IJRCS
International Journal of Robotics and Control Systems
Vol. 5, No. 4, 2025, pp. 2161-2177
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs
ASCEE

Deep Neural Network and KNN (CNN-KNN) Based Approach to Classify Mango Leaf Diseases

Ravi Anand [a,b,1,*], Ritesh Kumar Mishra [a,2], Rijwan Khan [c,3]

[a] National Institute of Technology, Patna, India
[b] Noida Institute of Engineering & Technology, Gr. Noida, India
[c] ABES Institute of Technology, Ghaziabad, India
[1] ravia.phd20.ec@nitp.ac.in; [2] rkmishra73@gmail.com; [3] rijwankhan786@gmail.com
* Corresponding Author

ARTICLE INFO

Article history
Received May 19, 2025
Revised July 18, 2025
Accepted November 04, 2025

Keywords
K-Nearest Neighbors Algorithm;
CNN;
Plant Disease Detection;
Agricultural AI;
Image Classifications

ABSTRACT

The ecological relevance of plants and the goods they produce is also diminished by fungal infections, impacting their economic value. The mango tree is severely afflicted by a fungus called anthracnose, most noticeable on the fruits and leaves. This paper's primary goal is to stipulate a viable system for a premature and reasonable solution for mango leaf disease detection by designing an appropriate and effective method. In recent years, digital image processing and deep neural network-based approaches have gained popularity in categorizing various mango leaf infections due to their high computational performance and identification accuracy. This paper proposed an algorithm based on deep neural network-based feature extraction and K-Nearest Neighbors Algorithm-based classification task to classify mango leaf diseases. This paper describes the possibility for CNN to extract mango leaf features on leaf images taken from the MangoleafBD dataset. The used dataset contains three types of leaf images, including healthy leaf images. The proposed method gives an accuracy level of 99.37 % at K-fold value 20. The obtained result shows that the developed model can be recommended for precise farming practices as a secondary opinion tool for mango leaf disease detection.

1. Introduction

Image processing is a broad term that encompasses a variety of techniques. Many steps, such as image capturing, pre-processing, segmentation, analysis, interpretation, recognition, and classification, are helpful in developing an automated system. Recent developments in technology have made the interplay between image processing and machine learning responsible for the automation of several computational processes. This combination has made way for the automatic extraction and characterization of information for a variety of applications such as precision agriculture, environmental surveillance, human computer interaction, brain machine interfaces, process control, and others in diverse areas where appropriate decision-making is desired.

The prominent research areas in precision agriculture are Soil and Crop Monitoring, Automated Irrigation Systems, Yield Prediction and Mapping, Weed, Pest, and Disease Detection, Agricultural Robotics and Automation, Decision Support Systems, Farm Management Information Systems

(FMIS), Big Data Analytics and AI Integration, Sustainable and Resource-Efficient Farming, and so on. Among the mentioned domains, Weed, Pest, and Disease Detection are considered for the present work. In the traditional farming approach, disease detection was identified manually, which is time-consuming and completely dependent on the expertise of the cultivator. To overcome the limitation or dependency of the experts, precise farming applied a computer vision and machine learning approach to detect the disease timely manner and accurately [1].

However, the number of diseases occurring in different parts of the plant showed visible symptoms, an important factor in diagnosing these diseases. The diagnosis of mango plant diseases is an excellent and skillful skill. Mango plant diseases were once defined artificially. In this circumstance, people are seeking new methods to preserve rice production, safeguard the environment, and decrease the use of hazardous pesticides and spraying. As a result, agricultural specialists, scientists, and researchers are primarily concerned with identifying, diagnosing, classifying, and guiding regulatory situations. As a result, the only approach to ensure better results is to diagnose the disease early and prevent it. As a result, it is unavoidable that cutting-edge technologies such as information science, artificial intelligence, digital imaging, and processing will aid the national economy's development to attain higher yields. The photos are of a fictitious pathogen provider in agriculture and associated sectors. Until recently, the sole option was to duplicate, reproduce, analyse, and show these recorded photos. Various disciplines such as digital image processing, space and mass image filtering technology, picture enhancement, and restoration technology are vital since it is simple and easy to comprehend [2]. Fungal diseases are very common in plant leaves. The diseases in plants are a cause for the decline in the quality and quantity of agricultural production [3].

As a result, we suggest a modified Convolutional Neural Network (CNN) in this work to categorize mango leaves affected by the fungus Anthracnose. Images taken in the real world validate the model's performance. The histogram of equalization, which balances the invariability across photographs taken under actual conditions, is used in the preprocessing of the photos. The central square crop technique resizes these photographs to a standard size image. The proposed model based on CNN-KNN is then trained and tested for the detection of sick mango leaves, considering the following key factors:

- Effective and timely treatment can be started in advance thanks to automatic and early detection of a disease and its severity.
- Understanding the disease's nature and life cycle can also help people understand their vulnerabilities.
- As a result, a deep learning model called Modified CNN is proposed for categorizing anthracnose-infected leaves.
- Images of healthy and infected leaves from real mango trees with fungal disease are gathered for this project. In addition, the model's performance is assessed against other cutting-edge methods using a compiled and standardized database.
- The suggested approach is automatic, computationally effective, and affordable, which can assist in maintaining the significance of the mango tree and its yields from an ecological and economic standpoint.

The key contributions of the study are as follows:

- The proposed model uses the concept of hybrid model preparation. According to this, the CNN-KNN model has been proposed.
- The proposed model uses 11106 samples, comprising 7773 training samples and 3333 testing samples.
- The proposed model achieved 99.37 % classification accuracy. The model has the capability to identify healthy, Anthracnose, and Black Sooty Mold classes with an accuracy of 99.51%, 99.34%, and 99.18%, respectively.

2. Literature Review

Mango disease control is an essential aspect of horticulture ecological management, as it is connected to the health and productivity of fruit plants. It is particularly significant in India, a rapidly developing country. Low productivity and vulnerability to several severe illnesses present substantial disease prevention and control problems. As a result, the current study has practical implications.

Numerous studies have suggested identifying and classifying leafy plants. In this review, the authors discuss all aspects of disease control, including the concepts, techniques, challenges, advantages, and disadvantages of image processing [1]. Several methods of neural networks to identify and classify diseases in plant leaf images. This work includes the models, types, mechanics, and classes used, and many other visual representations related to hyperspectral images. In the work presented by Ma et al., four types of poultry were classified from the leaves, namely, anthracnose, fast pink, powdery mildew, and leaf soup [2], [3].

All images are collected in real-time and categorized through deep convolutional neural networks (DCNN) [4]. VGG convolutional neural network to identify the index for plant leaves. The presentation method categorizes an image into patients and non-patients. The results are validated in a large data set, showing the validity of the in-depth study method [5]. Four convolutional network systems, such as VGG 16, Inception, V4, ResNet, and DenseNets, are used to classify the disease in images. The images were taken from the rice field documentation, which consists of 38 categories of health. Compared with other media outlets, the DenseNets network can achieve high visibility with high classification and is not time-consuming [6]. The advancement of an expert system for the identification of plant diseases in the Barracuda Mango (Nam-Dok Mai), one of Thailand's biggest agricultural exports, was described in this paper. However, because Thailand is a tropical nation, climate change is producing changes in plant diseases and impacting mango tree development. Many forms of agricultural production exist as a result of agriculturists' lack of understanding on how to correctly identify plant species. Furthermore, no decision-making mechanism is in place in regard to deciding how to prevent or cure disease on the farm. It resulted in several errors in the maintenance of the plants impacted. As a result, the method was created to assist agriculturists in identifying the damaged plants and resolving the problem as soon as possible. Agricultural professionals should use their advanced degrees to the disease process in question. The plant diagnostics application uses model-based data mining techniques to create a knowledge system. A model set for a rule-based leaf picture is presented in this article.

