A Machine Learning Hybrid Approach for Diagnosing Plants Bacterial and Fungal Diseases

Ahmed BaniMustafa[1*], Hazem Qattous[2], Ihab Ghabeish[3] and Muwaffaq Karajeh[4]
Data Science & Artificial Intelligence Department, Isra University, Amman, Jordan[1*]
Software Engineering Department, Princes Sumaya University for Technology, Amman, Jordan[2]
Department of Agricultural Sciences-As-Shoubak University College, Al-Balqa' Applied University, Al-Salt, Jordan[3]
Department of Plant Protection and Integrated Pest Management, Mutah University, Karak, Jordan[4]

***Abstract*—Bacterial and Fungal diseases may affect the yield of stone fruit and cause damage to the Chlorophyll synthesis process, which is crucial for tree growth and fruiting. However, due to their similar visual shot-hole symptoms, novice agriculturalists and ordinary farmers usually cannot identify and differentiate these two diseases. This work investigates and evaluates the use of machine learning for diagnosing these two diseases. It aims at paving the way toward creating a generic deep learning-based model that can be embedded in a mobile phone application or in a web service to provide a fast, reliable, and cheap diagnosis for plant diseases which help reduce the excessive, unnecessary, or improper use of pesticides, which can harm public health and the environment. The dataset consists of hundreds of samples collected from stone fruit farms in the north of Jordan under normal field conditions. The image features were extracted using a CNN algorithm that was pre-trained with millions of images, and the diseases were identified using three machine learning classification algorithms: (1) K-nearest neighbour (KNN); (2) Stochastic Gradient Descent (SGD); and (3) Random Forests (RF). The resulting models were evaluated using 10-fold cross-validation, with CNN-KNN achieving the best AUC performance with a score of 98.5%. On the other hand, the CNN-SGD model performed best in Classification Accuracy (CA) with a score of 93.7%. The results shown in the Confusion Matrix, ROC, Lift, and Calibration curves also confirmed the validity and robustness of the constructed models.**

Keywords*—Deep Learning; Machine Learning; Classification; Plant Diseases; Disease Diagnosis

I. INTRODUCTION

In Jordan and other Mediterranean countries, agriculture participates with a high share of the annual GDP, and stone fruit trees are among the most important sources of crop yield [1]. However, Bacterial and Fungal diseases, which appear as shot holes in the leaves, may affect the yield of the trees and may eventually cause damage to the Chlorophyll synthesis process, which is crucial for plant growth and fruiting.

Shot hole disease is caused by several pathogens. The fungal shot hole, or the *Coryneum* blight, is caused by *Wilsonomyces carpophilus (Lev.)* [2]. This fungus is common in the Mediterranean region, but it also spreads in America, Australia, Africa, Asia, Oceania, and Europe [3]. It infects stone fruit trees of Prunus, mainly apricot, peach, nectarine, and cherry. However, the first three are the most commonly affected hosts [4], [5]. It damages twigs, buds, blossoms, fruits, and leaves. The damage is most noticeable on the leaves, however. The causal agent thrives in cool and moist conditions between summer and fall, mainly in early spring and anytime during wet weather conditions [3], [6]. It develops rapidly under warmer temperatures [3] and overwinters on blossom buds, and cankers of branches [7]. The symptoms appear on the infected leaves as small, round reddish to purplish-brown specks with light green or yellow rings around them. Lesions can be circular to slightly ellipsoid; their tissues can become raised and scurfy and will tear along the lesion margins and may hang on at one attached point. It often dries up and falls away, giving the shot hole; as if someone fired a shotgun at the leaf [5].

As the disease spreads, more leaves get damaged until they fall. Significant infections can reduce photosynthesis, weaken the plant, and decrease the fruit yield eventually, which makes it a major concern for the stone fruit industry worldwide [8]. The other causal agent of the shot hole disease is the bacterium *Xanthomonas Arboricola PV. Pruni (Xap)* [9]. The symptoms of this bacterium are circular to irregular, water-soaked lesions on leaves. Later, lesions turn purple or brown. Usually, halos and cracks can be seen between the affected tissue and the surrounding healthy tissue. In the later stage, the infected tissue will be broken away under various natural forces, especially wind, and finally drop out, leaving a hole. Leaves with many holes or lesions will turn yellow or prematurely drop off [9]. It infects leaves, twigs, and fruits of the same stone fruit hosts of the *W. Carpophilus* fungus. The favourable conditions for the disease occurrence and spread are largely the same as the fungal shot hole, but it overwinters mainly in twig lesions (OEPP/EPPO, 2003). The disease is considered the most limiting factor to cherry Laurel production in landscapes and nursery production. Losses exceed 75% in the nursery due to this disease (unpublished data, J. Williams-Woodward). However, these diseases, which affect different parts of the trees and produce holes in leaves, are caused by different pathogens such as*Wilsonomyces Carpophilus* fungus and *Xanthomonas Arboricola PV. Pruni* bacteria. Sometimes it is difficult to differentiate the symptoms to identify the pathogen, even further, to identify if the symptoms are caused by a pathogen or insect feeding. This makes it difficult to specify suitable pesticides or implement a control plan to avoid damage to the trees.

Nonetheless, the infections caused by these diseases are difficult to identify and cure as they are usually confused with viral pest infections and with other soil and nutrition deficiencies. The accurate identification and diagnosis of these diseases infections through visual signs and symptoms may require deep knowledge and thorough experience, which is usually lacking in most ordinary farmers, and yet they may also require laboratory tests and identification procedures, which are usually expensive to perform for most farmers. Yet, fighting these bacterial and fungal infections without a proper diagnosis may fail or may require using a wide spectrum of pesticides, placing additional financial costs which add to the short-term losses in the fruit yield and the long-term damages made to the trees, which may affect their lifespan. In addition, the use of pesticides may harm the trees and may also place concerns over food safety, which may cause several marketing challenges due to allowable limits imposed by food and drug administrations and may ultimately affect customer confidence. This all adds to other environmental concerns that are related to pesticide leaks into the soil and water and their short-term and long-term influence on wildlife.

Novice agriculturalists and farmers not knowing much about the disease may confuse the infections with other viral and pest diseases and deficiencies caused by factors related to soil, water, or other environmental conditions. They might also confuse these diseases with symptoms associated with other issues related to the incorrect use of pesticides and fertilisers. Therefore, the false identification and diagnosis of fungal and bacterial may waste farmers' time and effort in solving irrelevant issues. This in fact, lay additional financial pressure on farmers and may also cause other serious environmental and health-related issues, such as those related to the improper use of pesticides or unintentional insect feeding. Moreover, the spread of fungal and bacterial diseases in various environments may also contribute to the evolution of new races of causal agents with different symptoms, thus making them hard to diagnose under normal field conditions [6].

