliability of the proposed method for calculating known lines and visual accuracy. The proposed method achieves 98.75% accuracy on YoloV5, 97% accuracy on InceptionV3, and 97% accuracy on CNN (in that order). Among them, the YoloV5 algorithm is suitable for instant search due to the efficiency of its resolution and the ability to operate at 53 frames per second. In addition, a database of crop insects was created by combining images taken using different devices. This search method makes the producer's job easier because the search method is rare and gives better results. The authors of [2] proposed the classification and detection of leaf diseases using deep

learning. All photos in the PlantVillage online collection. In the proposed method, they used CNN to classify leaf diseases. Twelve groups deal with plant diseases (bacteria, fungi, etc.), and three groups deal with healthy leaves. A total of fifteen classes participated. As a result, they are able to achieve very high accuracy in the training and testing phase, with 98.29% training accuracy and 98.029% testing accuracy across all data.

Studies in the literature [3] have suggested a good method to identify and describe rice diseases by analyzing the size, shape and color of lesions in leaf images. The proposed model uses image binarization to remove image noise from Otsu's global thresholding technology. The scheme was plugged into CNN and trained using 4,000 images of healthy and diseased foods to identify all three diseases. The results show that the proposed CNN method is efficient and fast, achieving 99.7% data accuracy. This method is light years ahead of other methods used to detect and isolate plant pathogens. In [4], the authors present a CNN-based model that can detect and classify tomato leaf diseases using publicly available data. Enhance that with photos taken at State Farm. To avoid overfitting, a different correlation coefficient is used to create the model compared to the training data. The results showed that the proposed model achieved more than 99% accuracy on both training and testing data and performed well in detecting and classifying tomato leaves. The authors of

[5] identified four diseases, two infectious diseases, two fungal diseases and one related disease in the "PlantVillage" database. Images of healthy leaves for all twelve crops are also included. Machine learning techniques such as support vector machine (SVM), convolutional neural networks (CNN), and gray-scale co-occurrence matrices (GLCM) are used to develop predictive models. Advances in backpropagation of artificial neural networks have coincided with the development of artificial intelligence. KMC study has also been conducted to diagnose diseases using rapid leaf imaging. Finally, the proposed method determined an accuracy rate of 98% for apples, 99% for rice trees, and 96%, 94%, 95%, and 97% for tomatoes, respectively. This work uses precision, recall, and f-test to evaluate the multi-class problem where each class has symptoms. To detect rice diseases, the authors of [6] suggested using the CNN development method. Image classification is a particularly useful task for DNNs. This work demonstrates how DNNs work in the context of image classification for plant disease diagnosis. Finally, this study compares and evaluates the accuracy of existing methods: 80% for TL, 85% for CNN+TL, 90% for ANN, and 95% for ECNN+GA. The paper cited in [7] discusses various ML and DL methods. This research is dedicated to predicting plant diseases using machine learning methods such as SVM, KNN, RF, LR and CNN. The next step is to compare ML and DL methods. Although the DL model presented in this study is the most accurate model, the CNN method surpasses all other ML methods and reaches an accuracy rate of 98.43%. However, RF ranks second.

IV. *SOME TECHNIQUES OF MACHINE LEARNING*

A. Support Vector Machine

A basic machine learning model, support vector machines (SVMs) use the idea of kernels to handle non-linear data and, when dealing with linear data, generate the best hyperplane. With the input photographs in hand, the model determines how many features are retrievable from the data. A number of hyperplanes are used to segment the data into categories before it is shown [8]. The support vector machine (SVM) kernel is a useful tool for non-linear data, as it uses a hyperplane to transform data from a lower-dimensional space to a higher-dimensional one. The two main categories of kernels applicable to this data are as follows: first, those that provide the most accurate results, and second, those that yield less accurate results. To find the best kernel to use with this data, one must be familiar with the idea of hyperparameter tuning [9,10].

