Journal of Agricultural Sciences (Tarim Bilimleri Dergisi) 2024, 30 (3) : 464 – 476 DOI: 10.15832/ankutbd.1288298

Journal of Agricultural Sciences
(Tarim Bilimleri Dergisi)

J Agr Sci-Tarim Bili
e-ISSN: 2148-9297
jas.ankara.edu.tr

JAS

# Machine Learning-based for Automatic Detection of Corn-Plant Diseases Using Image Processing

**Khaled Adil Dawood Idress[a*], Omsalma Alsadig Adam Gadalla[b], Yeşim Benal Öztekin[a], Geofrey Prudence Baitu[c]**

[a]*Ondokuz Mayis University, Faculty of Agriculture, Department of Agricultural Machinery and Technologies Engineering, Samsun, TÜRKİYE*
[b]*University of Khartoum, Faculty of Agriculture, Department of Agricultural Engineering, Khartoum, SUDAN*
[c]*University of Dar es Salaam, College of Agriculture and Food Technology, Department of Agricultural Engineering, Dar es Salaam, TANZANIA*

ARTICLE INFO
Research Article
Corresponding Author: Khaled Adil Dawood Idress, E-mail: adilkhaled850@gmail.com
Received: 26 April 2023 / Revised: 06 January 2024 / Accepted: 15 January 2024 / Online: 23 July 2024

Cite this article
Idress K A D, Gadalla O A A, Öztekin Y B, Baitu G P (2024). Machine Learning-based for Automatic Detection of Corn-Plant Diseases Using Image Processing. *Journal of Agricultural Sciences (Tarim Bilimleri Dergisi),* 30(3):464-476. DOI: 10.15832/ankutbd.1288298

ABSTRACT
Corn is one of the major crops in Sudan. Disease outbreaks can significantly reduce maize production, causing huge damage. Conventionally, disease diagnosis is made through visual inspection of the damage in fields or through laboratory tests conducted by experts on the affected plant parts of the crop. This process typically requires highly skilled personnel, and it can be time-consuming to complete the necessary tasks. Machine learning methods can be implemented to rapidly and accurately detect disease and reduce the risk of crop failure due to disease outbreaks. This study aimed to use traditional machine learning techniques to detect maize diseases using image processing techniques. A total of 600 images were obtained from the open-source Plant Village dataset for experimentation. In this study, image segmentation was done using K-means clustering, and a total of 4 GLCM texture features and two statistical features were extracted from the images. In this study, four traditional machine learning algorithms were applied to detect diseased maize leaves (common rust and gray leaf spot) and healthy maize leaves. The results showed that all the algorithms performed well in identifying the diseased and healthy leaves, with accuracy rates ranging from 90% to 92.7%. The highest accuracy scores were obtained with support vector machine and artificial neural networks, respectively.

Keywords: Maize Disease, Traditional Machine Learning, Image Processing, Feature Extraction

## 1. Introduction

Maize (Zea mays L.) is one of the most widely cultivated and consumed cereal crops worldwide, after wheat and rice, and is recognized as the "queen of cereals". It holds significant economic importance for resource-limited farmers in developing countries. Maize is not only utilized as a primary food source but also for industrial purposes, including biofuel production, starch, and oil extraction. Unfortunately, like other crops, plant diseases are a significant challenge that farmers face worldwide. Maize diseases occur yearly and significantly impede maize production (Subramanian et al. 2022). The diseases affecting maize crops have the potential to cause varying degrees of harm, from moderate to severe, to the overall production of the crop. Reports indicate that the annual damage caused by pathogenic diseases alone ranges from 4% to 14% of total maize production (Oerke & Dehne 2004). During its growth, maize leaves are exposed to several disease risks such as grey leaf spot, blight, and common rust. Early detection and management of these diseases are essential for maintaining the health of the maize crop and ensuring optimum yield. In order to manage diseases effectively, it is crucial to accurately diagnose and identify them before implementing any control measures. Conventionally, disease diagnosis is made through visual inspection of the damage in fields or through laboratory tests conducted by experts on the affected plant parts of the crop (Donatelli et al. 2017). Despite their effectiveness, the traditional approaches to disease diagnosis have some inherent limitations, which can make them impractical in some situations. These methods typically require highly skilled personnel and it can be time-consuming to complete the necessary tasks.