Machine learning techniques in general [10], and deep learning algorithms in particular such as convolutional neural networks [11], [12], [13] have recently witnessed impressive success in several scientific and commercial applications, particularly in the fields of computer vision, image recognition [14], [15], [16], and other classification applications [17], [18], [19]. One of the potential applications of machine learning is identifying and diagnosing plant diseases' infections [20], and their other related issues [21]. In this research, we have conducted an end-to-end empirical study that involves collecting, diagnosing, and identifying many stone fruit leaf samples that suffer from bacterial and fungal infections. The sample leaves have been photographed, categorised and then used to train and test three machine learning classifier models, which are created by combining convolutional neural networks with three classical machine learning algorithms: (1) K-nearest neighbour (KNN); (2) Stochastic Gradient Descent (SGD); and (3) Random Forests (RF). These algorithms were selected due to their reported success in image classification. All the constructed models will be evaluated using Classification Accuracy (CA), Area Under the Curve (AUC), Precision and F1 metrics, in addition, the Confusion Matrix, ROC, Lift and Calibration curves will also be used to confirm the validity and robustness of the created models.

The following four subsections are dedicated to investigating the related work and defining the research hypothesis, question, and aims. Section II provides details regarding the research methodology, materials and data collection, and an overview of the techniques applied and their performance metrics. Section III presents the results of model building and evaluation which is followed by Section IV, which analyses the results obtained in this work in light of the research hypothesis, question, and objectives. It also compares the results with those reported by others. Section V provides a conclusion of the findings and comments of the result on the limitation of the study and highlights the contribution of this study. It also comments on the significance of the results and their potential applicability, as well as on the possible future work that is related to the study.

A. Related Work

In this subsection we investigate nine of the most popular related works and most cited studies that have been published in the last six years.

A study that was published in [22] reported 78% accuracy in classifying diseases that infect apple trees using the field-collected dataset, while [23] reported accuracy of 81% in classifying four grapes diseases using images that have been captured under field normal condition. On the other hand, [24] reported accuracy of 82% in classifying cucumber diseases using a research centner dataset, while [25] reported an accuracy of 83% in classifying disease that infects potato using images of their tubers which have been collected from potato fields. [26] reported yet another field-based study that involved classifying diseases that infarct tea plants with an accuracy of 90%. [27] reported accuracy of 88% in classifying banana diseases using the plant village public dataset which is very close to the results reported by [28] which aimed at detecting and classifying diseases that infect tomato leaves and which also depended on using a public dataset that was downloaded from plant village. [29] reported an accuracy of 99% in classifying soybeans disease also using plant village public datasets. [30] also reported a 99% of classification accuracy using the plant village dataset, but this time for classifying twenty-six diseases that hit fourteen different crops. A recent study reported in [31] achieved classification accuracy of 96.63% using the ResNet-50 deep learning algorithm to predict 15 disease classes using 20,000 public dataset that is published by plant village. Another study published by [32] reported an accuracy of 96% using unsupervised learning. However, the study provides no details regarding the number of the samples in the dataset. Another study published in [33], reported an accuracy of 92.57% using YOLOv5 model for classifying 61 categories of plant diseases that hit ten different plant species. The dataset consisted of 36,258 images. A more recent study that used Multiple Linear Regression (MLP) reported 91% accuracy in identifying blister blight in tea plants[34], while another recent study reported the use of CNN algorithm to detect 13 different diseases achieved precision of 96.3% [35].

The analysis of the related work shows that most of the reported applied techniques scored a classification accuracy which ranged between 78% and 99%. It was noticed that the majority of the investigated works reported a classification accuracy above 90% depended on huge public datasets which involved thousands of images that have been captured in a controlled environment with a standardised camera type, images orientation, images resolution, images aspect ratio, lightning and sometimes the growth conditions of the plant itself. On the other hand, the studies that reported performance of less than 90% depended on using a less number of images (a few hundred images) captured using ordinary cameras and for leaves of plants that were grown in field normal conditions. Furthermore, only few of the reviewed studies spent effort in diagnosing the plant diseases and investigating their cause and pathogens. Most of the other studies depended only on the disease's visual symptoms, which shade doubt on the validity of their findings. In addition, few of the surveyed studies commented on the utilisation and the practical use of their obtained results.

B. Research Hypothesis

The research suggests that deep learning pre-trained algorithms can be used for the automatic identification of image features and then be classified using other machine learning classification algorithms to provide an accurate diagnosis of stone fruit fungal and bacterial diseases.

C. Research Question

The research question seeks to answer the question: Can we build a successful machine learning model to diagnose stone fruit fungal and bacterial diseases?

D. Research Aim

Creating a machine learning model for diagnosing stone fruit diseases will pave the way towards creating a generic deep learning-based model that can be embedded in a mobile phone application or a web service for the purpose of providing a fast, reliable, and cheap diagnosis of plant diseases, which helps to reduce the excessive, unnecessary, or improper use of pesticides in agriculture, which can be harmful to public health and the environment.

II. MATERIALS AND METHODS

This section provides a brief description of the technology applied and the methodology implemented in order to automate and tackle the problem of stone fruit disease diagnosis in this research to tackle.

A. Materials

The dataset consists of 500 sample images. 294 of the images belong to the Fungal class, while 106 images belong to the Bacterial class. 99 belong to the control healthy class. The dataset creation involved collecting and photographing hundreds of images of stone fruit that are infected with the bacterium *Xanthomonas arboricola Pv. Pruni (Xap)* and *Wilsonomyces Carpophilus* fungus.

The leaves were collected from various farms in the north of Jordan during the mid of August, which are located in a region with a moderate Mediterranean mountainous climate with an elevation between 630-950 Metres above sea level. The average temperature between April and September is 29.7 °C during the day and 17.2 °C, while the average humidity is 45% and the average pressure is 1008 mBar. The rain is very scarce during the data collection period. Fig 1 shows sample leaves that suffer from bacterial and fungal infections. The samples were collected and photographed in the field normal conditions.

The dataset has an equal number of images for each infection class. The stone fruit leaves were first collected and then photographed using a mobile phone camera with an 8M pixels resolution. The leaves were then examined by two domain experts who work as plant diseases experts and professors at al-Balqa applied University and Mutah University. The disease experts diagnosed the infection in each leaf and placed the leaves' images in their corresponding folders. The diagnosis of each leaf was then confirmed using plant disease literature and also by using the plant disease database sponsored by the European and Mediterranean Plant Protection Organisation (EPPO).

B. Methods

The methods applied in this research involve performing three major steps: (1) The extraction of significant image features using deep learning convolutional neural networks VGG16 algorithms ; (2) the classification of images into Bacterial and Fungal classes using the KNN, SGD and RF machine learning algorithms; and (3) models evaluation based on 10-folds cross-validation and using confusion matrix, ROC, AUC, classification Accuracy and Precision metrics. Fig. 2 illustrates the three major steps in the proposed research methodology.

