Optimizing Performance in Mango Plant Leaf Disease Classification through Advanced Machine Learning Techniques

Sarika Khandelwal
Department of CSE, G H Raisoni College of Engineering, Nagpur, India
sarikakhandelwal@gmail.com (corresponding author)

Archana Raut
Department of CSE, G H Raisoni College of Engineering, Nagpur, India
archana.kakade5@gmail.com

Harsha Vyawahare
Department of CSE, Sipna College of Engineering and Technology, Amravati, India
harsha.vyawahare@gmail.com

Dipti Theng
Department of CSE, Symbiosis Institute of Technology, Pune, India
deepti.theng@gmail.com

Sheetal Dhande
Department of CSE, Sipna College of Engineering and Technology, Amravati, India
sheetaldhandedhandge@gmail.com

Received: 25 June 2024 | Revised: 21 August 2024 and 28 September 2024 | Accepted: 4 October 2024

 DOI: https://doi.org/10.48084/etasr.8220

ABSTRACT

Leaf diseases pose a significant threat to the productivity and quality of mango crops, necessitating effective detection and management strategies. This study presents an automated system for the detection of mango leaf diseases using machine learning techniques. Using image processing methods to extract relevant features from leaf images, various machine learning models were trained to accurately classify common mango leaf diseases. This approach involved using a comprehensive dataset of diseased and healthy mango leaves, preprocessing images, and extracting features such as color, texture, and shape. Features were extracted using MobileNetV2 and EfficientNetV2. Feature fusion was performed using a dense layer. Principal component analysis was used to reduce dimensionality. These reduced features were then fed to a support vector classifier to classify the mango leaf images. Eight different classes were considered, including the seven most common diseases in mango leaves and one class for healthy ones. The proposed model achieved a remarkable accuracy of 99.83 %. These results demonstrate that machine learning models can achieve high accuracy in the early detection of mango leaf diseases. Implementing this system in agricultural practices can significantly help farmers in timely disease management, reducing crop losses, and improving mango production.

Keywords-SVM; CNN; MobileNetV2; EfficientNetV2; MSE; precision; recall; f1-score

I. INTRODUCTION

Mango, also known as Magnifier Indicia (MI), is a prominent tropical fruit crop renowned for its palatability, nutritional benefits, and substantial economic value. Mango contributes significantly to the revenue of agricultural areas. India is the leading country in mango production, with approximately 40% of the world's production [1]. Pests and diseases are anticipated to destroy about 30% to 40% of overall crop yield [2]. However, the productivity and quality of mangoes are significantly compromised by various pathological conditions that affect different anatomical parts of the tree, including leaves, stems, and fruits. Among these, leaf diseases are particularly harmful as they inhibit photosynthetic efficiency, decrease tree vitality, and result in considerable yield reductions. To manage and control mango plant leaf diseases in the early stages, it is essential to accurately and promptly detect disease conditions. Current approaches to diagnosing and treating diseases are time-consuming and tedious. They also require professional examination and hence may not be cost-effective. Despite the best efforts of farmers to identify diseases based on visual inspection, many farmers are still unable to identify the diseases correctly, leading to lower-grade or lower mango production [4]. To improve mango yield, which will ultimately help the country's economy, it is necessary to identify leaf diseases promptly and accurately. The field of machine learning and computer vision can help develop an automated solution to predict mango diseases in the early stages.

This study utilized machine learning algorithms to detect mango leaf plant diseases. Different diseases of mango leaf plants can be identified and classified using machine learning algorithms by examining the attributes derived from leaf images, such as color, texture, etc. Deep Learning (DL) can automatically extract features from raw leaf image data and use them for disease classification and identification [5]. DL methods have various uses in the agriculture field, particularly for plant leaf disease classification. However, detection accuracy can be considerably improved by training DL models on plant leaf diseases that are specific to a given area [6, 7]. Convolutional Neural Networks (CNN) are the most popular deep learning approach for image data classification, as they are low in processing complexity and efficient in computations. CNN can produce good results with fewer neurons and may require less training time. Farmers can use these detection models to minimize crop losses and increase mango yield by adopting corrective measures. This study aimed to identify seven types of mango leaf disease. The mango leaf images were taken from the dataset in [3]. Table I shows the number of mango leaf samples for each class.