The most promising strategy for overcoming the limitations of traditional disease diagnosis methods and increasing production seems to be the development of automated systems based on computer vision (Subramanian et al. 2022). The progress made in artificial intelligence (AI) techniques has led to the development of highly advanced computerized systems that can accurately identify diseases (Jiang et al. 2017). Two classification techniques, machine learning (ML) and deep learning (DL), are used for categorizing crop leaf diseases. The key distinction between traditional machine learning and deep learning lies in how features are extracted. In traditional ML, features aren't automatically extracted, while in DL, feature extraction happens automatically and is regarded as learning weights. Hence, in DL, the system learns the required features autonomously by being exposed to adequate data (Vasavi et al. 2022). The use of machine learning-based AI applications has experienced significant growth in recent times, aided by the development of computational systems, particularly processors with embedded graphical processing units (GPUs). Leaf diseases can be precisely identified and categorized by utilizing digital image processing approaches, support vector machine (SVM), neural networks, and other techniques (Zhang et al. 2018). Panigrahi et al. (2020) employed supervised machine learning methodologies such as Naive Bayes (NB), Decision Tree (DT), K-Nearest Neighbor (KNN), Support Vector Machine (SVM, and Random Forest (RF) to detect maize plant diseases, encompassing northern leaf blight, Common rust, and Cercospora leaf spot. The aforesaid classification techniques were analyzed and compared in order to select the most suitable model with the highest accuracy for plant disease prediction. Classification accuracies of 77.56%, 77.46%, 76.16%, 74.35%, and 79.23% were achieved using SVM, NB, KNN, DT, and RF, respectively. Padol & Yadav (2016) used a Linear Support Vector Machine classification approach to identify various types of diseases in grape leaves. They applied pre-processing techniques to the grape leaf images and employed K-means clustering to detect the affected regions. They extracted color and texture features from these regions. SVM yielded better results, with an accuracy of 93.33%, compared to an accuracy of 83.33% achieved with powdery class grape leaf images. Singh et al. (2015) introduced an automatic leaf disease detection method using genetic algorithms. The proposed approach involves pre-processing the input image using image processing techniques and segmenting it using genetic algorithms for disease classification. To enhance the recognition rate in the classification process, the authors suggest utilizing Artificial Neural Networks (ANN), fuzzy logic, and hybrid algorithms. They proposed that these techniques will be effective in improving the accuracy of the classification process. Linear SVM, medium tree, quadratic SVM, and cubic SVM were employed to identify diseases in maize leaves. The model was initially trained using labeled image data, and the classification model with the highest accuracy rate was chosen to test new image data for disease detection. After testing, quadratic SVM performed best in detecting maize disease, with an accuracy rate of 83.3% (Chokey & Jain 2019). Mukhopadhyay et al. (2021) demonstrated a method for classifying diseases on tea plant leaves. Their approach involved using a non-dominated sorting genetic algorithm (NSGA-II) based on image segmentation to identify the infected regions on tea leaves. Principal Component Analysis (PCA) and multi-class (SVM) were then applied to classify and detect disease in the tea leaves, respectively. The results showed that this algorithm was able to accurately identify the type of disease present in tea leaves, with an average accuracy of 83%. Ramakrishnan & Sahaya (2015) created a method for classifying diseases on groundnut leaves by utilizing the backpropagation algorithm. Mokhtar et al. (2015) utilized gray-level co-occurrence matrix (GLCM) texture features to differentiate between healthy and contaminated tomato leaves and classify their respective status. They applied the SVM algorithm along with established kernel functions for the classification process. Their dataset encompassed 800 samples, consisting of both healthy and diseased tomato leaves. To assess the effectiveness of their approach, they employed N-fold cross-validation, achieving impressive classification accuracy of 99.83%. Ramesh & Vydeki (2019) proposed a machine learning-based classification method employing KNN and ANN. They extracted statistical features such as mean and standard deviation and GLCM texture features to detect blast disease in Indian rice. Their experimentation led to 63% and 88% test accuracies for normal images using the KNN and ANN algorithms, respectively. Moreover, for blast-affected images, test accuracies were 79% with KNN and 90% with ANN algorithms, respectively. In another study SVM and a convolutional neural network (CNN) were used to detect and categorize various plant diseases. Data was sourced from the "new plant diseases dataset" on Kaggle, which includes a collection of more than 12,949 images displaying both healthy and unhealthy crop leaves. Initially, the diseased segment of the leaf was isolated from the image, and diverse features were extracted using a GLCM. This segmented portion is then identified using SVM, achieving an accuracy of 80%. In order to enhance accuracy further, they employed CNN for the identification of plant diseases, which notably improved performance, achieving an accuracy rate of 97.71% (Fulari et al. 2020).