1) Features Extraction: The features in each image are extracted using the convolutional neural network VGG16 implementation that was pre-trained using 14 million images from the ImageNet database which correspond to 1000 labelled classes[36]. The algorithm was originally proposed by Simonyan and Zisserman to introduce several improvements to AlexNet which was developed earlier by google [37], [38].

The applied implementation was based on the Keras framework using Python and Caffe framework. Keras is a deep learning Python Library that focuses on simplifying deep learning model construction and visualising its using Tensor Flow, while Caffe is a deep learning framework that was developed by the BAIR AI research group at the University of California, Berkeley [39]. It enables constructing expressive, modularised and fast deep learning models [40], [41]

The constructed vectors for each image feature were then preprocessed, first by normalising the image features values within the interval between -1 and 1 and then, by removing the sparse features for each feature with missing values exceeding the ratio of 5%.

Deep learning is one of the most popular contemporary technology in machine learning that has recently witnessed unprecedented success in several applications, particularly in image and voice recognition. as it requires no or minimal feature engineering[11]. Deep learning is simply, a back propagation neural network algorithm that is based on creating a machine learning model using several layers, where each represents a level of abstraction[11].

Convolutional neural networks (CNN) are a class of deep learning algorithms that uses feed-forward neural networks to extract and learn image features using multiple layers[11], [42]. CNN has achieved success in several machine learning applications, particularly in image processing, recognition and classification. In convolutional neural networks, and each layer,

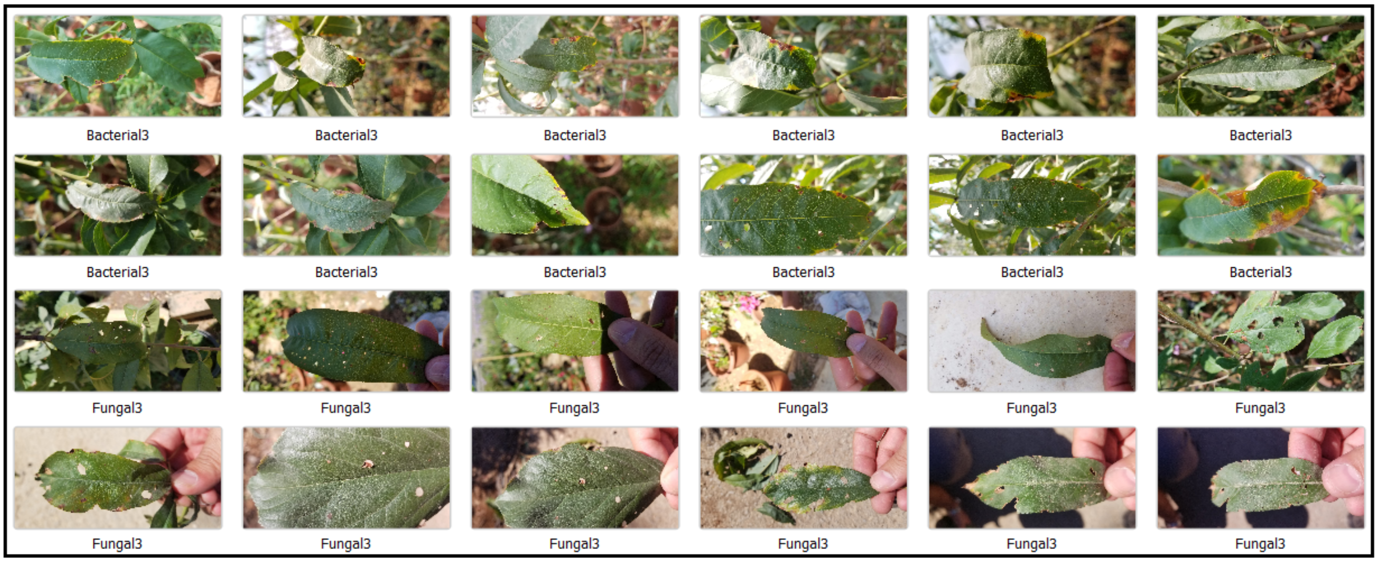

Fig. 1. Image samples that were collected and photographed in the field normal conditions for stone fruit leaves that suffer from fungal and bacterial infections.

the image dimensionality is reduced and its features are automatically identified, extracted, stored in vectors and then passed to a more specialised deeper layer which extracts even more features using its own ”convolutions”. A convolution does a job in terms of feature extraction that is analogous to light filters in optics[42], [43], [11]. Convolutional Neural Network (CNN) algorithms were used for extracting and analysing image features. The identified feature was then used for creating three machine learning classification models for identifying the stone fruit fungal and bacterial diseases using a dataset that consists of hundreds of images that have been captured for the infected leaves. Convolutions are used to learn the image features in a fashion analogous to filters, in each layer the dimensionality of the image is reduced using convolution and then fed into the lower layer [43]. The convolution at the top layer is used to filter, extract and learn the most generic features in the image and then pass them down toward the convolutions in the deeper layers that are used to filter, extract and learn the more specific features as illustrated in Fig. 2.

2) Classification: Machine learning (ML) concerns enabling computers to learn and improve from previous examples and prior experience using a wide spectrum of artificial intelligence techniques and algorithms[44], [45], [46]. Modern technology and recent advances in machine learning are the driving force behind what is called ”The fourth industrial revolution.”

Machine learning tries to mimic the ability of human beings and other intelligent species to learn and attain knowledge from prior experience and examples and then generalise and use this knowledge to respond to future situations and process unknown data. In machine learning classifiers, a model is built and then trained to predict the outcome or responses in a precooked dataset which is called ”training data.” The model is then validated by testing its ability to predict the unknown outcomes and responses in unknown datasets or situations, which is called ”validation data.”

Selecting the appropriate machine learning modelling technique is important and must take into consideration the nature of the dataset, study objectives, and the potential of the applied technique [46], [45], [47]. In this research, three algorithms have been applied to the targeted images and then evaluated to find the best machine learning algorithm that can be coupled with the deep learning convolutional neural networks VGG16 algorithm to diagnose the targeted leaves and classify their infection. The applied algorithms involve KNN, SGD, and RF. These algorithms will be introduced and discussed in detail in this section.

K-Nearest Neighbours Algorithm (KNN): An instance-based learning technique that is used for classifying data samples by measuring their proximity to neighbouring data points that belong to a set of pre-labelled classes. This technique was first introduced by Evelyn Fix and JL Hodges, Jr. in their unpublished technical report while working at the USAF School of aviation [48]. KNN measures the Euclidean distance between predicted and training values belonging to a predefined class in a two-dimensional space using equation 1. The predicted classes for a point are determined based on a popularity vote regarding its distance from other neighbouring data points that belong to the neighbouring classes. KNN was used in several image classification applications that included classifying brain CT Scan [49] and recognising images of various objects [50]. In both applications, it achieved a classification accuracy of 80%.

$$\Delta(x_i, x_j) = \sqrt{\sum_{i=1}^{n} (\mid x_{in} - x_{jn} \mid)^2} \tag{1}$$

Stochastic Gradient Descent (SGD): This algorithm classifies samples iteratively. In each iteration, the weights of the classification model are updated using equation 2.