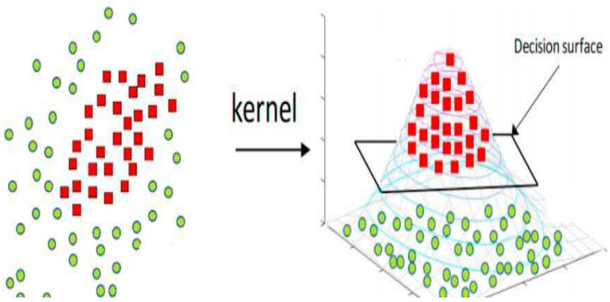

Fig. 1. Procedure of Kernel

Using of kernel method, which transforms 2D image into 3D one, non-linear decision boundary is created, as shown in Figure 1. After determining which hyperplane best fits the data, it returns the projection to 2D.

B. K Nearest Neighbors

After sorting the data into classes, this supervised learning model aligns new data points with their respective classes by first viewing their K-nearest neighbours, then the most often occurring neighbor, and finally the data point itself. After each data point has been correctly classified and added to the model, this process is repeated [11]. If we want an exact k-value, we'll have to run the model with a range of values. When k is odd, data categorization is a breeze. Once we have all the numbers we need, we can either see the data graphically or, more often, figure out the best value [12].

C. CNN

Algorithms in this subdomain of Deep Learning learn to behave and perceive the world as humans do when put into a real-time setting; they are trained to consider every way a person might approach a particular problem. Every part of our lives now uses one of these technologies, from simple mobile phones to very precise supercomputers. There has been an improvement in labor productivity, so much so that they are replacing people with sophisticated machinery [13].

The three layers that make up this CNN are as follows: • Input Layer • Hidden Layer • Output layer

Image data is stored as pixels in nodes within the input layer, and all operations in a CNN take place within these nodes. This layer is used to feed the model.

Instructions for processing data, extracting features, and transforming data are handled by the hidden layer of this layer. More buried tiers in a CNN's design equals a more complicated architecture. Its performance with real-time data is comparable to that of CNN models, and the model can learn features thanks to the data processing in the hidden layer. Node values inform the CNN's activation functions, which decide which neurons to activate. We identified and predicted the presence of illness on leaves using features recovered from the hidden layer in this study. The flattened outputs from the hidden layer are inputted into the fully linked output layer. This layer takes in data and parses it into the classes you specify [14].

V. *CLASSIFICATION AND IDENTIFICATION OF PLANT DISEASES*

Computer vision, a branch of artificial intelligence, enables computers to mimic human faces and extract, analyze, and recognize real-world images in the same way as humans [15].

Some success has been achieved using ML techniques for disease diagnosis and classification, but recent advances in DL (a subset of ML) indicate that this field has untapped potential to improve realism. Many deep learning and multi-technology tools have been developed to detect and classify symptoms of plant diseases [16].

Many fast-growing industries are already demonstrating the benefits of computer vision, including diagnostics, surveillance, satellite imagery, and agricultural businesses. Agricultural use of computer vision includes detection and classification of plant diseases based on various extractions and symptoms. It works on image processing, including scaling, filtering, segmentation, feature extraction and selection, starting from image acquisition, and then uses ML or DL search and classification [17].

VI. *CAUSES OF PLANT DISEASES*

Many agricultural diseases can occur at different stages of plant growth, affecting plant growth and ultimately reducing crop yield [15 , 18 , 19]. Plant diseases are caused by events that occur at different stages of plant growth [20]. Crop diseases are divided into two groups: biotic factors and abiotic factors, as described in reference [18]. While microbial organisms cause the formation of biotic organisms such as bacteria, fungi, viruses, mites and slugs in plants, abiotic organisms such as water, temperature, electricity and inadequate nutrition prevent the growth of the plant. The study included images of plant leaves suffering from various diseases from the PlantVillage dataset, as well as images of healthy and diseased plants from other sources [17]. Other images of healthy and diseased plants from other sources are also discussed in the works of [21] and [22]. This work in reference [17] provides a clear overview of computer vision methods and methods for diagnosis and classification of plant diseases, such as field crop, images acquisition, page image datasets, image pre-processing (test set, training set and validation set), data provides the definition. segmentation and performance evaluation. Figure 2 shows the factors that cause plant diseases. Figures 3 and 4 show samples of diseased leaves from the Plant Village dataset, as well as photos from other datasets displaying both healthy and sicked leaves.