In the literature, many studies have used deep-learning models to detect maize diseases. Waheed et al. (2020) introduced an optimized DenseNet architecture and conducted training on several models, including VGGNet, XceptionNet, EfficientNet, and NASNet, to identify and classify maize leaf diseases. Their study demonstrated that the proposed CNN architecture possessed fewer parameters, making it computationally efficient compared to other architectures. The accuracy achieved by the model proposed in their work stands at an impressive 98.06%. Subramanian et al. (2022) utilized pre-trained models, including VGG16, ResNet50, InceptionV3, and Xception, to classify three common maize leaf diseases using a dataset comprising 18 888 images of healthy and diseased leaves. Their methodology involved employing Bayesian optimization for hyperparameter tuning and implementing image augmentation techniques to enhance the models' ability to generalize. The study conducted a comparative analysis of the proposed models. The results revealed that all the trained models showcased an accuracy exceeding 93% in accurately classifying maize leaf diseases. Notably, VGG16, InceptionV3, and Xception stood out by achieving exceptional accuracy rates, surpassing 99% in their classification performance for these diseases. (Yang et al. 2023) proposed an adversarial training collaborating multi-path context feature aggregation network for maize disease density prediction. Their study focused on utilizing adversarial training to address issues related to network overfitting, thereby enhancing the robustness and generalization capabilities of the model. They conducted both quantitative analysis and interpretability analysis using the Plant Village dataset. The outcomes of their research demonstrated exceptional results with high-quality performance metrics. Specifically, their model achieved impressive recognition accuracies of 98.4%, 99.60%, 98.62%, and 99.80% for classifying leaf blight, gray leaf, healthy leaf, and leaf rust images, respectively. This signifies the effectiveness and accuracy of their approach in predicting maize disease densities across various categories within the dataset. Xu et al. (2021) introduced a multiscale convolutional global pooling neural network to enhance the accuracy of identifying maize diseases. By employing transfer learning techniques, they addressed the challenge of limited sample data leading to overfitting. The experimental results revealed notable classification accuracies using different models: 86.17% for VGGNet-16, 88.8% for DenseNet, 90.48% for ResNet, 90.86% for AlexNet, and 93.28% for TCI-ALEXN. Their findings suggest that the proposed model exhibited superior performance compared to conventional convolutional neural network models like VGGNet-16, DenseNet, ResNet-50, and AlexNet in terms of recognition accuracy, signifying its efficacy in accurately identifying maize diseases. A new method was introduced to predict the severity of maize common rust disease using CNN deep learning models. The approach involved using threshold-segmentation on images of infected maize leaves to calculate the percentage of affected leaf area. This information was used to create fuzzy decision rules for categorizing the disease into four severity classes. A VGG-16 network was trained using these severity classes to automatically classify test images of common rust disease. The VGG-16 network achieved high validation accuracy (95.63%) and testing accuracy (89%) when tested on images categorized into early stage, middle stage, late stage, and healthy based on the proposed approach (Sibiya & Sumbwanyambe 2021). A proposed novel 15-layer deep convolutional neural network model was developed to distinguish images of three maize crop diseases - grey leaf spot (GLS), common rust (CR), and northern corn leaf blight (NCLB) - in addition to healthy images. The researchers utilized the maize dataset available in the Plant Village data repository for training, validation, and testing of their model. Their model was trained using 90% (1,376 images) of the dataset and then evaluated with the remaining 10% of the data. The experimental outcomes demonstrated strong performance in categorizing previously unseen maize images. The proposed model achieved an impressive overall classification accuracy of 99.10%, along with an F1-score of 97.49% on the testing set of the maize dataset (Haque et al. 2023). CNN was utilized for the automated detection of maize leaf diseases, trained on the Plant Village maize crop dataset. Various preprocessing techniques were applied to the dataset, and the efficiency of different model configurations was assessed based on metrics like accuracy, recall, precision, and F1 score. The proposed CNN model achieved a high accuracy of 96.76% by employing contrast limiting adaptive histogram equalization (CLAHE) on each RGB channel, subsequently converting the enhanced image into hue, saturation, value (HSV) color space (Jasrotia et al. 2023). From previous studies, we have noticed that using deep learning models to detect healthy and diseased maize leaves achieved very high accuracy, ranging from 79.74% to 99.84%. However, because of requiring a large amount of data, instead of just focusing on one approach, we focused on improving the accuracy of existing traditional machine learning methods to classify diseased and healthy maize leaves.