The model-based model contains 129 leaf pictures gathered from the mango area under quality control and standardization standards set by Maeda University, with findings of category 3 tablets (anthracnose, algal spot, Normal) with an accuracy of 89.92 percent. The findings of the experiments demonstrate that a principle-based model may be used to apply plant marches [7]. Fungal infections not only reduce the weight of plants and their products, but they also have a negative impact on the environment. Anthracnose is a disease that affects fruiting trees, particularly their fruits and foliage. The goal of this article is to come up with an effective and efficient approach to treat the condition and its symptoms, or to seek assistance for quick and effective remedies to this problem. Computer vision and in-depth learning methods have earned a reputation in the categorization of various diseases in recent years due to their better performance in calculus and accuracy [8]. Deep Learning (DL) is one of the most rapid learning machines. In-depth studies utilise CNN to categorise pictures, since they give the most correct answers to issues throughout the world. CNN contains architectural elements such as AlexNet, Google Net, Dense Net, Squeeze Net, ResNet, VGGNet etc. pretrained. In this work, CNN and AlexNet laboratories were used to identify and compare the precision and efficacy of blue and potato leaves. For this project, a 4004-page diagram was used. The Village Village website included a picture of the potato, and a picture of the blue was taken on the GBPUAT page. The results demonstrate that the AlexNet application is more accurate than the CNN architecture [9]. By identifying pests and diseases of leaves and fruits, this research has improved the management of specific crops used in insect farming, thus addressing one of the major causes of pests. declining mango production in the Philippines and helping blue farmers in Pampanga. Farmers' distrust of using pesticides in times of pests. Multi-SVM and GLCM are used in image processing to detect anthracnoses, fruit flies and raspberry molds at 85% accuracy. It is done through contrasting kurtosis, rubbing, and entropy.

This research project can be used as a model for other fruit trees, or as a basis for local agricultural management (especially blue farming using data science) [10]. Anthracnose or foliage (red leaves), many diseases affecting blue plants. Mangoes are economically important, and it is necessary to test these diseases to prevent epidemics and reduce production. It is said to use machine vision to use blue-leaf imaging to identify plant diseases. The method is available using the YCbCr converted image and creates vectors of the font and font of the displayed image, which are provided to the participants in the test process. GLCM, a color-based technology and launch Gabor is used for writing shapes and colors. Comparing the results obtained by the most advanced steering class and the vector support machine (SVM) was completed. Analyze the application techniques used to obtain individual results for each technique. The overall results are in an image file with 86 images. The accuracy of categorizing the participants at the farthest distance and the support vector machines are 79.16% and 83.34%, respectively [11]-[13].

Fast and accurate identification of harmful microorganisms is critical for public safety biomonitoring to avert foodborne disease and ensure food security. Microorganism detection techniques have improved over time [14], [15]. Researchers used the ABC and CS algorithms to optimise the distribution route for fresh food delivery in the time window while also considering other considerations such as a fixed number of delivery trucks, fixed cost, and fixed fuel by covering all service locations [16]. In any situation where human intervention is problematic, the robot can protect warehouses from invaders. This robot would be inexpensive, efficient, and safe, as it would keep the food intact and ensure its high quality [17].

In 2008, a security system was presented that used sensors and actuators to regulate monitoring. It contained two communication systems, one for communicating with modules and the other for regulating processes [18]. Data mining techniques with ANN and Hybrid approach with Fuzzy logic is used for improving food quality [19], [20]. In 2013, a gadget used fuzzy logic conclusions and a computer vision system to locate eggs with spots or cracks and grade them depending on the defect or size. It was possible thanks to a specially devised algorithm in which the eggs were given a 100 % rating if they were normal and an 83 % rating if they had blood spots [21]. Various novel optimization algorithms have been developed in recent years, including ant colony optimization (ACO) [22], [23], particle swarm optimization (PSO) [24], cuckoo search (CS) [25], [26], lion optimization algorithm (LOA) [27], grasshopper optimization algorithm [28], Levy flight algorithm [29], tree growth algorithm [30], grey wolf optimizer [31], bees pollen optimization algorithm [32], Harris Hawks optimization (HHO) [33], Researchers have been interested in using the butterfly optimization (BO) method [34]-[40] and hybrid algorithms to solve various unconstrained and constrained optimization issues. Various approaches discussed in the literature survey are summarized in Table 1.

Table 1. Summarized literature survey

Reference No	Used Model	Measurement of Performance in %
[41]	Support Vector Machine [SVM]	87.50
[42]	CCA fusion + SVM	95.50
[43]	Fully connected Convolutional Network	98.90
[44]	K-means based clustering algorithm	93.12
[45]	Hybrid based CNN	98.70
[46]	CNN	97.04
[47]	CNN	91.83
[48]	SVM + XGBoost	86.58
[49]	CNN	79.64
[50]	CNN + RNN	94.27
[51]	CNN + ReLu	95.30

From the literature review, it has been observed that many models built on CNN architecture surpass classical machine learning approaches such as SVM and clustering algorithms. The optimal performance is achieved by hybrid models or with deeper architectures of CNNs. It has also been noticed that models that integrate different approaches, like CCA with SVM or CNN with RNN, perform better than classical methods when applied in isolation. The performance of several CNN models is lacking, as mentioned in the study [49], where the accuracy was only 79.64%. This indicates an issue with the optimization process applied or uniformity within the applied CNN architectures. On the basis of the identified gaps and trends, it has been decided that the hybrid approach is more suitable for the mango leaf disease detection. Therefore, a CNN-KNN-based approach is applied in which CNN is used for the feature extraction and KNN is used for classification.

3. Materials and Methodology

There should be enough detail in the materials and methods section for all procedures to be repeated. Two subsections divide this section. Initially, there is the dataset that is utilized along with the suggested approach.

3.1. Data Set Description

The dataset used in this paper is taken from the public source MangoleafBD [52]. Fig. 1 depicts different types of images used in this research. This dataset contains Mango Leaves image samples, which are divided into three categories named Healthy, Anthracnose, and Black Soothy Mold. The MangoleafBD dataset consists of 4000 images of seven different classes. For this study, three class samples are considered, and further, these sample images are augmented, and similar types of images are generated through different augmentation techniques. In the augmentation process, horizontal flip, vertical flip, and random rotation of 15 degrees are applied. The generated set of image details for each class is given in Table 2.

Table 2. Dataset description

Class Name	Sample Count	Training Sample	Testing Sample
Healthy	2024	1416	608
Anthracnose	5025	3517	1508
Black Soothy Mold	4057	2840	1217
Total sample	**11106**	**7773**	**3333**

A total of 11106 images is generated of healthy, anthracnose and black Soothy mold class. Among this set of samples, 2024 samples belong to the healthy class, 5025 samples belong to the anthracnose class, and 4057 samples belong to the black Soothy mold classes.

3.2. Proposed Methodology

The primary aim of the procedure is to establish an acceptable and practical approach to identifying the disease and its symptoms so that an appropriate system can be developed to handle the problem as soon as feasible and effectively as possible. Computer vision and in-depth research approaches have become more prevalent in the categorization of fungal infections in recent years because of the efficiency and accuracy of their calculations. The proposed methodology comprised of three main section (i) pre-processing section, (ii) Feature extraction using CNN and (iii) classification using KNN classifier. The brief description of each section is given consecutively. Fig. 2 shows the basic flow chart of the proposed methodology.