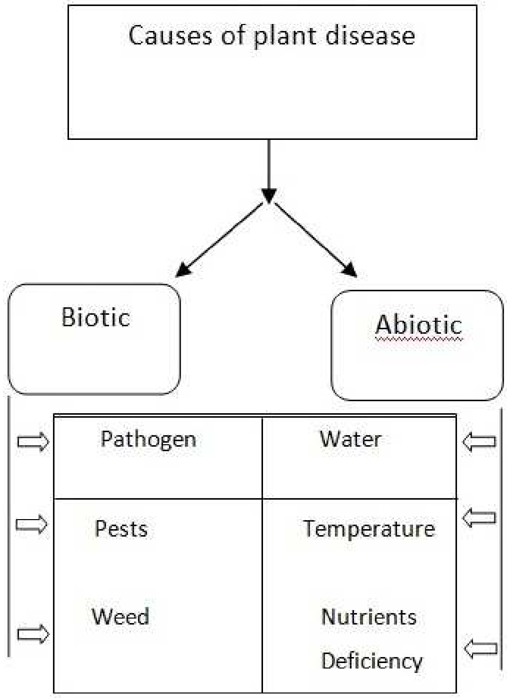

Fig. 2. Causes of Plant Disease

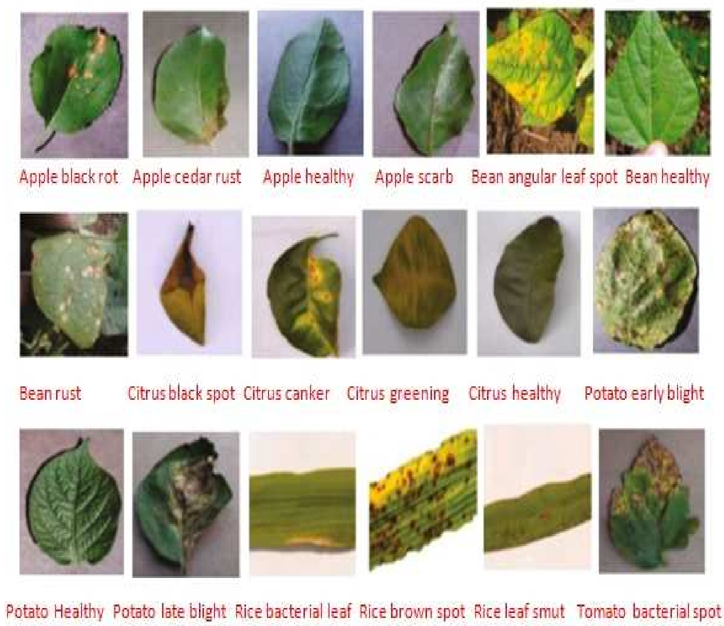

Fig. 3. Here are a few examples of plant disease photos from the Plant Village collection of leaf photographs. [17]

Fig. 4. Various pictures from various databases depicting damaged and healthy leaves of plants [21, 22]

VII. IMPLEMENTATION

To begin, a number of machine learning techniques are brought in for the soybean dataset taken from reference [13]. The SVM and KNN algorithms developed in python are employed to determine the kind of illness existing in the leaf.