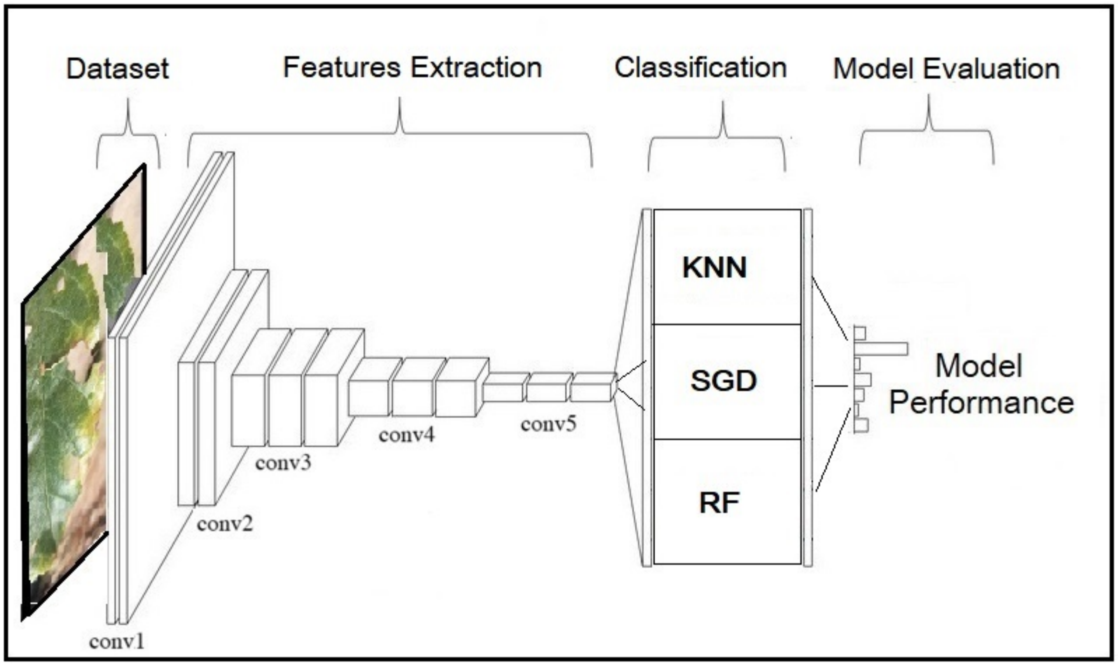

Fig. 2. An illustration of the applied research methodology, where CNN is used for extracting images' features and that are then fed to KNN, SGD, and RF classification algorithms.

$$\omega_{t+1} = \omega_t - \gamma \frac{1}{n} \sum_{i=1}^{n} \nabla_\omega Q(z_i, \omega_t) \tag{2}$$

Where γ denotes the model gain and ω denotes the model weights which have to be close to optimal.

The estimated value z is randomly selected given the weight w_t for each iteration t. This process helps to optimise the resulting value in each iteration provided that the selected values are based on the ground truth distribution. The convergence of the SGD model assumes a gradual decline in the value of model gain γ which must be neither too fast nor too slow. The optimal value for the convergence is achieved when γ_t is close to the value t^{-1} which causes a decline in the error rate at the same speed [51].

SGD has been successfully used in several image classification applications. These applications involved introducing some minor modifications to improve their performance by optimising their parameters and also by combining it with other machine learning techniques such as Artificial Neural Networks (ANN) [52], [53], [54].

Random Forests (RF): A non-parametric, powerful machine learning technique that is used for both regression and classification [55]. The random Forests model consists of a number of regression trees that are built concurrently and then voted on selecting the best-performing model [56].

Random Forests was reported successful for image classification in several applications, which covers a wide spectrum of image processing domains [57], [58], [55]. This makes Random forests a good candidate for classifying stone fruit diseases in this study.

3) Model Evaluation: Cross-validation is the most widely accepted technique for evaluating the performance of machine learning classifiers [59], [60]. In this method, samples are split into several equal-size folds, and then the validation process is carried out considering one fold for testing the classification model each time, while the rest are used for training the classification model. The model validation is then repeated several times, which is equal to the number of folds. At the end of the validation process, the performance of the model is calculated by averaging the model performance for each fold.

The confusion matrix is a common evaluation method that is used to measure the performance of a machine learning model across classes. The confusion matrix is drawn in a tabular form that shows the numbers or percentage of the classified samples as illustrated in Fig. 3. It shows four types of information: (A) True positive values, which refer are the number of samples that were predicted positive and are indeed positive; (B) False-positive, which refers to the number of samples that are predicted positively by the model but they are in fact negative; (C) False-negative which refers to the number of samples that are predicted negative, but they are in fact positive, and (D) True negative which refers to the number of samples that are predicted negative and they are indeed negative.

In addition to the Classification Accuracy (CA), four other performance measures were used for evaluating the classification model: (1) Area Under the Curve (AUC); (2) F1; and (4) Precision. The confusion matrix, Receiver Operating Characteristic (ROC), Lift, and Calibration curves were also used to confirm the validity and robustness of the models performance. The ROC curve maps the false positive rate to the true positive rate of the class prediction. This curve is widely accepted as an excellent and accurate metric of the machine learning model's performance [61], [62]. In the ROC curve, the

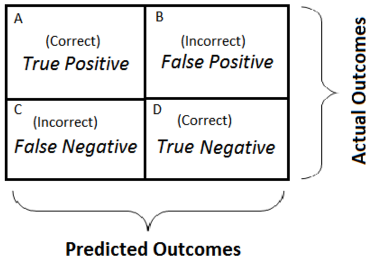

Fig. 3. An illustration of the confusion matrix. Rows show the predicted classes, while columns show the actual classes

further is a line drawn from the middle of the chart (diagonal), indicating a better performance achieved by the model. The ROC curve is also used as a basis for calculating the AUC performance metric, which is simply defined as the Area Under the ROC curve (AUC). The larger the area is, the better performance is achieved by the model [63]. The calibration curve measure the model's performance in predicting true positive values versus false positive values, while the Lift chart measure the model's performance in predicting positive values [64].

Classification Accuracy (CA) is one of the most commonly used performance metrics in machine learning [65]. It is used to evaluate the model's performance in predicting the learned classes by calculating the ratio between the number of the correctly classified samples compared to the total samples, which are expressed equation 3.

$$CA = \frac{(TP + TN)}{Total} \tag{3}$$

The precision metric is also another common measure that is used for evaluating the classification model's performance by calculating the ratio between the number of true positive values and the total number of both true positive and false positive values [65], which is expressed by equation 4.

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

III. Results

The results obtained in this empirical study involved creating and evaluating three machine learning models using KNN, SGD, and RF algorithms. Each of these models has been created based on the feature vectors that have been extracted using the pre-trained deep learning convolutional neural networks VGG16 algorithm and using images that have been captured for stone fruit leaves that are infected with bacterial and fungal diseases.