The figure describes the classification of image based leaf diseases using convolutional neural network (CNN) that is integrated with KNN classifier for the ultimate prediction. The initial stage of the system is to capture images of mango leaves which have been afflicted by various diseases (depicted in the top-left portion). These images depict both the healthy and the diseased states like: Soothy Mould, Anthracnose, and healthy class images.

The preprocessing stage focuses on further refining the deep learning workflow by performing the following actions: Image Scaling: Changing the mango leaf images to a specific measurement (example 224×224) so that the CNN can work with it. Denoising: Improving the quality of an image by removing convincing noise, blur, or artifacts. Normalization: Rendering of pixel values (normally

0-255) of 0-1 enhances training efficiency. The next step is Feature Extraction (Convolutional Neural Network - CNN). A deep CNN makes use of layers to preprocess images in order to extract meaningful features. The CNN model consisted of Convolution + ReLU layers used to extract edges, texture, and shapes using filters alongside detecting patterns. Adding non-linearity is achieved through ReLU. Then pooling layer cuts down on computation burdens while allowing for the preservation of important features. Multiple Conv + Pool Layers: Stacked layers allow better shallow-to-deep (low- to high-level) feature learning. The entire working of the CNN model is briefly described through mathematical expression in given equation (1) to equation (5). Here, X stands for input image, W_1 stands for weight at layer 1, and b_1 stands for bias at layer 1.

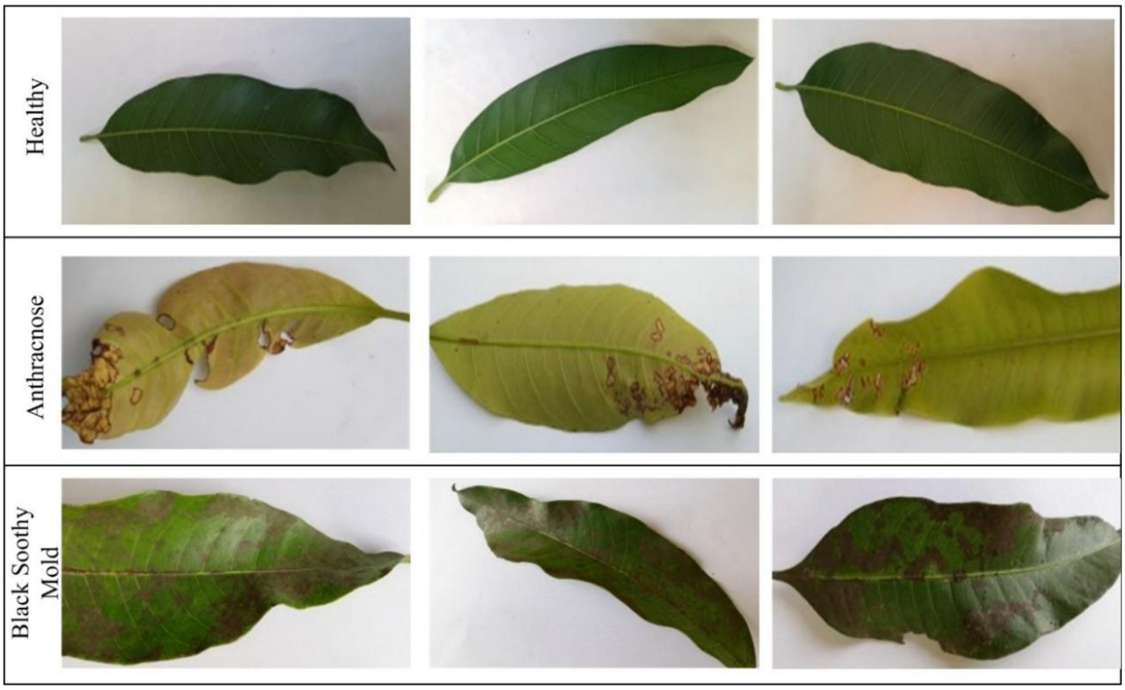

Fig. 1. Different types of sample images taken from the dataset

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} W_{(1)} = \begin{bmatrix} w_{11}\, w_{12} \\ w_{21}\, w_{22} \\ w_{31}\, w_{32} \\ w_{41}\, w_{34} \end{bmatrix} b_1 = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$Z^{(1)} = \begin{bmatrix} w_{11} & w_{21} & w_{31} & w_{41} \\ w_{12} & w_{22} & w_{32} & w_{42} \end{bmatrix} \bullet \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \tag{1}$$

$$Z^{(1)} = W^{(1)T} \bullet X^{(1)} + b^{(1)} \tag{2}$$

$$A^{(1)} = ActivationFunction(Z^{(1)}) \tag{3}$$

$$Z^{(2)} = W^{(2)T} \bullet A^{(1)} + b^{(2)} \tag{4}$$

$$A^{(2)} = activationFunction(Z^{(2)}) \tag{5}$$

3.3. Pre-Processing Section

Image pre-processing is very important and crucial step for the proposed model. In this stage, image resizing and image denoising are performed. The input samples are collected from various sources therefore all images are not in the same size. So that image resizing is applied to get the similar size images i.e., 224 × 224 × 3. After that image denoising is performed to make images noise free. Lastly image normalization according to equation (6), is performed for better performance of the proposed system.

$$r_{scaled} = \frac{(r - r_{min}}{(rmin_{max}} \quad (6)$$

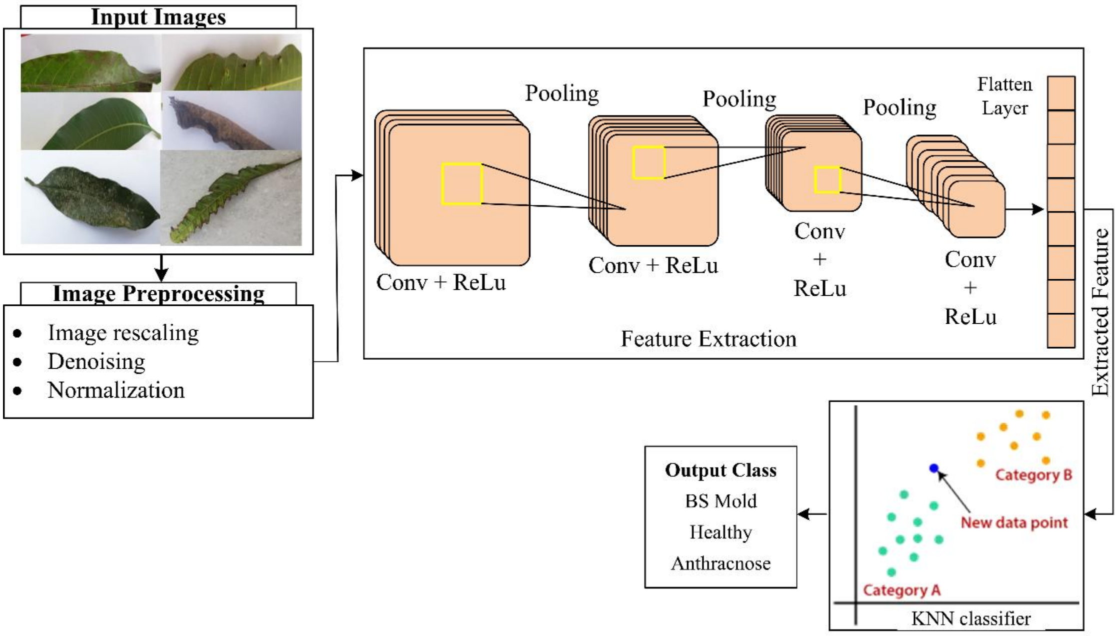

Fig. 2. Proposed model for mango leaf disease detection

3.4. Feature Extraction

In this work, an automated feature extraction approach is used that involves Convolutional Neural Networks (CNN). The used CNN model is composed of four convolutional layers, a pooling layer, a ReLu activation function, and a lastly a flatten layer. Firstly, pre-processed images are passed to the conv1 layer having an input size of 224 × 224 × 3. Here ReLu activation function is used for activating the desired neurons. The pooling layer is used to pool the extracted features for the next conv2 layer. A similar kind of operations are applied at conv2, conv3, and conv4 layers. After this, one flat layer is used for converting each feature into the proper format. The block diagram of the feature extraction CNN model is given in Fig. 3.