A. Implementation through SVM

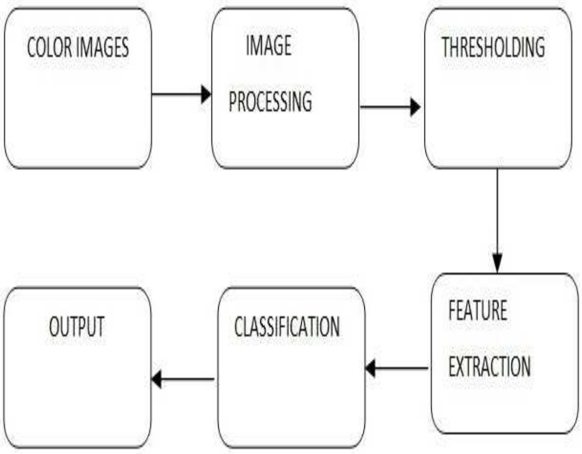

Fig.5. SVM Implementation demonstrates how SVM is used

B. Implementation through KNN

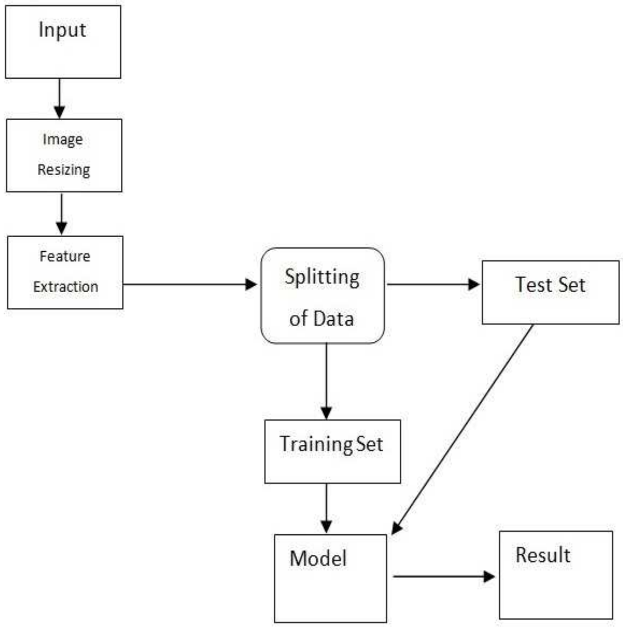

Fig-6 Implementation of KNN and demonstrated how KNN is used

VIII. RESULTS

A. Result of SVM

A 80% success rate was achieved with the "Soyabean leaf" dataset using the support vector machine (SVM) model. Predictions made using a small set of test photos show promising results.

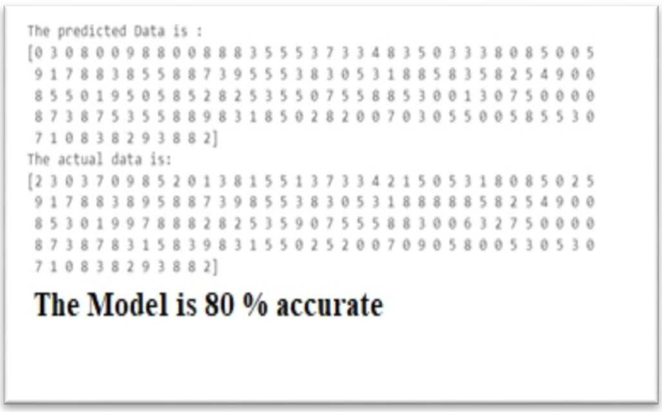

Fig. 7. SVM Model Result shows the result obtained from SVM model.

B. Result of KNN

Using the "Soyabean leaf" dataset to train the KNN model yields a 64% accuracy rate with a k-value of 6. K is calculated here using both the Manhattan distance and the Euclidean distance. In the plot, the outcomes are examined.

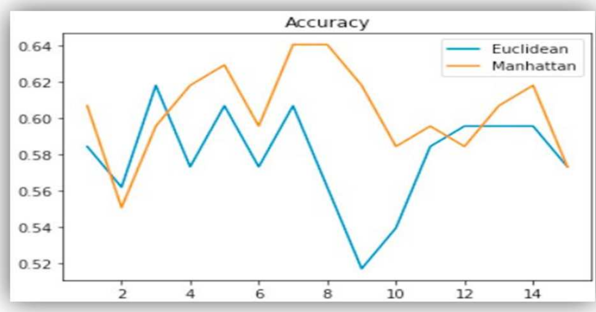

Fig 8. KNN Accuracy Result

IX. CONCLUSION

In this study, we apply SVM and KNN, two models for multi-class classification, to the problem of leaf disease identification. It aids farmers in the correct use of pesticides for diseases found on leaves by providing real-time image categorization and detection. We may conclude that SVM is superior to KNN based on the results obtained; KNN has an accuracy of 64% while SVM has an accuracy of 80%. The KNN models' accuracy might see an improvement with a larger dataset. Thus, support vector machines (SVMs) are superior models for leaf disease detection.

TABLE 1. COMPARISON TABLE

Model	**SVM**	**KNN**
Accuracy	80%	64%

Shows the accuracy of two different model

X. FUTURE SCOPE

It is possible to improve the accuracy of leaf disease diagnoses by developing numerous new machine learning methods. Computers with extra GPUs or a cluster of many computers can be used when dealing with massive datasets. Instant software that uses up-to-date images to accurately diagnose leaf diseases can be developed to assist and improve the lives of farmers, especially when assessing the quality of their products.