TABLE I. DATASET DESCRIPTION

Class	Common disease name	Scientific disease name	No. of samples
A	Anthracnose	Colletotrichum Gloeosporioides	500
B	Bacterial Canker	Xanthomonas Campestris pv. Mangiferaeindicae	500
C	Cutting Weevil	Sternochetus Mangiferae	500
D	Die Back	Lasiodiplodia Theobromae	500
E	Powdery Mildew	Oidium Mangiferae	500
F	Sooty Mould	Generally associated with Capnodium spp.	500
G	Gall Midge	Procontarinia Matteiana	500
H	Healthy		500

Various techniques have been proposed to improve accuracy and performance, mainly using image processing, pattern recognition [8], computer vision, neural networks [9], DL, etc. [10-12]. Some other extensively applied techniques include Polymerase Chain Reaction (PCR) based molecular approaches and spectroscopic imaging. These techniques are reliable and accurate, but they are time-consuming and require more expensive tools, time, and labor [13]. The various methods used to classify plant leaf diseases include SVM, NN, and K-means clustering, etc. For these techniques, the Grey Level Co-occurrence Matrix (GLCM) is frequently used for segmentation, color transformation, and feature extraction [9]. In [14], transfer learning techniques, based on VGG16 and ResNet50 pre-trained models, were used to identify potato plant leaf diseases. In [15], ANN and decision tree classifiers were used along with several machine learning models and image processing to focus on the potential of model fusion in disease classification. In [2], four different mango leaf diseases were identified using a customized pre-trained ResNet model. The performance of three different ResNet architectures was compared, and their accuracy was more than 90%. Among all the investigated models, the highest accuracy was 91.50% for ResNet50. In [16], transfer learning was used with the DenseNet201, InceptionResnetV2, and Resnet50 models to identify and classify mango leaf diseases, reporting an accuracy of 98% for DenseNet201.

In [17], a Neural Network Ensemble (NNE) model was combined with SVM for Mango Leaf Disease Recognition (MLDR). The proposed MLDR system achieved an average accuracy of 80%. In [18], image processing methods were used for the early recognition of plant diseases. A Raspberry Pi was used to connect a camera to a display unit and transmit data to the cloud. This approach involved capturing leave images and analyzing them through data preprocessing and segmentation followed by clustering. In [19], models such as ResNet, VGG Net, Inception V4, and DenseNets were examined, concluding that DenseNets achieved state-of-the-art performance due to their fewer parameters and realistic computation time. In [20], various CNN models were evaluated in mango and potato leaf disease identification, concluding that deeper networks such as VGG16 and ResNet50 provided higher accuracy than shallower models.

In [13], leaf diseases were detected on mango and grape plant images using the AlexNet transfer learning approach. The pre-training of this architecture was performed on a dataset named ImageNet. This model achieved an impressive accuracy of 99% for grape leaf classification, while for mango leaves, the accuracy stood at 89%. In [19], a deep learning approach was proposed using a CNN to identify five different mango leaf diseases. This model demonstrated remarkable performance, achieving an accuracy of 96.67% in correctly classifying various disease conditions. This approach first segments the leaf vein pattern using a leaf vein-seg method. Subsequently, the salient features were extracted from the input data and integrated through a fusion technique based on canonical correlation analysis followed by an application of SVM. The model in [21] achieved an accuracy of 95.5%. In [22], an ML-based clustering model was used on a self-captured dataset to detect fading leaf diseases. Table II shows a comparative analysis of various methods for mango plant leaf disease detection.