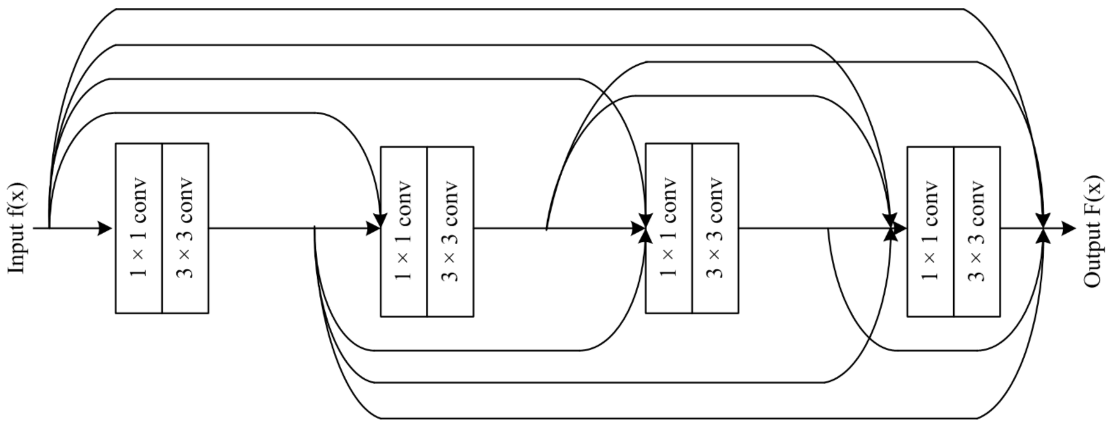

Fig. 3. Feature extraction using CNN

The proposed hybrid algorithm combining Convolutional Neural Networks (CNN) and k-Nearest Neighbors (KNN) is an interesting approach that leverages the strengths of both techniques. CNNs are powerful for feature extraction in spatial data like images, while KNN is a non-parametric method often used for classification and regression tasks.

4. KNN Classifier

If we talk about an instance-based classifier, then it is a kNN classifier. kNN classifier determines the class of a testing instance based on k-nearest neighbor in the testing instance with the use of Euclidean distance between neighboring instances. Altering the parameter k has an impact on classification performance. The value of k is optimized through repeated experimentation for classifier design by incrementing it in steps of 1 from 1 to 20. If multiple values of k yield the same performance, the smallest k is selected.

The instance illustrated in Fig. 4 represents the unknown test ROI (⅃), which should belong to either the cross class (×) or the dash class (–). When k equals 5, the algorithm searches for the five closest neighbors. The test sample is classified as a dash (–) in this instance due to the presence of three dashes and two crosses within the circle.

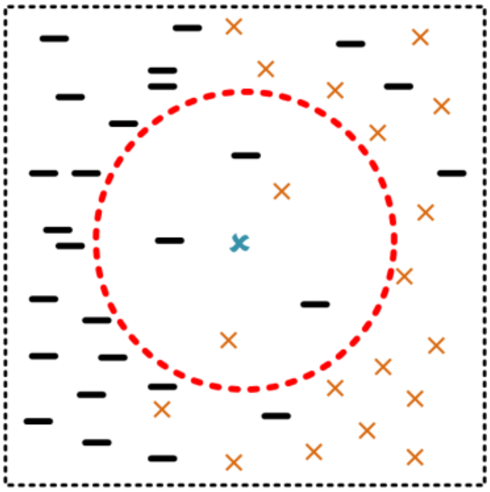

Fig. 4. Example of *k*NN classifier for k = 5

The KNN classifier is an example of a non-parametric, instance-based learning algorithm that can be employed for both classification and regression tasks. Here, we apply it to classify different leaf diseases with the aid of features retrieved through a CNN. Each input sample is represented as a feature vector as shown in equation (7).

$$X = [x_1, x_2, \dots, x_n] \quad (7)$$

The feature vector of the new data points is represented as given in equation (8). It contains the respective feature set and class label as y^i.

$$X^i = [({x^i}_1, {x^i}_2, \dots, {x^i}_n), y^i] \quad (8)$$

To find the nearest neighbors, calculate the distance between the test point and all training samples. The most common distance metric used in KNN is Euclidean distance calculated according to the formula given in equation (9).

$$d(x, x^i) = \sqrt{\sum_{j=1}^{n} (x_j - x_j^i)^2} \quad (9)$$

Other distance metrics are Manhattan distance and Minkowski distance. The formulas for the above methods are given in equations (10) and (11), respectively.

$$d(x, x^i) = \sum_{j=1}^{n} (x_j - x_j^i) \tag{10}$$

$$d(x, x^i) = \left(\sum_{j=1}^{n} \left|(x_j - x_j^i)\right|^p \right)^{1/p} \tag{11}$$

where p=1 gives Manhattan, p=2 gives Euclidean distance.

Now, compute the distance between the test point x and every training sample x^i. After obtaining the distance in ascending order, find the closest k samples. Among the k nearest neighbors, count the frequency of each class label. The class with the highest vote becomes the predicted label. The label is predicted according to the major voting scheme as given in equation (12).

$$y' = max \sum_{i \in N_k(x)} \Pi(y^i = C) \tag{12}$$

Here, C is the set of all possible class labels, Nk(x) is the indices of the k nearest neighbors of x, and $\prod$ is the indicator function (1 if the condition is true, 0 otherwise).

5. Proposed Model (CNN-KNN) Algorithm

Proposed algorithm is divided into six basic steps as follows:

Algorithm 1:

Initialization:

num_epochs=200; batch_size=50

Step 1: Train CNN on image data

def train_cnn(X_train, y_train):
cnn_model = create_cnn_model()

Step 2: Define and compile your CNN model

cnn_model.fit(X_train, y_train, epochs=num_epochs, batch_size=batch_size)
return cnn_model

Step 3: Extract features using CNN using function

def extract_features_cnn(cnn_model, X_data):
features = cnn_model.predict(X_data)
return features

Step 4: Train KNN on CNN-extracted features

Repeat for k_neighbors=1 to 20, increment by 1{
def train_knn(features, y_train_knn, k_neighbors):
knn_classifier = KNeighborsClassifier(n_neighbors=k_neighbors)
knn_classifier.fit(features, y_train_knn)
return knn_classifier}

Step 5: Proposed model prediction using the given function

X_new_data = load_new_data() # Assuming you have new data
prediction = hybrid_predict(cnn_model, knn_classifier, X_new_data)

Step 6: Print or use the prediction as needed print(prediction)

6. Experimental Setup and Performance Evaluation

Two subsections are considered here, one for experimental setup and the second is for performance Evaluation Metrics.