Plant diseases, such as grey leaf spot and common rust, can cause significant damage to corn crops, resulting in reduced yield and economic losses for farmers. Traditional methods for disease detection require trained professionals and can be time-consuming and expensive. In contrast, automated disease detection using machine learning-based techniques can provide a more efficient and cost-effective solution to this problem. Previous studies have used various image-processing techniques to detect plant diseases, including color, shape, and texture feature extraction, and classification algorithms. However, the proposed study differs from previous research in several ways. Firstly, statistical features such as mean and standard deviation were used, in addition to GLCM texture features, to improve the accuracy of the detection system. Secondly, a gap was noted in prior research; only two studies concentrated on utilizing GLCM texture features and statistical features for distinguishing between tomato and Indian rice leaves, with no studies conducted about classifying maize leaves.

The proposed study is needed to address the limitations of previous research and provide a more robust and accurate solution for automated plant disease detection. The study has the potential to benefit farmers and agricultural industries by enabling early detection and prevention of plant diseases, which can help to reduce crop losses and increase food security.

This study aims to evaluate the performance of various machine learning algorithms, including SVM, DT, RF, and ANN in detecting diseased maize leaves (common rust and gray leaf spot), as well as healthy maize leaves. MATLAB was employed to segment the images, extract the features, and perform the classification.

## 2. Material and Methods

### *2.1 Image acquisition*

Image detection research requires a suitable dataset at every step, from training to evaluating the algorithms. A total of 600 images were obtained from the open-source Plant Village dataset (https://github.com/spMohanty/PlantVillage-Dataset) for experimentation. The images were divided into three categories, with two being diseased and one being healthy. The categories included 200 images each of healthy maize leaves, grey leaf spot, and common rust. Example images of disease symptoms for each class are presented in Figures 1, 2, and 3.