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] Jasim MA, Al-Tuwaijari JM. Plant leaf diseases detection and classification using image processing and deep learn- ing techniques. Int Comput Sci Soft Eng Conf. 2020. https://doi.org/10.1109/CSASE48920.2020.9142097

[3] Upadhyay SK, Kumar A. A novel approach for rice plant diseases classification with deep convolutional neural network. Int J Inf Technol. 2022;14(1):185–99. https://doi.org/10.1007/s41870-021-00817-5.

[4] Guerrero-Ibañez A, Reyes-Muñoz A. Monitoring tomato leaf disease through convolutional neural networks. Electron. 2023;12(1):1–15. https://doi.org/10.3390/electronics12010229

[5] Ahmed I, Yadav PK. A systematic analysis of machine learning and deep learning based approaches for identifying and diagnosing plant diseases. Sustain Oper Comput. 2023;4:96–104. https://doi.org/10.1016/j.susoc.2023.03.001

[6] Balaji V, et al. Deep transfer learning technique for multimodal disease classification in plant images. Contrast Media Mol Imaging. 2023;2023:5644727. https://doi.org/10.1155/2023/5644727

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

[8] N. R. Bhimte and V. R. Thool, "Diseases Detection of Cotton Leaf Spot Using Image Processing and SVM Classifier," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 340-344

[9] P. Krithika and S. Veni, "Leaf disease detection on cucumber leaves using multiclass Support Vector Machine," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2017, pp. 1276-1281.

[10] Y. K. Dubey, M. M. Mushrif and S. Tiple, "Superpixel based roughness measure for cotton leaf diseases detection and classification," 2018 4th International Conference on Recent Advances in Information Technology (RAIT), 2018, pp. 1-5.

[11] M. P. Vaishnnave, K. S. Devi, P. Srinivasan and G. A. P. Jothi, "Detection and Classification of Groundnut Leaf Diseases using KNN classifier," 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), 2019, pp. 1-5.

[12] S. Veni, R. Anand, D. Mohan and P. Sreevidya, "Leaf Recognition and Disease Detection using Content based Image Retrieval," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 243-247

[13] A. Jenifa, R. Ramalakshmi and V. Ramachandran, "Cotton Leaf Disease Classification using Deep Convolution Neural Network for Sustainable Cotton Production," 2019 IEEE International Conference on Clean Energy and Energy Efficient Electronics Circuit for Sustainable Development (INCCES), 2019, pp. 1-3.

[14] S. Kumar, K. Prasad, A. Srilekha, T. Suman, B. P. Rao and J. N. Vamshi Krishna, "Leaf Disease Detection and Classification based on Machine Learning," 2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE), 2020, pp. 361-365

[15] Kumar R, Chug A, Singh AP, Singh D. A systematic analysis of machine learning and deep learning based approaches for plant leaf disease classification: a Review. J Sensors. 2022. https://doi.org/10.1155/2022/3287561.

[16] Saleem MH, Potgieter J, Arif KM. Plant disease classification: a comparative evaluation of convolutional neural networks and deep learning optimizers. Plants. 2020;9(10):1–17. https://doi.org/10.3390/plants9101319.

[17] 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

[18] Vishnoi VK, Kumar K, Kumar B. Plant disease detection using computational intelligence and image processing. Berlin Heidelberg: Springer; 2021.

[19] Faizal Azizi MM, Lau HY. Advanced diagnostic approaches developed for the global menace of rice diseases: a review. Can J Plant Pathol. 2022;44(5):627–51. https://doi.org/10.1080/07060661.2022.2053588.

[20] Shoaib M, et al. An advanced deep learning models-based plant disease detection: a review of recent research. Front Plant Sci. 2023;14:1–22. https://doi.org/10.3389/fpls.2023.1158933.

[21] Ahmed I, Yadav PK. A systematic analysis of machine learning and deep learning based approaches for identifying and diagnosing plant diseases. Sustain Oper Comput. 2023;4:96–104. https://doi.org/10.1016/j.susoc.2023.03.001.

[22] Dhiman P, Kaur A, Balasaraswathi VR, Gulzar Y, Alwan AA, Hamid Y. Image acquisition, preprocessing and classifica- tion of citrus fruit diseases: a systematic literature review. Sustainability. 2023;15(12):9643. https://doi.org/10.3390/ su15129643.