6.1. Experimental Setup

The suggested model was implemented using python. KNN approach using Python, TensorFlow/Keras for the CNN, and scikit-learn for the KNN.

6.2. Performance Evaluation

With the use of different metrics performance evaluation of model is done [53], [54]. Such that the model's accuracy (Acc) is evaluated as per equation number (13):

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

Precision (Pre) is evaluated as per equation number (14):

$$Pre = \frac{TP}{TP + FP} \tag{14}$$

Recall (Rec) is evaluated as per equation number (15):

$$Rec = \frac{TP}{TP + FN} \tag{15}$$

“TP,” “TN,” “FP,” and “FN” are abbreviations for “true positive,” “true negative,” “false positive,” and “false negative,” respectively.

7. Results and Discussion

7.1. Result Analysis

An extensive experiment has been carried out for the identification of mango leaf disease using CNN-based feature extraction. The classification task has been performed using a KNN classifier. For this work, the MangoleafBD dataset is considered and passed through the proposed model. Firstly, the input dataset is divided into a training and a testing set at a ratio of 70:30. The Training set is used for the training of the proposed model. The testing sample is passed to the trained model for 10-fold validation and 20-fold validation. The distribution of samples for each trial using k-fold validation is shown in Fig. 5.

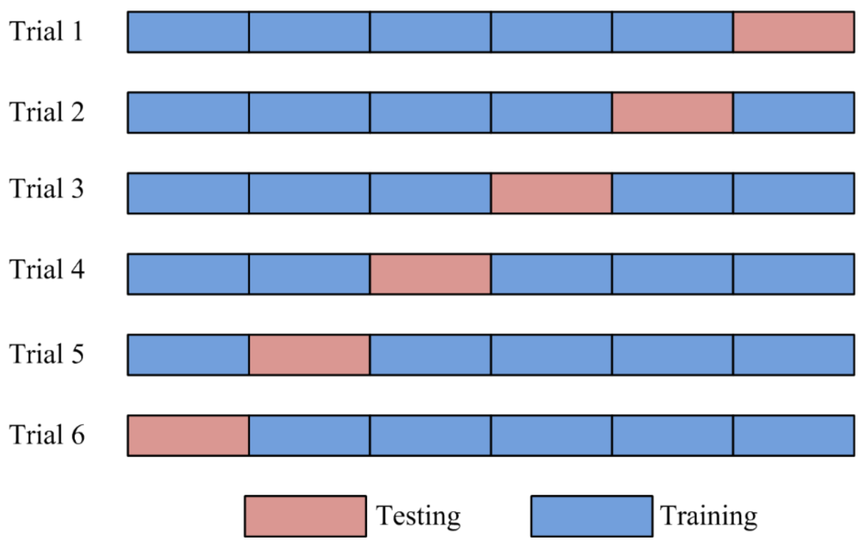

Fig. 5. Distribution of sample for each trial using k-fold validation

Based on the proposed algorithm gives an accuracy level of 99.3 % at K fold value 10 and 99.37 at K fold value 20. Comparison of proposed model with KNN and CNN based on accuracy, precision and Recall is tabulated in Table 3 for K fold value 10 and in Table 4 for k fold value 20.

Table 3. Comparison at 10 K folds

Model	Accuracy (%)	Precision (%)	Recall (%)
KNN	97.30	97.10	97.10
CNN	98.90	98.89	98.89
Proposed Model	99.30	99.28	99.27

From Table 4, it has been observed that mango leaf diseases classification has been performed using only KNN model, CNN model and proposed (CNN-KNN) model. The obtained accuracy for KNN classifier is 97.30 % and CNN classifier is 98.90 %. The proposed model yields 99.30 % of accuracy with 99.28 % of precision and 99.27 % of recall. Table 4 presents the obtained results for 20 k-folds validation results.

Table 4. Comparison at 20 K folds

Model	Accuracy	Precision	Recall
KNN	97.30	97.30	97.30
CNN	98.92	98.91	98.90
Proposed Model	99.37	99.36	99.35

It has been observed from Table 4 that KNN classifier yields 97.30 % of accuracy, 97.30 % of precision and 97.30 % of recall. Similarly, CNN yields 98.92 % of accuracy, 98.91 % of precision and 98.90 % of recall. Although, proposed model yields 99.37 % of accuracy, 99.36 % precision and 99.35 % recall. Therefore, it can be concluded that the proposed model (hybrid model) achieved outstanding accuracy for the classification of mango leaf disease identification. The obtained confusion matrix for the validation set is shown in Fig. 6. From the confusion matrix given in Fig. 6, it can be noted that out of 608, 605 test instances of the healthy class are correctly detected, and the remaining 3 instances are misclassified as anthracnose and sooty mold. Among 1508 samples of anthracnose class, 1498 samples are correctly identified, and the remaining 10 are misclassified as healthy (7) and sooty mold (3). Similarly, 1207 samples of the BS Mold class are correctly identified, and the remaining 10 samples are misclassified as the anthracnose class and the healthy class. The detailed description of the mis-classification samples is given in Table 5.

Table 5. Misclassification analysis

No. of Testing samples	Testing class labels	Label predicted by the proposed model		
		Correctly classified	Misclassified Samples	Description
3333	{608∈Healthy} {1508∈Anthracnose} {1217∈BS Mold}	{605∈Healthy} ✓	{3∈Healthy} ×	3 misclassified samples belonging to Healthy are further classified as {1∈BS Mold} and {2∈anthracnose}.
		{1498∈Anthracnose} ✓	{10∈Anthracnose} ×	10 misclassified samples belonging to Anthracnose are further classified as {7∈Healthy} and {3∈BS Mold}.
		{1207∈BS Mold} ✓	{10∈BS Mold} ×	10 misclassified samples of BS Mold class are further identified as {2∈Healthy} and {8∈ Anthracnose}.

The performance of the proposed model is also plotted in terms of the ROC curve. The obtained ROC Curve at K Fold 10 and ROC Curve at K Fold 20 are shown in Fig. 7 and Fig. 8, respectively. On the basis of the obtained confusion matrix as shown in Fig. 6, the individual class accuracy for healthy is 99.51 % (605/608), for class anthracnose is 99.34 % (1498/1508), and for class BS mold is 99.18 % (1207/1217). After this, we can perform a t-test using equation (16).

Vol. 5, No. 4, 2025, pp. 2161-2177

$$\tau = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{16}$$

Where x- is the sample mean, s is the standard deviation, and n is the sample size.

$$\bar{x} = \frac{0.9951 + 0.9934 + 0.9918}{3} = 0.9934$$
$$s = 0.00165$$
$$n = 3$$
$$\mu_0 = 0.95$$
$$t = \frac{0.9934 - 0.95}{0.00165/\sqrt{3}} = 45.54$$

With t=45.54, and degrees of freedom df = 2, the p-value will be extremely small (p << 0.05), meaning we reject the null hypothesis. It has been concluded that the classification model performs significantly better than 95 % accuracy in all classes (statistically significant, p < 0.05), based on a one-sample t-test on per-class accuracy.