II. DATASET DESCRIPTION

The dataset used consists of 4000 images, including 1800 images of distinct leaves, while the remaining images were constructed by zooming and rotating the existing images. Each image is in the JPG format with 240×320 pixels. Each of the eight classes contains 500 samples. Figure 1 shows sample images for each class.

TABLE II. COMPARATIVE ANALYSIS OF MANGO PLANT LEAF DISEASE DETECTION METHODS

Ref. no	Disease	Model used	Dataset source	Dataset size	Accuracy %
[9]	Dag, Golmachi, moricha, Shutimold	SVM of NNE	Self - captured	Not mentioned	80
[17]	Powdery mildew, Red rust, Gall machi, Anthracnose	DenseNet 201	Self-captured	1500	98.00
[20]	Anthracnose	AlexNet	Plant Village	4004	98.33
[22]	Anthracnose	M-CNN	Plant Village	2200	97.13
[23]	Multiple	SVM	Self-Plant Village	510	76
[24]	Bacteria canker disease, Powdery mildew, Scab	Transfer learning using AlexNet	Plant Village	8438	89

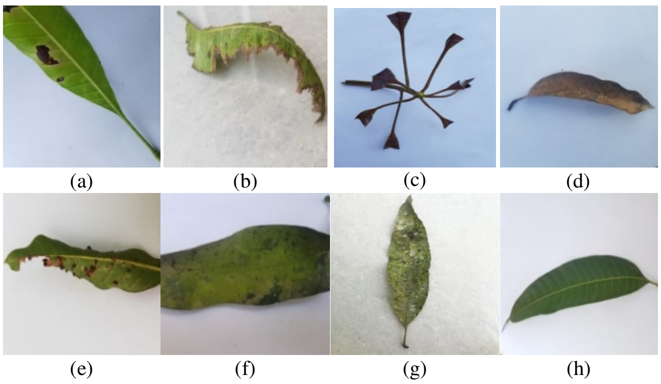

Fig. 1. Sample dataset images for each class: (a)Anthracnose, (b) Bacterial canker, (c) Cutting weevil, (d) Die back, (e) Powdery mildew, (f) Sooty mould, (g) Gall midge, (h) Healthy.

III. PROPOSED METHOD

This study used the MobileNetV2 and EfficinetNetV2 pre-trained models. MobileNetV2 is a compact CNN framework developed for vision applications on mobile and embedded devices [26]. MobileNetV2 features depth-wise separable convolution, inverted residuals, bottleneck design, linear bottlenecks, and Squeeze-and-Excitation (SE) blocks. These features enhance the model's performance and efficiency in classifying images, in addition to reducing computational demands without compromising accuracy. At the same time, MobileNetV2 is a small model that allows quicker inference times, which makes it appropriate for real-time applications as well. In its architecture, depth-wise separable convolutions help decrease computational cost. It also divides the typical convolution into two distinct processes: depth-wise convolution and point-wise convolution. Some other features of MobileNetV2 are:

- Inverted residuals help in improving the model's accuracy. Its bottleneck structure increases channels before depth-wise separable convolutions, enhancing the model's ability to capture complex features.
- Its bottleneck design reduces the computational cost by using 1×1 convolutions to decrease the number of channels before depth-wise separable convolutions.
- Linear Bottlenecks: Uses linear activations to prevent information loss and capture fine-grained details. SE blocks improve feature representation by adaptively recalibrating channel-wise feature responses.

EfficientNetV2 is a new series of convolutional networks with faster training and improved efficiency compared to earlier models [25].