In the first phase, the features in each image were by submitting the captured images using Keras API to a remote VGG16 algorithm implementation that is pre-trained using millions of images from the ImageNet database. Fig. 4 shows a set of images that have been generated by the VGG16 algorithm, for example, a leaf using the algorithm convolutions, pooling, and normalisation procedures. The image is a sample of the images that are generated by each convolution in the model layers. The image was first converted into a grey-scale and was then passed to the VGG16 layers. The extracted features were stored in vectors that contain the most important descriptors for each image and saved on the local machine. The image features in the vectors were then loaded into the classifier models.

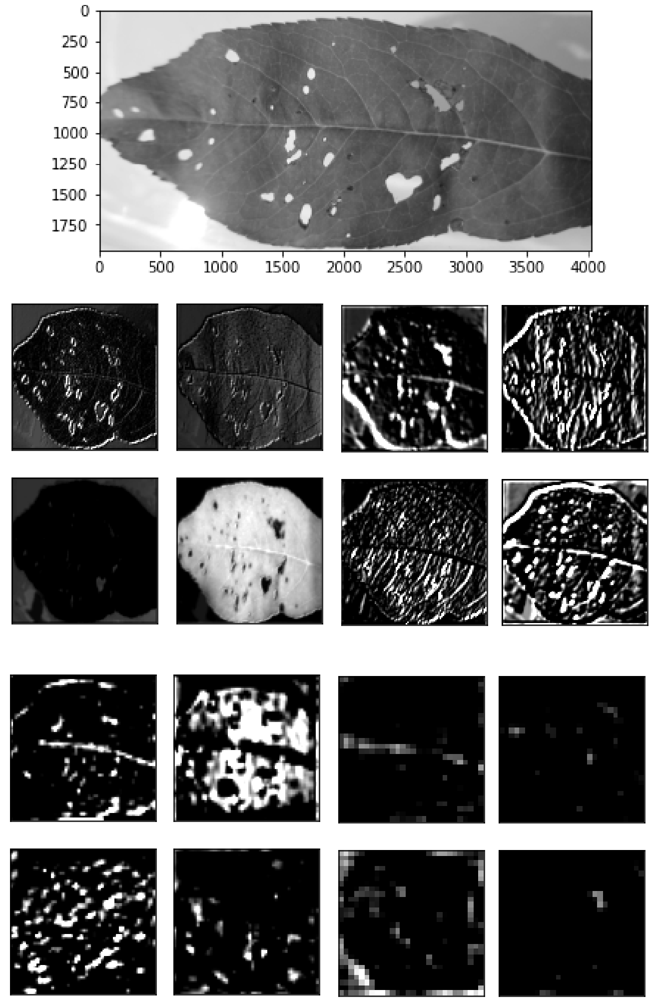

Fig. 4. Filtered image features of a leaf that suffers from fungal infection.

In the second phase, three classification models were constructed using the features extracted in the first phase. The first model was constructed using the KNN algorithm and then trained using the training dataset to predict the image classification into the two pre-labelled classes. The second model was constructed using the SGD algorithm, while the third was built using the RF algorithm.

In the third phase, the models constructed in the second phase were trained using the image dataset that was described earlier in Section II-A, which consisted of tens of images of leaves that suffer bacterial and fungal disease infections. In the fourth stage, the models were then evaluated using 10-fold cross-validation. The evaluation of each model involved measuring the performance of the classifier using: (1) Classification Accuracy; (2) Precision test; and (3) Area Under the Curve (AUC). In addition, a confusion matrix was constructed to validate the performance of each model that was visualised using the ROC, Lift and Calibration curves [66].

The confusion matrix for the three constructed models is shown in Table IV, II, and III. The values on the diagonal of the table –highlighted with dark shadow– represent the image samples that have been correctly classified in each class, while the values to the left and the right of the table diagonal highlighted with bright shadow–, represent those images that have been wrongly classified. The values in the rows represent the actual values that, in fact, belong to the corresponding class, while the values in the columns show the values that are predicted to belong to the corresponding class.

The evaluation of the classifier models has shown good performance based on the results of the confusion matrix, ROC, Lift and calibration curves, in addition to classification accuracy and sensitivity metrics. All the applied techniques have scored more than 87.4% in classification accuracy and precision performance metrics and more than 93.9% in the AUC metric, which is an even more efficient metric for evaluating the performance of the classifiers.

The CNN-KNN outperformed the other two models in the AUC metric. It scored 98.5% in AUC, while the CNN-SGD and CNN-RF models scored 95.4% and 93.9%, respectively. On the other hand, the CNN-SGD achieved the highest score in the classification accuracy metric, which was 93.6%. In contrast, the CNN-KNN and CNN-RF models scored classification accuracy of 92.4% and 87.4%, respectively. The results of the precision and F1 support the model's scores in classification accuracy. The CNN-SGD model scored 93.7% in F1 and 94.% in Precision metrics. The CNN-KNN model scored an F1 of 92.3% and a Precision of 92.3%, while the CNN-RF scored 85.9% and 88% in F1 and Precision metrics. In the AUC metric, the CNN-KNN model scored an even higher result, just over 95.6%, while the CNN-SGD scored 89.4%. The CNN-RF model scored 88.3%. Table I compares the performance of the three models in the applied metrics.

TABLE I. A COMPARISON OF THE PERFORMANCE OF THE THREE CREATED CLASSIFICATION MODELS: CNN-KNN, CNN-SGD AND CNN-RF BASED ON AUC, ACCURACY AND PRECISION METRICS

	AUC	**CA**	**F1**	**Precision**
CNN-KNN	98.5%	92.4%	92.3%	92.3%
CNN-SGD	95.4%	93.6%	93.7%	94.1%
CNN-RF	93.9%	87.4%	85.9%	88.0%

TABLE II. THE CONFUSION MATRIX FOR THE CNN-KNN MODEL

		Predicted			
		Bacterial	Fungal	Healthy	Total
Actual	Bacterial	**84**	22	0	106
	Fungal	15	**279**	1	295
	Healthy	0	0	**99**	99
	Total	99	301	100	500

The performance curves of the ROC charts also confirmed the accuracy and precision of the metric result. It showed that the CNN-KNN model by far outperformed both the SGD and RF models. These results were also apparent in predicting the fungal class and, to a lesser extent, in predicting the bacterial class. However, while the CNN-SGD model performed relatively well according to the ROC chart, the performance of the RF model lagged. Fig. 5 illustrate the performance of the three constructed models using the ROC chart. Furthermore, a calibration curve was created for all the constructed models, which compares them in classifying true and false positive classes. The threshold of the P value was set to 0.50, which is shown in Fig. 6. The SGD classifier scored a rate of 91.5% and 10% in predicting true positive and 8.8% in predicting false positive samples. The KNN classifier scored 79.2% and 3.8% in predicting the true positive and false positive samples, respectively, while the Random Forests model scored 39.6% and 10% in predicting the true positive and false positive samples, respectively. These results, yet again, show that both the SGD and KNN significantly outperformed the Random Forest model. However, the SGD model achieved the best performance, particularly in predicting true positive values. The Lift curve in Fig. 7 shows the performance of the three classifiers in predicting the positive values (Positive rate). Based on the Lift curve, the SGD model achieved the best performance, followed by the KNN and SGD models.