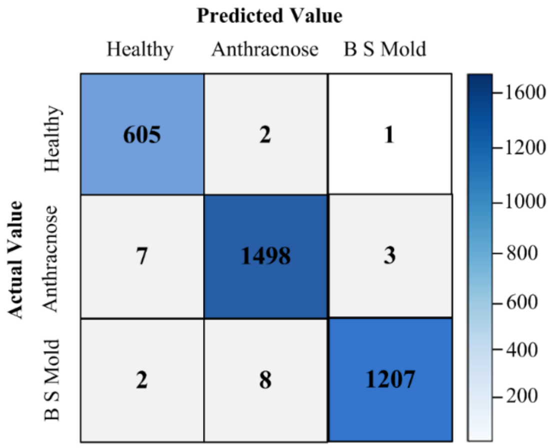

Fig. 6. Obtained confusion matrix for the proposed model

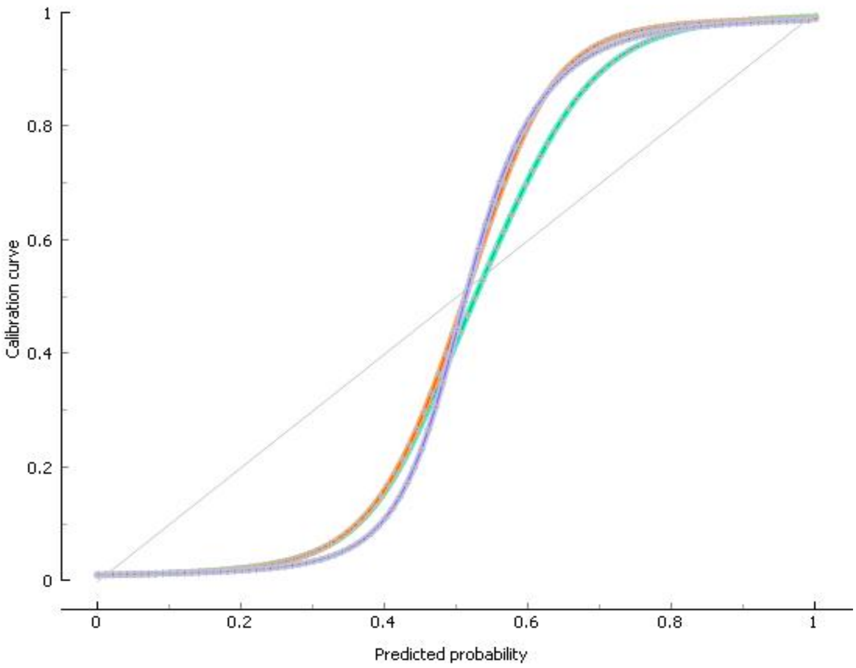

Fig. 7. ROC Cure at K Fold 10

7.2. Comparative Analysis

Comparison of proposed approach is done with other approaches of literature survey is tabulated in Table 6. Based on the results proposed approach outperform with other state of art algorithms. Table 6 presents the comparative analysis of the previously published work and the proposed work. From Table 6, it is found that the highest accuracy for previous work for the mango leaf disease detection is 99.00 % by using AlexNet. The proposed work attained 99.37 % of accuracy at 20 k-fold validation, which shows that the proposed model is suitable for the agro-engineering purpose.

Table 6. Comparison with previous work

Reference No	Used Model	Performance (%)
[41]	Support Vector Machine (SVM)	87.50
[42]	CCA fusion + SVM	95.50
[43]	FC-CNN	98.90
[44]	K-means clustering algorithm	93.12
[45]	Hybrid CNN	98.70
[46]	CNN	97.04
[47]	CNN	91.83
[48]	SVM + XGBoost	86.58
[49]	CNN	79.64
[50]	CNN + RNN	94.27
[51]	CNN + ReLu	95.30
[55]	MobileNetV3	98.00
[56]	VGG-16	78.40
[56]	AlexNet	99.00
[57]	YoloV8 with ABO	98.00
[58]	EfficientNetV2	97.13
Proposed	CNN + KNN	99.37

Note: SVM- Support Vector Machine; CNN- Convolutional neural network; RNN- Recurrent neural network; FC-CNN- Fully connected CNN.

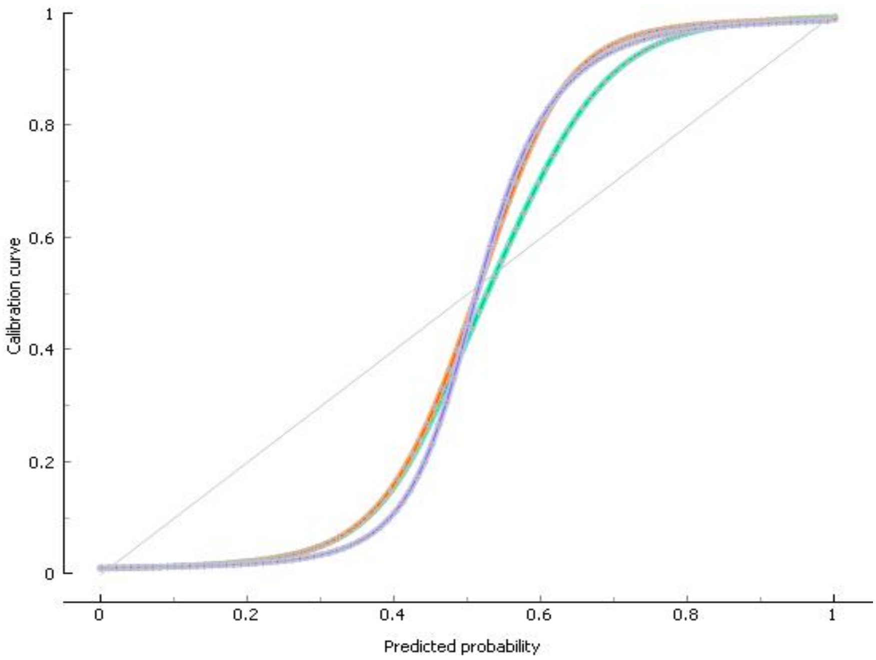

Fig. 8. ROC Curve at K Fold 20

8. Conclusions and Future Scope

The system's most powerful technique is to identify illnesses of the mango plant in an agricultural setting. The proposed algorithm gives an accuracy level of 99.3 % at K fold value 10 and 99.37% at

K fold value 20, which is enough to claim that the proposed model is yielding good results. This paper presents the results of a study that used a convolutional neural network and feature selection to detect and classify diseases in leaves. A combination of a deep neural network and feature selection has been used to detect diseases with great performance. These methods can aid farmers in spotting diseases and protecting their crops from harm.

The agricultural industry stands to gain significantly from the upcoming directions of this research, which will cover a lot of ground. We may start by building on the concept of a hierarchical combination of labels to make it even better. One possibility is to be an early adopter of an algorithm that can figure out the best balance weights on its own. This breakthrough has tremendous promise, especially in situations where there are many publicly accessible datasets pertaining to different plant species and illnesses. At various prediction levels, we can also investigate the possibility of using deep supervision to mitigate training-level gradients. The same work has been extended to the other types of disease occurring in the mango leaf. By using such a type of concept, precise farming gains more popularity.

Data Availability: The data used to support the results of the study may be obtained from the corresponding author.

Funding: This study did not receive any funding.

Conflict of Interest: There is no conflict of interest regarding the publication of this paper.

References

[1] M. T. S. S. Sharma, H. Kumar, "Detection and Classification of Plant Diseases by Alexnet and Googlenet Deep Learning Architecture," *International Journal of Scientific Research & Engineering Trends*, vol. 8, no. 1, pp. 218-223, 2022, https://ijsret.com/wp-content/uploads/2022/01/IJSRET_V8_issue1_120.pdf.