The proposed model consists of 4 phases, as described below:

- Preprocessing: The dataset images were preprocessed before applying the pre-trained models. For EfficinetNetV2, the images were resized to 300×300, whereas for the MobileNetV2 were resized to 224×224.
- Feature extraction: The Efficient V2 and MobileNetV2 pre-trained models were used for feature extraction. These features were concatenated, and then a dense layer was used in the feature fusion step. For both models, a batch size of 64 was applied. The extracted features of both models were concatenated and one dense layer was used with the Rectified Linear Unit (ReLU) activation function and 1024 neurons. Adam optimization was used to optimize the dense layer.
- Dimensionality reduction: Principal Component Analysis (PCA) was used to reduce the dimensions of the features resulting from the previous phase. PCA was applied to reduce the feature set to 700, which is 97% of the variance ratio.
- Support Vector Classifier (SVC): SVC was used to classify the reduced feature set into 8 different classes, namely A, B, C, D, E, F, G, and H. Figure 2 shows the proposed classifier. SVC with radial basis function was used for classification. Its performance was tested on a dataset of 4000 images. The entire dataset was divided into training and testing sets with a ratio of 70% (2800 images) to 30% (1200 images).

IV. FEATURE EXTRACTION AND FEATURE FUSION

The most popular kernel function among the various kernelized algorithms is the Radial Basis Function (RBF). RBF was used, which is normally used in SVM classification. For two samples x and x', the RBF kernel is represented as a feature vector in some input space given in:

$$K(x,x') = \exp\left(-\frac{||x-x'||^2}{2\,\sigma^2}\right) \quad (1)$$

where x and x' are two data points. $||x-x'||^2$ is known as the squared Euclidean distance among the two feature vectors. σ is a free parameter. The same definition includes a parameter $\gamma = \frac{1}{2\,\sigma^2}$ as given in:

$$K(x,x') = \exp(-\gamma||x-x'||^2) \quad (2)$$

In (2), γ defines the influence of a single training sample. The lower the value of γ, the larger the radius of influence, while a high value means a small radius. The summary of this feature extraction and fusion model is given in Table III, and the layered architecture of the proposed model is shown in Figure 3.

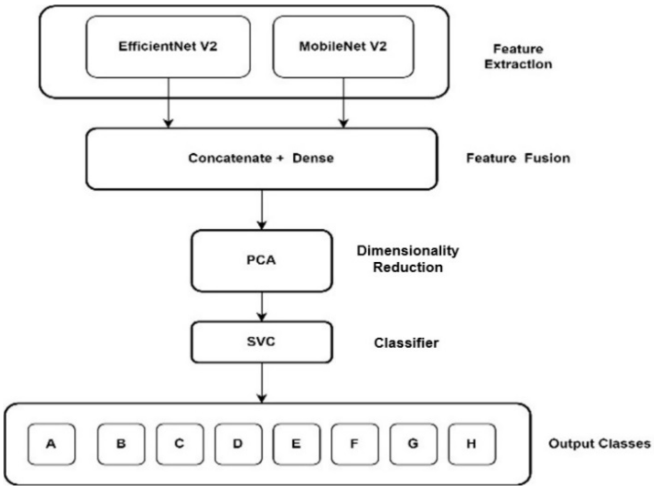

Fig. 2. The proposed classifier.

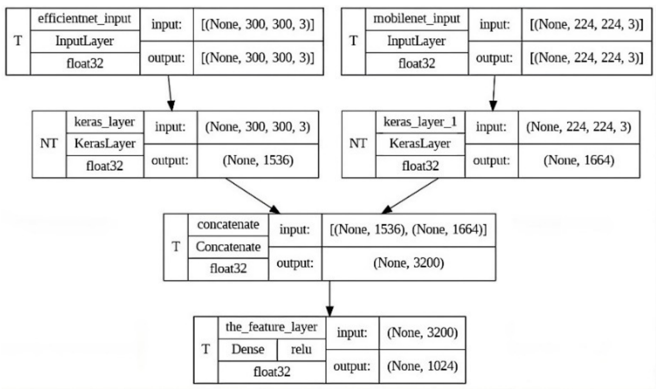

Fig. 3. Layered architecture of the proposed model.