**Figure1- Samples in healthy condition**

**Figure 2- Samples of leaves with common rust**

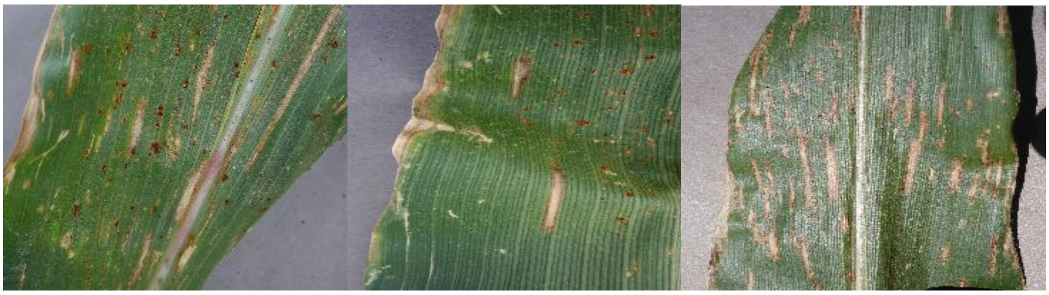


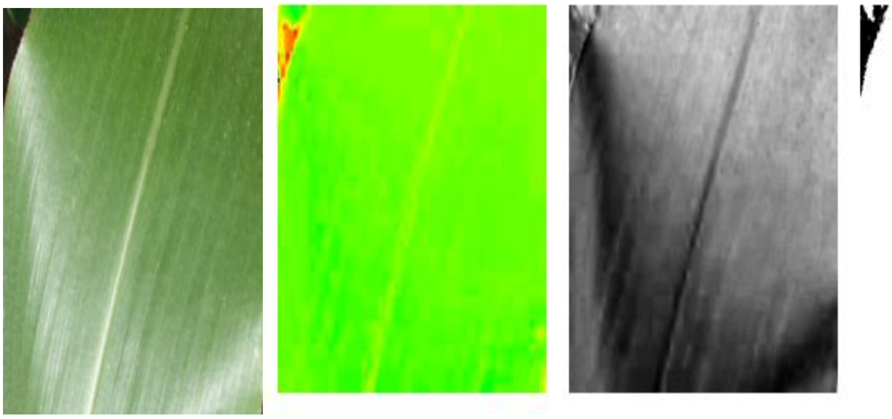
**Figure 3- Samples of leaves with gray leaf spot**

*2.2 Image processing*

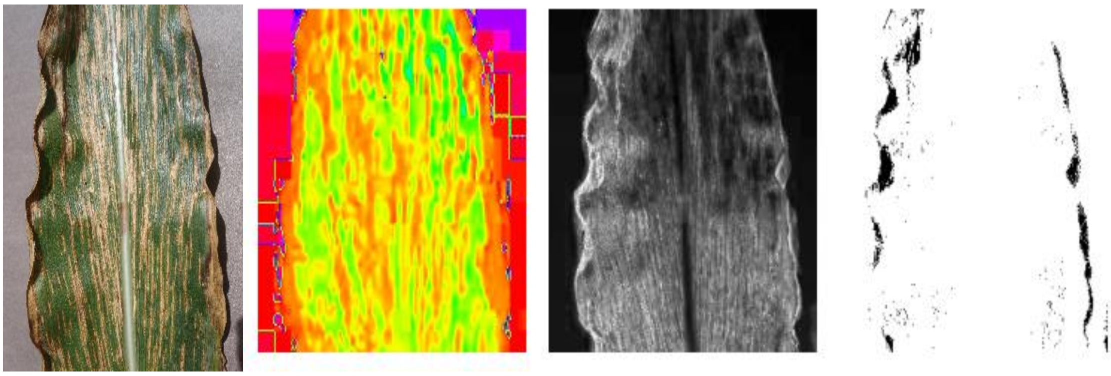
The color of the leaves can indicate the overall health of a maize plant. Diseased plants often have leaves that turn yellow or brown or have spots or decaying areas. The goal is to gather this information from image data. In feature extraction, the aim is to obtain features that differentiate well in the feature space, have a low number of dimensions, and are resilient to data variability. Red-green-blue (RGB) features extract color information and are widely used for image processing and recognition. RGB is particularly useful for object detection in images that display significant color changes. For object detection in an image with significant color changes, it is recommended to use RGB as the color format. The values of RGB range from 1 to 255.

*2.3 Image segmentation*

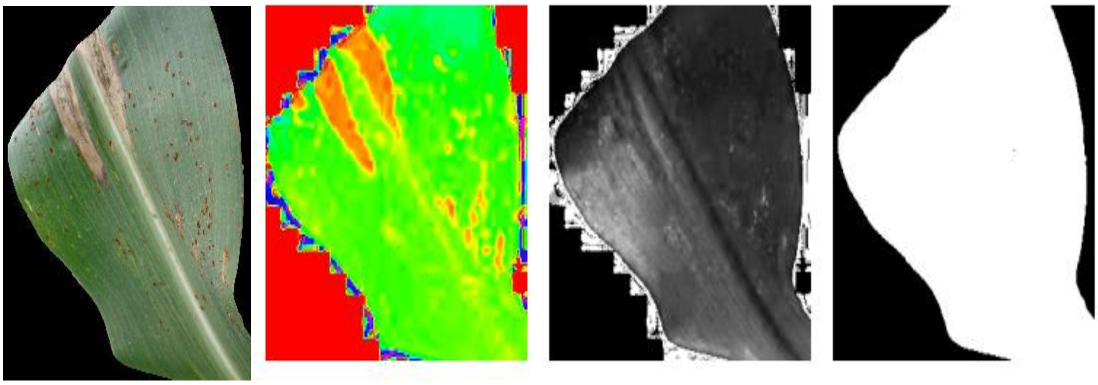
K-means clustering, which is a computer vision approach for dividing an image into multiple segments or clusters based on the distribution of color and intensity among its pixels, was used in this study for image segmentation. Following the pre-processing of images, different K values were used for image segmentation with K-means clustering. According to experimental findings, K=3 produced a clearer and more precise image (Dhingra et al. 2019). Figure (a-d) shows the original images (healthy and infected images), hue, saturation, and K-means clustered images. In this study, the image was converted from RGB to gray-level color space. Then, the images were resized to 400 pixels, and finally, the images were clustered using K-means clustering.