[2] S. Arivazhagan, R. N. Shebiah, S. Ananthi, S. V. Varthini, "Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features," *Agricultural Engineering International: CIGR Journal*, vol. 15, no. 1, pp. 211-217, 2013, https://cigrjournal.org/index.php/Ejounral/article/view/2338.

[3] Z. Iqbal, M. A. Khan, M. Sharif, J. H. Shah, M. H. ur Rehman, and K. Javed, "An automated detection and classification of citrus plant diseases using image processing techniques: A review," *Computers and electronics in agriculture*, vol. 153, pp. 12-32, 2018, https://doi.org/10.1016/j.compag.2018.07.032.

[4] J. K. Patil, R. Kumar, "Advances in image processing for detection of plant diseases," *Journal of Advanced Bioinformatics Applications and Research*, vol. 2, no. 2, pp. 135-141, 2011, https://bipublication.com/files/JABARv2i2201103.pdf.

[5] A. Camargo and J. S. Smith, "An image-processing based algorithm to automatically identify plant disease visual symptoms," *Biosystems Engineering*, vol. 102, no. 1, pp. 9-21, 2009, https://doi.org/10.1016/j.biosystemseng.2008.09.030.

[6] 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, 2010, https://doi.org/10.1016/j.compag.2010.02.007.

[7] C. Trongtorkid and P. Pramokchon, "Expert system for diagnosis mango diseases using leaf symptoms analysis," *2018 International Conference on Digital Arts, Media and Technology (ICDAMT)*, pp. 59-64, 2018, https://doi.org/10.1109/ICDAMT.2018.8376496.

[8] 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.

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

[10] G. S. Tumang, "Pests and Diseases Identification in Mango using MATLAB," *2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST)*, pp. 1-4, 2019, https://doi.org/10.1109/ICEAST.2019.8802579.

[11] S. B. Ullagaddi and S. V. Raju, "Disease recognition in Mango crop using modified rotational kernel transform features," *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pp. 1-8, 2017, https://doi.org/10.1109/ICACCS.2017.8014610.

[12] Y.-N. Wang, L. Jin, and H. Mao, "Farmer cooperatives' intention to adopt agricultural information technology-mediating effects of attitude," *Information Systems Frontiers*, vol. 21, no. 3, pp. 565-580, 2019, https://doi.org/10.1007/s10796-019-09909-x.

[13] B. Zhang, Z. Wang, C. Ye, H. Zhang, K. Lou, and W. Fu, "Classification of infection grade for anthracnose in mango leaves under complex background based on CBAM-DBIRNet," *Expert Systems With Applications*, vol. 260, p. 125343, 2025, https://doi.org/10.1016/j.eswa.2024.125343.

[14] T. Varma, P. Mate, N. A. Azeem, S. Sharma, and B. Singh, "Automatic mango leaf disease detection using different transfer learning models," *Multimedia Tools and Applications*, vol. 84, no. 11, pp. 9185-9218, 2025, https://doi.org/10.1007/s11042-024-19265-x.

[15] P. Patel, A. Prajapati, and V. Koradia, "and Diagnosis of Mango Leaf Diseases Using Machine Learning," *ICT Systems and Sustainability*, vol. 2, pp. 247-256, 2025, https://doi.org/10.1007/978-981-97-8591-9_22.

[16] S. P. S. Shivam, I. Kumar, "Rice plant infection recognition using deep neural network systems," *International Semantic Intelligence Conference (ISIC 2021)*, pp. 384-393, 2021, https://ceur-ws.org/Vol-2786/Paper47.pdf.

[17] V. Gautam, R. K. Ranjan, P. Dahiya, and A. Kumar, "ESDNN: A novel ensembled stack deep neural network for mango leaf disease classification and detection," *Multimedia Tools and Applications*, vol. 83, no. 4, pp. 10989-11015, 2024, https://doi.org/10.1007/s11042-023-16012-6.

[18] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, pp. 237-285, 1996, https://doi.org/10.1613/jair.301.

[19] X. Fu, J. Zhang, X. Bai, X. Chang, and Y. Xue, "Review and outlook on reinforcement learning: Its application in agricultural energy internet," *IET Renewable Power Generation*, vol. 18, no. 16, pp. 3678-3690, 2024, https://doi.org/10.1049/rpg2.13019.

[20] G. Mohyuddin, M. A. Khan, A. Haseeb, S. Mahpara, M. Waseem and A. M. Saleh, "Evaluation of Machine Learning Approaches for Precision Farming in Smart Agriculture System: A Comprehensive Review," *IEEE Access*, vol. 12, pp. 60155-60184, 2024, https://doi.org/10.1109/ACCESS.2024.3390581.

[21] M. Omid, M. Soltani, M. H. Dehrouyeh, S. S. Mohtasebi, and H. Ahmadi, "An expert egg grading system based on machine vision and artificial intelligence techniques," *Journal of Food Engineering*, vol. 118, no. 1, pp. 70-77, 2013, https://doi.org/10.1016/j.jfoodeng.2013.03.019.

[22] M. Dorigo, M. Birattari and T. Stutzle, "Ant colony optimization," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28-39, 2006, https://doi.org/10.1109/MCI.2006.329691.

[23] A. H. Gandomi, X.-S. Yang, and A. H. Alavi, "Cuckoo search algorithm: a metaheuristic approach to solve structural optimization problems," *Engineering with Computers*, vol. 29, no. 1, pp. 17-35, 2013, https://doi.org/10.1007/s00366-011-0241-y.

[24] I. Kumar, H. S. Bhadauria, J. Virmani, and S. Thakur, "A hybrid hierarchical framework for classification of breast density using digitized film screen mammograms," *Multimedia Tools and Applications*, vol. 76, pp. 18789-18813, 2017, https://doi.org/10.1007/s11042-016-4340-z.

[25] M. Mareli and B. Twala, "An adaptive Cuckoo search algorithm for optimisation," *Applied Computing and Informatics*, vol. 14, no. 2, pp. 107-115, 2018, https://doi.org/10.1016/j.aci.2017.09.001.

[26] A. S. Joshi, O. Kulkarni, G. M. Kakandikar, and V. M. Nandedkar, "Cuckoo search optimization-a review," *Materials Today: Proceedings*, vol. 4, no. 8, pp. 7262-7269, 2017, https://doi.org/10.1016/j.matpr.2017.07.055.

[27] M. Yazdani and F. Jolai, "Lion optimization algorithm (LOA): a nature-inspired metaheuristic algorithm," *Journal of Computational Design and Engineering*, vol. 3, no. 1, pp. 24-36, 2016, https://doi.org/10.1016/j.jcde.2015.06.003.

[28] S. Saremi, S. Mirjalili, and A. Lewis, "Grasshopper optimisation algorithm: theory and application," *Advances in Engineering Software*, vol. 105, pp. 30-47, 2017, https://doi.org/10.1016/j.advengsoft.2017.01.004.

[29] C. Charin, D. Ishak, M. A. A. Mohd Zainuri, and B. Ismail, "Modified levy Flight optimization for a maximum power point tracking algorithm under partial shading," *Applied Sciences*, vol. 11, no. 3, p. 992, 2021, https://doi.org/10.3390/app11030992.

[30] A. Cheraghalipour, M. Hajiaghaei-Keshteli, and M. M. Paydar, "Tree Growth Algorithm (TGA): a novel approach for solving optimization problems," *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 393-414, 2018, https://doi.org/10.1016/j.engappai.2018.04.021.

[31] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Advances in Engineering Software*, vol. 69, pp. 46-61, 2014, https://doi.org/10.1016/j.advengsoft.2013.12.007.