TABLE III. SUMMARY OF THE FEATURE EXTRACTION AND FUSION MODEL

Layer (type)	Output shape	Param #	Connected to
efficientnet_input (InputLayer)	[(None,300,300,3)]	0	[ ]
Mobilenet_input (InputLayer)	[(None,224,224,3)]	0	[ ]
Keras_layer (KerasLayer)	(None,1536)	12930622	[efficientnet_input[0][0]]
Keras_layer_1 (KerasLayer)	(None,1664)	3766048	[mobilenet_input[0][0]]
Concatenate (Concatenate)	(None,3200)	0	[keras_layer[0][0], keras_layer_1[0][0]]
The_feature_layer (Dense)	(None,1024)	3277824	[concatenate[0][0]]
Total params:19974494 (76.20 MB) Trainable params: 3277824 (12.50 MB) Non-trainable params: 16696670 (63.69 MB)			

V. RESULTS AND DISCUSSION

This study used a dataset comprising 4,000 images of 1,800 distinct leaves, including 8 classes: 7 most common diseases and one healthy class. Data augmentation was applied to the collected dataset to increase the number of images for model training and testing. The data set was divided into 70% for training and 30% for testing the model. A model based on a feature fusion extractor and an SVC classifier was developed to classify the images, as previously described. The experiment was carried out on Google Colab, using a free tier subscription with NVIDIA K80 12 GB GDDR5 VRAM, up to 16 GB of RAM, and approximately 33 GB of disk space. Figure 4 shows the generated confusion matrix. The performance of the model was evaluated using precision, recall, and F1 score, as shown in Table IV. The accuracy of the model was calculated using:

$$Accuracy = (TP + TN) / (Total\ Cases)$$

The accuracy of the proposed system was 0.998. Figure 7 shows a comparative analysis of the accuracy of the proposed model over existing state-of-the-art models available in the literature. This comparative analysis shows that the proposed model is superior to the existing models considering accuracy.

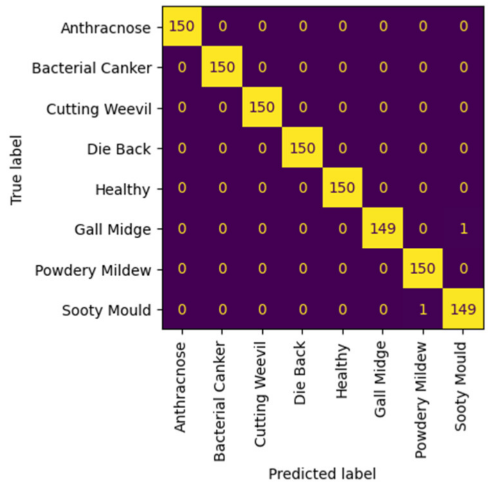

Fig. 4. Confusion matrix.

TABLE IV. PERFORMANCE OF THE PROPOSED MODEL

	Disease classes							
	A	B	C	D	E	F	G	H
Precision	1.0	1.0	1.0	1.0	0.99	0.99	1.0	1.0
Recall	1.0	1.0	1.0	1.0	1.0	0.99	0.99	1.0
F1 score	1.0	1.0	1.0	1.0	1.0	0.99	1.0	1.0

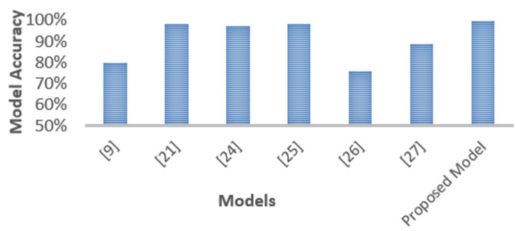

Fig. 5. Accuracy comparison.