TABLE III. THE CONFUSION MATRIX FOR THE CNN-SGM MODEL

		Predicted			
		Bacterial	Fungal	Healthy	Total
Actual	Bacterial	**97**	8	1	106
	Fungal	23	**272**	0	295
	Healthy	0	0	**99**	99
	Total	120	280	100	500

TABLE IV. THE CONFUSION MATRIX FOR THE CNN-RF MODEL

		Predicted			
		Bacterial	Fungal	Healthy	Total
Actual	Bacterial	**49**	57	0	106
	Fungal	6	**289**	0	295
	Healthy	0	0	**99**	99
	Total	55	346	99	500

The champion two models were constructed by combining the pre-trained convolutional network algorithm VGG16 with the KNN and SGD algorithms. The CNN-KNN model scored 98.5% in the AUC metric, while the CNN-SGD scored a classification accuracy of 93.6%. However, the AUC metric is considered by many data mining practitioners as the best performance metric for evaluating machine learning models [63].

IV. DISCUSSION

This research involved collecting and photographing hundreds of leaves from stone fruit trees that are grown in the north of Jordan that suffer from bacterial and fungal diseases. The leaves were photographed and then classified based on their identified infections. The images were then fed into a deep learning convolutional network neural network algorithm that was used to identify the significant features involved in each of them. The resulting features were then used as input for three machine learning algorithms that have been used for classifying the leaves' images according to their infected diseases. The constructed models were then trained, and their performances were measured and evaluated based on a 10-fold cross validation.

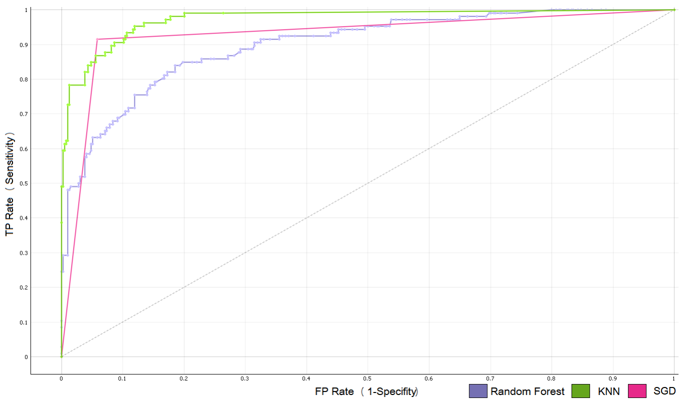

Fig. 5. An ROC chart that compares the performance of the three created classification models.

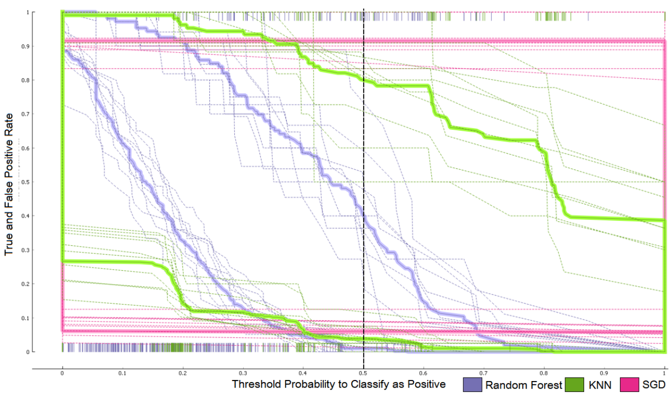

Fig. 6. A Calibration curve showing the performance of the three created classification models in predicting true positive and false positive classes

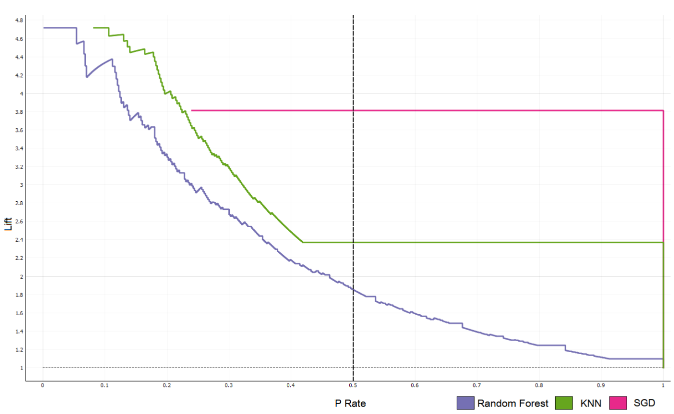

Fig. 7. A Lift curve showing the performance of the three created classification models in terms of the relationship between lift and the classification positive rate

The results confirmed the research hypothesis, which suggested that pre-trained deep learning algorithms can provide an efficient method for diagnosing stone fruit fungal and bacterial diseases by identifying image features and then using them in classifying the diseases' images using machine learning classification algorithms.

The results also provides a positive answer to the research question by creating classification models that were able to diagnose the stone fruit fungal and bacterial infections with an accuracy performance of 93.6% using the CNN-SGD model and an AUC performance of 98.5% using the CNN-KNN model.

The study was successful in achieving the preset research objectives, which aimed at creating a computational model that would help create an automated prediction model that can be embedded in a mobile phone application or a web service to provide a fast, cheap, and reliable diagnosis of fungal and bacterial infections through photographing the visual symptoms that on plant leaves using standard mobile cameras and under the field's normal conditions. The findings of the research confirm the validity of our proposed approach to achieve the research objectives through combining CNN algorithm for feature extraction with three classical machine learning classifiers: KNN, SGD and Random Forest.

The performance of the classification model created in this study was above the average performance of other models that have been reported in similar and previous studies, particularly when compared to other studies that involved collecting, photographing, and analysing hundreds of plant leaf images under normal field conditions [22], [23], [25], [26], [27]. However, other studies that used public datasets with controlled conditions and larger dataset achieved better results as reported in [29], [30], [31], [32], [35].

The CNN-KNN model outperformed both CNN-SGD and CNN-RF in the AUC metric, as it scored a performance of 98.5%, while they scored an AUC performance of 95.4% and 93.9%, respectively. On the other hand, the CNN-SGD model outperformed the other two models based on the classification accuracy metric as its scored performance of 93.6%. The CNN-KNN and CNN-RF models scored a classification accuracy of 92.8% and 87.4%, respectively. However, most of the other studies that reported better results used thousands of images from public benchmark datasets that were acquired under controlled environments and conditions. In contrast, our study did not suffer some of the problems that are related to the degradation of model performance when leaves were taken on a different background, as reported in [30].