[32] T.-K. Dao, J.-S. Pan, T.-S. Pan, and T.-T. Nguyen, "Optimal path planning for motion robots based on bees pollen optimization algorithm," *Journal of Information and Telecommunication*, vol. 1, no. 4, pp. 351-366, 2017, https://doi.org/10.1080/24751839.2017.1347765.

[33] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849-872, 2019, https://doi.org/10.1016/j.future.2019.02.028.

[34] H. Abderazek, F. Hamza, A. Yildiz, L. Gao, S. Sait, "A comparative analysis of the queuing search algorithm, the sine-cosine algorithm, the ant lion algorithm to determine the optimal weight design problem of a spur gear drive system," *Materials Testing*, vol. 63, no. 5, pp. 442-447, 2021, https://doi.org/10.1515/mt-2020-0075.

[35] D. Gürses, N. Pholdee, S. Bureerat, S. Sait, A. Yıldız, "A novel hybrid water wave optimization algorithm for solving complex constrained engineering problems," *Materials Testing*, vol. 63, no. 6, pp. 560-564, 2021, https://doi.org/10.1515/mt-2020-0093.

[36] H. Abderazek, F. Hamza, A. Yildiz, S. Sait, "Comparative investigation of the moth-flame algorithm and whale optimization algorithm for optimal spur gear design," *Materials Testing*, vol. 63, no. 3, pp. 266-271, 2021, https://doi.org/10.1515/mt-2020-0039.

[37] G. Wang, "Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems," *Memetic Computing*, vol. 10, no. 2, pp. 151-164, 2018, https://doi.org/10.1007/s12293-016-0212-3.

[38] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: a new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300-323, 2020, https://doi.org/10.1016/j.future.2020.03.055.

[39] A. H. Gandomi and A. H. Alavi, "Krill herd: a new bio-inspired optimization algorithm," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 12, pp. 4831-4845, 2012, https://doi.org/10.1016/j.cnsns.2012.05.010.

[40] M. Tubishat, M. Alswaitti, S. Mirjalili, M. A. Al-Garadi, M. T. Alrashdan and T. A. Rana, "Dynamic Butterfly Optimization Algorithm for Feature Selection," *IEEE Access*, vol. 8, pp. 194303-194314, 2020, https://doi.org/10.1109/ACCESS.2020.3033757.

[41] M. R. Mia, S. Roy, S. K. Das, and M. A. Rahman, "Mango leaf disease recognition using neural network and support vector machine," *Iran Journal of Computer Science*, vol. 3, pp. 185-193, 2020, https://doi.org/10.1007/s42044-020-00057-z.

[42] 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, p. 11901, 2021, https://doi.org/10.3390/app112411901.

[43] R. Saleem, J. H. Shah, M. Sharif, and G. J. Ansari, "Mango leaf disease identification using fully resolution convolutional network," *Computers, Materials & Continua*, vol. 69, no. 3, pp. 3581-3601, 2021, https://doi.org/10.32604/cmc.2021.017700.

[44] S. Khan and M. Narvekar, "Novel fusion of color balancing and superpixel based approach for detection of tomato plant diseases in natural complex environment," *Journal of King Saud University-Computer and Information Sciences*, vol. 43, no. 6, pp. 3506-3516, 2020, https://doi.org/10.1016/j.jksuci.2020.09.006.

[45] P. Kaur, S. Harnal, R. Tiwari, S. Upadhyay, S. Bhatia, A. Mashat, and A. M. Alabdali, "Recognition of leaf disease using hybrid convolutional neural network by applying feature reduction," *Sensors*, vol. 22, no. 2, p. 575, 2022, https://doi.org/10.3390/s22020575.

[46] S. Bensaadi and A. Louchene, "Low-cost convolutional neural network for tomato plant diseases classifiation," *IAES International Journal of Artificial Intelligence*, vol. 12, no. 1, p. 162, 2023, http://doi.org/10.11591/ijai.v12.i1.pp162-170.

[47] J. Lu, L. Tan, and H. Jiang, "Review on convolutional neural network (CNN) applied to plant leaf disease classification," *Agriculture*, vol. 11, no. 8, p. 707, 2021, https://doi.org/10.3390/agriculture11080707.

[48] M. A. Azim, M. K. Islam, M. M. Rahman, and F. Jahan, "An effective feature extraction method for rice leaf disease classification," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 2, pp. 463-470, 2021, http://doi.org/10.12928/telkomnika.v19i2.16488.

[49] L. C. Ngugi, M. Abdelwahab, and M. Abo-Zahhad, "A new approach to learning and recognizing leaf diseases from individual lesions using convolutional neural networks," *Information Processing in Agriculture*, vol. 10, no. 1, pp. 11-27, 2021, https://doi.org/10.1016/j.inpa.2021.10.004.

[50] M. Altalak, M. Ammad uddin, A. Alajmi, and A. Rizg, "Smart agriculture applications using deep learning technologies: A survey," *Applied Sciences*, vol. 12, no. 12, p. 5919, 2022, https://doi.org/10.3390/app12125919.

[51] N. E. M. Khalifa, M. H. N. Taha, A. El-Maged, M. Lobna, and A. E. Hassanien, "Artificial intelligence in potato leaf disease classification: a deep learning approach," *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges*, pp. 63-79, 2021, https://doi.org/10.1007/978-3-030-59338-4_4.

[52] S. I. Ahmed, M. Ibrahim, M. Nadim, M. M. Rahman, M. M. Shejunti, T. Jabid, and M. S. Ali, "MangoLeafBD: A comprehensive image dataset to classify diseased and healthy mango leaves," *Data in Brief*, vol. 47, p. 108941, 2023, https://doi.org/10.1016/j.dib.2023.108941.

[53] I. Kumar, A. Bisht, and J. Rawat, "Identification of Tomato Plant Disease Using Faster R-CNN and RetinaNet," *Convergence of Cloud Computing, AI, and Agricultural Science*, pp. 306-327, 2023, https://www.irma-international.org/viewtitle/329141/?isxn=9798369302002.

[54] P. Kohli, I. Kumar, and V. Vimal, "Plant leaf disease identification using unsupervised fuzzy c-means clustering and supervised classifiers," *Cyber security in intelligent computing and communications*, pp. 281-293, 2022, https://doi.org/10.1007/978-981-16-8012-0_22.

[55] S. S. Puranik *et al*., "MobileNetV3 for Mango Leaf Disease Detection:An efficient Deep Learning Approach for Precision Agriculture," *2024 5th International Conference for Emerging Technology (INCET)*, pp. 1-7, 2024, https://doi.org/10.1109/INCET61516.2024.10593318.

[56] R. A. Rizvee *et al*., "LeafNet: A proficient convolutional neural network for detecting seven prominent mango leaf diseases," *Journal of Agriculture and Food Research*, vol. 14, p. 100787, 2023, https://doi.org/10.1016/j.jafr.2023.100787.

[57] V. K. Pratap and N. S. Kumar, "Deep Learning based Mango Leaf Disease Detection for Classifying and Evaluating Mango Leaf Diseases," *Fusion: Practice & Applications*, vol. 15, no. 2, pp. 261-277, 2024, https://doi.org/10.54216/FPA.150222.

[58] P. V, S. Venkatraman, A. A, A. S. A and P. K. S, "Mango Leaf Disease Detection and Classification Using Spatial Attention Enabled Ensemble Classification," *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*, pp. 1-8, 2024, https://doi.org/10.1109/ADICS58448.2024.10533513.