VI. CONCLUSION AND FUTURE SCOPE

In India, mangoes are among the most widely grown fruit crops. Therefore, it is essential to protect its cultivation and identify the numerous diseases in their early stages. This study proposed a model based on a feature fusion extractor and an SVC classifier, using a deep learning and image processing method to recognize eight plant leaf classes, including seven disease classes and one healthy. To strengthen the proposed model's accuracy, PCA was used for dimensionality reduction, and features were extracted using pre-trained MobiNetV2 and EfficientNetV2 models. Feature fusion with the results of the two pre-trained models made the proposed model more accurate. The use of an RBF kernel optimization and PCA for dimensionality reduction enhanced model performance. This method identified seven different diseases using a minimum set of layers. The Mango Leaf Plant Diseases dataset [3] was used to test the performance of the model, where it achieved a notable accuracy of 99.83%. The accuracy of the proposed model was far better than previous models in the same domain. Increasing the quantity of datasets and utilizing transfer learning are two ways to improve model performance and efficiency. In the future, some other disease classes can be added to expand the proposed model.

REFERENCES

[1] S. Arivazhagan and S. V. Ligi, "Mango Leaf Diseases Identification Using Convolutional Neural Network," *International Journal of Pure and Applied Mathematics*, vol. 120, no. 8, pp. 11067–11079, 2018.

[2] S. Gulavnai and R. Patil, "Deep learning for image based mango leaf disease detection," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3S3, pp. 54–56, 2019.

[3] Sawkat Ali, "MangoLeafBD Dataset." Mendeley, Aug. 30, 2022, https://doi.org/10.17632/HXSNVWTY3R.1.

[4] L. F. Arauz, "Mango Anthracnose: Economic Impact and Current Options For Integrated Management," *Plant Disease*, vol. 84, no. 6, pp. 600–611, Jun. 2000, https://doi.org/10.1094/PDIS.2000.84.6.600.

[5] A. N. Tusher, Md. T. Islam, Mst. S. R. Sammy, S. A. Hasna, and N. R. Chakraborty, "Automatic Recognition of Plant Leaf Diseases Using Deep Learning (Multilayer CNN) and Image Processing," in *Third International Conference on Image Processing and Capsule Networks*, 2022, pp. 130–142, https://doi.org/10.1007/978-3-031-12413-6_11.

[6] J. Wu, "Introduction to Convolutional Neural Networks," May 2017.

[7] R. C. Ploetz, "The major diseases of mango: strategies and potential for sustainable management," *Acta Horticulturae*, no. 645, pp. 137–150, Feb. 2004, https://doi.org/10.17660/ActaHortic.2004.645.10.

[8] J. Francis, Anto Sahaya Dhas D, and Anoop B K, "Identification of leaf diseases in pepper plants using soft computing techniques," in *2016 Conference on Emerging Devices and Smart Systems (ICEDSS)*, Namakkal, India, Mar. 2016, pp. 168–173, https://doi.org/10.1109/ICEDSS.2016.7587787.

[9] Md. R. Mia, S. Roy, S. K. Das, and Md. A. Rahman, "Mango leaf disease recognition using neural network and support vector machine," *Iran Journal of Computer Science*, vol. 3, no. 3, pp. 185–193, Sep. 2020, https://doi.org/10.1007/s42044-020-00057-z.

[10] M. Bhange and H. A. Hingoliwala, "Smart Farming: Pomegranate Disease Detection Using Image Processing," *Procedia Computer Science*, vol. 58, pp. 280–288, Jan. 2015, https://doi.org/10.1016/j.procs.2015.08.022.

[11] U. Obu, Y. Ambekar, H. Dhote, S. Wadbudhe, S. Khandelwal, and S. Dongre, "Crop Disease Detection using Yolo V5 on Raspberry Pi," in *2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN)*, Salem, India, Jun. 2023, pp. 528–533, https://doi.org/10.1109/ICPCSN58827.2023.00092.

[12] P. M. Mainkar, S. Ghorpade, and M. Adawadkar, "Plant leaf disease detection and classification using image processing techniques," *International Journal of Innovative and Emerging Research in Engineering*, vol. 2, no. 4, pp. 139–144, 2015.