The models created in the present study is quite rigourous, even when images were taken in different aspect ratios and at different orientations. In addition, most of the studies that we have referred to in the related work section lack consideration for the prospect of the acquired knowledge and for the practical deployment and realisation of the generated model, which was considered from the beginning and to the end of this study.

V. Conclusion

In this research, the power of deep learning in feature extraction and detection was combined with the classification power of three machine learning algorithms: KNN, SGD, and RF, for analysing leaf infections visually using image recognition and classification technology.

The study concerned conducting end-to-end empirical research, which involved collecting hundreds of stone fruit leaf samples from fields that suffer from bacterial and fungal infections. The samples were collected and then photographed from farms in various locations in the north of Jordan. The bacterial and fungal infections were diagnosed and identified by two experts in plant diseases and confirmed using literature and using the European and Mediterranean Plant Protection Organisation (EPPO) database repository, which covers 1,700 pests and 88,000 plant species.

The results confirm the applicability and validity of combining the deep learning CNN algorithm with other machine learning classification algorithms for identifying and diagnosing plant bacterial and fungal diseases based on their visual signs and symptoms. The CNN-KNN model outperformed both the CNN-SGD and CNN-RF models based on the AUC metric and ROC curve, as it scored a performance of 98.5% in AUC. On the other hand, the CNN-SGD model's performance was the best in the classification accuracy metric, scoring 93.6%. The CNN-KNN model scored 92.4%, while the CNN-RF model scored 87.4%. Nevertheless, the limitations of this study come from its specialisation in one crop rather than others, as well as in its limited coverage of bacterial and fungal infections rather than other diseases and deficiencies. Nevertheless, this limitation was noticed in most of the previous works, including those that used public and field datasets. Expanding the research to cover more plants and more diseases would provide a more comprehensive image of other crops and other diseases.

The future work of this research may involve conducting further research to cover a wider range of crops and diseases under the field's normal conditions and using thousands of images. It may also involve using more deep learning methods and machine learning algorithms to achieve better results. In addition, constructing a data repository for all the diagnosed diseases would also help in enhancing the accuracy of future models by providing more images for model training and testing.

Furthermore, the success of these applications can be utilised in deploying machine learning classification models that can be embedded in a mobile application that can provide fast, cheap, and reliable identification of plant diseases or can be integrated into a web-based service that can help farmers across the globe to diagnose their plant's diseases by photographing the leaf infections in their crops using their mobile phones and upload the images to the web service.

ACKNOWLEDGEMENT

We would like to thank Mr. Abu Hmiadan and other anonymous farmers who facilitated the collection of samples.

REFERENCES

[1] V. A. Gulhane and A. A. Gurjar, "Detection of diseases on cotton leaves and its possible diagnosis," *International Journal of Image Processing (IJIP)*, vol. 5, no. 5, pp. 590–598, 2011.

[2] J. Adaskaveg, J. Ogawa, and E. Butler, "Morphology and ontogeny of conidia in wilsonomyces carpophilus, gen. nov. and comb. nov., causal pathogen of shot hole disease of prunus species," *Mycotax*, vol. 37, pp. 275–290, 1990.

[3] C. Văcăroju, C. Ză1ă, and S. Cristea, "Research on the morphology and biology of the stigmina carpophila fungus," in *scientific conferences with international participation durable agriculture-Agriculture of future, 4th edition. Craiova (Romania): Faculty of Agriculture, University of Craiova*, vol. 110, 2008.

[4] J. M. Ogawa and H. English, *Diseases of temperate zone tree fruit and nut crops*. UCANR Publications, 1991, vol. 3345.

[5] A. Ahmadpour, Y. Ghosta, M. Javan-Nikkhah, K. Ghazanfari, and R. Fatahi, "Study on morphology, pathogenicity and genetic diversity of wilsonomyces carpophilus isolates, the causal agent of shot hole of stone fruit trees based on rapd-pcr in iran," *Archives of Phytopathology and Protection, Plant*, vol. 45, no. 17, pp. 2076–2086, 2012.

[6] H. Khlaif, I. Abu-Obeid, and B. Werikat, "Occurrence of plant bacterial diseases in jordan," *African Journal of Agricultural Research*, vol. 13, no. 40, pp. 2104–2117, 2018.

[7] L. Highberg and J. Ogawa, "Survival of shot-hole inoculum in association with dormant almond buds," *Journal of Plant Disease*, vol. 70, no. 9, pp. 828–831, 1986.

[8] L. Gianessi and N. Reigner, "The value of fungicides in us crop production. croplife foundation," *Journal of Crop Protection Research Institute*, 2005.

[9] D. Ritchie, "Sprays for control of bacterial spot of peach cultivars having different levels of disease susceptibility," *Fungicide Nematicide Tests*, vol. 54, p. 63, 1999.

[10] E. Alpaydin, *Machine learning*. MIT Press, 2021.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[12] G. Hinton, "Deep learning a technology with the potential to transform health care," *Jama*, vol. 320, no. 11, pp. 1101–1102, 2018.

[13] A. BaniMustafa, "Predicting software effort estimation using machine learning techniques," in *2018 8th International Conference on Computer Science and Information Technology (CSIT)*. IEEE, 2018, pp. 249–256.

[14] M. Pak and S. Kim, "A review of deep learning in image recognition," in *2017 4th international conference on computer applications and information processing technology (CAIPT)*. IEEE, 2017, pp. 1–3.

[15] R. Gajjar, N. Gajjar, V. J. Thakor, N. P. Patel, and S. Ruparelia, "Real-time detection and identification of plant leaf diseases using convolutional neural networks on an embedded platform," *The Visual Computer*, vol. 38, no. 8, pp. 2923–2938, 2022.

[16] M. M. Kabir, A. Q. Ohi, and M. F. Mridha, "A multi-plant disease diagnosis method using convolutional neural network," in *Computer Vision and Machine Learning in Agriculture*. Springer, 2021, pp. 99–111.

[17] A. BaniMustafa, "Enhancing learning from imbalanced classes via data preprocessing a data-driven application in metabolomics data mining," *The ISC International Journal of Information Security*, vol. 11, no. 3, pp. 79–89, 2019.

[18] A. BaniMustafa and N. Hardy, "Applications of a novel knowledge discovery and data mining process model for metabolomics," *arXiv preprint arXiv:1907.03755*, 2019.

[19] A. BaniMustafa, M. Baklizi, and K. Khatatneh, "Machine learning for securing traffic in computer networks," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 13, no. 12, pp. 426–435, 2022.

[20] M. H. Sheikh, T. T. Mim, M. S. Reza, A. S. A. Rabby, and S. A. Hossain, "Detection of maize and peach leaf diseases using image processing," in *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2019, pp. 1–7.

[21] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and electronics in agriculture*, vol. 147, pp. 70–90, 2018.

[22] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59 069–59 080, 2019.

[23] X. Xie, Y. Ma, B. Liu, J. He, S. Li, and H. Wang, "A deep-learning-based real-time detector for grape leaf diseases using improved convolutional neural networks," *Frontiers in plant science*, vol. 11, 2020.

[24] E. Fujita, Y. Kawasaki, H. Uga, S. Kagiwada, and H. Iyatomi, "Basic investigation on a robust and practical plant diagnostic system," in *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2016, pp. 989–992.

[25] D. Oppenheim and G. Shani, "Potato disease classification using convolution neural networks," *Advances in Animal Biosciences*, vol. 8, no. 2, p. 244, 2017.

[26] J. Chen, Q. Liu, and L. Gao, "Visual tea leaf disease recognition using a convolutional neural network model," *Symmetry*, vol. 11, no. 3, p. 343, 2019.

[27] J. Amara, B. Bouaziz, and A. Algergawy, "A deep learning-based approach for banana leaf diseases classification," in *Datenbanksysteme for Business, Technologie und Web (BTW 2017) - Workshopband*, B. Mitschang, D. Nicklas, F. Leymann, H. Schning, M. Herschel, J. Teubner, T. Hórder, O. Kopp, and M. Wieland, Eds. Bonn: Gesellschaft fúor Informatik e.V., 2017, pp. 79–88.

[28] K. Yamamoto, T. Togami, and N. Yamaguchi, "Super-resolution of plant disease images for the acceleration of image-based phenotyping and vigor diagnosis in agriculture," *Sensors*, vol. 17, no. 11, p. 2557, 2017.

[29] S. Wallelign, M. Polceanu, and C. Buche, "Soybean plant disease identification using convolutional neural network," in *The thirty-first international flairs conference*, 2018.

[30] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in plant science*, vol. 7, p. 1419, 2016.

[31] P. V. Raja, K. Sangeetha, N. B A, S. M, and S. S S, "Convolutional neural networks based classification and detection of plant disease," in *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)*, 2022, pp. 1484–1488.

[32] M. Pei, M. Kong, M. Fu, X. Zhou, Z. Li, and J. Xu, "Application research of plant leaf pests and diseases base on unsupervised learning," in *2022 3rd International Conference on Computer Vision, Image and Deep Learning and International Conference on Computer Engineering and Applications (CVIDL and ICCEA)*, 2022, pp. 1–4.

[33] H. Wang, S. Shang, D. Wang, X. He, K. Feng, and H. Zhu, "Plant disease detection and classification method based on the optimized lightweight yolov5 model," *Agriculture*, vol. 12, no. 7, p. 931, 2022.

[34] Z. Liu, R. N. Bashir, S. Iqbal, M. M. A. Shahid, M. Tausif, and Q. Umer, "Internet of things (iot) and machine learning model of plant disease prediction–blister blight for tea plant," *IEEE Access*, vol. 10, pp. 44 934–44 944, 2022.

[35] P. Praveen, M. Nischitha, C. Supriya, M. Yogitha, and A. Suryanandh, "To detect plant disease identification on leaf using machine learning algorithms," in *Intelligent System Design*. Springer, 2023, pp. 239–249.

[36] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 2009, pp. 248–255.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[38] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge, 2014," *arXiv preprint arXiv:1409.0575*, 2014.

[39] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[40] A. Gulli and S. Pal, *Deep learning with Keras*. Packt Publishing, 2017.

[41] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.

[43] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[44] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[45] A. BaniMustafa and N. Hardy, "A strategy for selecting data mining techniques in metabolomics," in *Plant Metabolomics: Methods and Protocols*, ser. Methods in Molecular Biology, N. Hardy and R. Hall, Eds. Springer Science, 2012, vol. 860, ch. 18, pp. 317–333.

[46] A. Banimustafa and N. Hardy, "A scientific knowledge discovery and data mining process model for metabolomics," *IEEE Access*, vol. 8, pp. 209 964–210 005, 2020.

[47] A. M. Bani Mustafa, "A knowledge discovery and data mining process model for metabolomics," Ph.D. dissertation, Aberystwyth University, 2012.

[48] E. Fix, *Discriminatory analysis: nonparametric discrimination, consistency properties*. USAF School of Aviation Medicine, 1951.

[49] R. Ramteke and K. Y. Monali, "Automatic medical image classification and abnormality detection using k-nearest neighbour," *International Journal of Advanced Computer Research*, vol. 2, no. 4, p. 190, 2012.

[50] J. Kim, B.-S. Kim, and S. Savarese, "Comparing image classification methods: K-nearest-neighbor and support-vector-machines," in *Proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics*, 2012, pp. 133–138.

[51] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT'2010*. Springer, 2010, pp. 177–186.

[52] A. El Mouatasim, "Fast gradient descent algorithm for image classification with neural networks," *SIGNAL IMAGE AND VIDEO PROCESSING*, 2020.

[53] B. Zhang, C. Chen, Q. Ye, J. Liu, D. Doermann *et al.*, "Calibrated stochastic gradient descent for convolutional neural networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 9348–9355.

[54] Z. Yang, C. Wang, Z. Zhang, and J. Li, "Accelerated stochastic gradient descent with step size selection rules," *Signal Processing*, vol. 159, pp. 171–186, 2019.

[55] N. Horning *et al.*, "Random forests: An algorithm for image classification and generation of continuous fields data sets," in *Proceedings of the International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences, Osaka, Japan*, vol. 911, 2010.

[56] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[57] B. Xu, Y. Ye, and L. Nie, "An improved random forest classifier for image classification," in *2012 IEEE International Conference on Information and Automation*. IEEE, 2012, pp. 795–800.

[58] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *2007 IEEE 11th international conference on computer vision*. Ieee, 2007, pp. 1–8.

[59] M. W. Browne, "Cross-validation methods," *Journal of mathematical psychology*, vol. 44, no. 1, pp. 108–132, 2000.

[60] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14. Montreal, Canada, 1995, pp. 1137–1145.

[61] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (roc) curve." *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.

[62] J. Fan, S. Upadhye, and A. Worster, "Understanding receiver operating characteristic (roc) curves," *Canadian Journal of Emergency Medicine*, vol. 8, no. 1, pp. 19–20, 2006.

[63] C. X. Ling, J. Huang, H. Zhang *et al.*, "Auc: a statistically consistent and more discriminating measure than accuracy," in *Ijcai*, vol. 3, 2003, pp. 519–524.

[64] M. Vuk and T. Curk, "Roc curve, lift chart and calibration plot," *Advances in methodology and Statistics*, vol. 3, no. 1, pp. 89–108, 2006.

[65] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.

[66] J. D. Novaković, A. Veljović, S. S. Ilić, Ž. Papić, and M. Tomović, "Evaluation of classification models in machine learning," *Theory and Applications of Mathematics & Computer Science*, vol. 7, no. 1, pp. 39–46, 2017.