[13] S. Sankaran, A. Mishra, R. Ehsani, and C. Davis, "A review of advanced techniques for detecting plant diseases," *Computers and Electronics in Agriculture*, vol. 72, no. 1, pp. 1–13, Jun. 2010, https://doi.org/10.1016/j.compag.2010.02.007.

[14] F. Islam, M. N. Hoq, and C. M. Rahman, "Application of Transfer Learning to Detect Potato Disease from Leaf Image," in *2019 IEEE International Conference on Robotics, Automation, Artificial-intelligence and Internet-of-Things (RAAICON)*, Dhaka, Bangladesh, Nov. 2019, pp. 127–130, https://doi.org/10.1109/RAAICON48939.2019.53.

[15] S. Sundari, "Leaf Disease Classification Using Artificial Neural Networks and Decision Tree Classifier," *Journal of Image Processing and Pattern Recognition Programming*, vol. 5, no. 1, Apr. 2018.

[16] A. Rajbongshi, T. Khan, Md. M. Rahman, A. Pramanik, S. M. T. Siddiquee, and N. R. Chakraborty, "Recognition of mango leaf disease using convolutional neural network models: a transfer learning approach," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 3, pp. 1681–1688, Sep. 2021, https://doi.org/10.11591/ijeecs.v23.i3.pp1681-1688.

[17] M. Merchant, V. Paradkar, M. Khanna, and S. Gokhale, "Mango Leaf Deficiency Detection Using Digital Image Processing and Machine Learning," in *2018 3rd International Conference for Convergence in Technology (I2CT)*, Pune, India, Apr. 2018, pp. 1–3, https://doi.org/10.1109/I2CT.2018.8529755.

[18] S. Aftab, C. Lal, S. K. Beejal, and A. Fatima, "Raspberry Pi (Python AI) for Plant Disease Detection," *International Journal of Current Research and Review*, vol. 14, no. 03, pp. 36–42, 2022, https://doi.org/10.31782/IJCRR.2022.14307.

[19] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Computers and Electronics in Agriculture*, vol. 161, pp. 272–279, Jun. 2019, https://doi.org/10.1016/j.compag.2018.03.032.

[20] S. Arya and R. Singh, "A Comparative Study of CNN and AlexNet for Detection of Disease in Potato and Mango leaf," in *2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, Ghaziabad, India, Sep. 2019, pp. 1–6, https://doi.org/10.1109/ICICT46931.2019.8977648.

[21] R. Saleem, J. H. Shah, M. Sharif, M. Yasmin, H.-S. Yong, and J. Cha, "Mango Leaf Disease Recognition and Classification Using Novel Segmentation and Vein Pattern Technique," *Applied Sciences*, vol. 11, no. 24, Jan. 2021, Art. no. 11901, https://doi.org/10.3390/app112411901.

[22] U. P. Singh, S. S. Chouhan, S. Jain, and S. Jain, "Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease," *IEEE Access*, vol. 7, pp. 43721–43729, 2019, https://doi.org/10.1109/ACCESS.2019.2907383.

[23] K. Kusrini *et al.*, "Data augmentation for automated pest classification in Mango farms," *Computers and Electronics in Agriculture*, vol. 179, Dec. 2020, Art. no. 105842, https://doi.org/10.1016/j.compag.2020.105842.

[24] U. Sanath Rao *et al.*, "Deep Learning Precision Farming: Grapes and Mango Leaf Disease Detection by Transfer Learning," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 535–544, Nov. 2021, https://doi.org/10.1016/j.gltp.2021.08.002.

[25] M. Tan and Q. Le, "EfficientNetV2: Smaller Models and Faster Training," in *Proceedings of the 38th International Conference on Machine Learning*, Jul. 2021, pp. 10096–10106.

[26] S. Anwar, S. R. Soomro, S. K. Baloch, A. A. Patoli, and A. R. Kolachi, "Performance Analysis of Deep Transfer Learning Models for the Automated Detection of Cotton Plant Diseases," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11561–11567, Oct. 2023, https://doi.org/10.48084/etasr.6187.