(a) Healthy image (b) Hue (c) Saturation ( d) Clustered image



(a) Gray leaf spot (b) Hue (c) Saturation (d) Clustered image



Common rust image (b)Hue (c) Saturation (d) Clustered image

*2.4 Feature extraction*

In this study, disease infection was detected using segmented images to extract various features. Statistical features such as mean and standard deviation were extracted along with GLCM texture features. The gray level co-occurrence matrix was utilized for the GLCM texture feature to identify the region affected by the disease. The study established the relationships between the features extracted with the GLCM algorithm for both the affected and normal regions. Furthermore, the average color value of the image was computed using the following equation.

1. Mean

Mean = Ei = $\sum_{j=1}^{N} pij$

Where; Ei, represents the mean of the dataset: N is the number of values in the dataset and $\sum$, denotes the summation of all values in the dataset: P(i,j), denotes the probability of co-occurrence of pixel values i and j at a particular offset within the GLCM.

2. Standard Deviation

$$S.D. = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(Pij - Ei)^2}$$

Where; S.D, represents the standard deviation: N is the number of values in the dataset: $\sum$, denotes the summation of squared differences between each value Pij and (Ei) is the mean.

3. Energy

$$Energy = \sum_{(i,j=0)}^{(N-1)}(Pij)^2$$

Where; P(i,j), denotes the probability of co-occurrence of pixel values i and j at a particular offset within the GLCM. Energy represents the sum of squared elements in the GLCM.

4. Homogeneity

$$Homogeneity = \sum_{i,j=0}^{N-1}\frac{Pij}{1+(i-j)^2}$$

Where; |i−j|, represents the absolute difference between i and j.

5. Contrast

Contrast = $\sum_{i,j=0}^{N-1} Pij(i-j)^2$

Where; P(i,j), represents the probability of co-occurrence of pixel values i and j in the GLCM, and i and j are the gray levels.

6. Correlation

Correlation = $\sum_{i,j=0}^{N-1}\frac{Pij\ (i-\mu)(i-\mu)}{\sigma^2}$

Where; P(i,j), represents the probability of co-occurrence of pixel values i and j in GLCM: μ, denotes is the mean of GLCM, and σ is the standard deviation of GLCM.

*2.5 Performance evaluation*

The confusion matrix and area under the receiver operating characteristic (ROC) curve are two commonly used metrics for evaluating the performance of binary classification models. The confusion matrix provides detailed information about the model's predicted and actual classes, while the ROC curve provides a visual representation of the model's trade-off between sensitivity and specificity at different classification thresholds.

The confusion matrix consists of four values: true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). These values can be used to calculate various performance metrics, such as accuracy, precision, recall, and F1-score. The confusion matrix is beneficial for identifying which types of errors the model is making.

The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at different classification thresholds. The TPR is the ratio of correctly predicted positive examples to the total number of positive examples, while the FPR is the ratio of incorrectly predicted positive examples to the total number of negative examples. The area under the curve (AUC) is a scalar value that summarizes the performance of a model across all possible classification thresholds. AUC represents the probability that a randomly chosen positive example will be ranked higher than a randomly chosen negative example by the model. A perfect model would have an AUC of 1.0, while a completely random model would have an AUC of 0.5. The AUC metric has several advantages over other classification performance metrics, such as accuracy, precision, and recall. One advantage is that it is insensitive to class imbalance, which is a common problem in many real-world classification problems. Additionally, the AUC metric provides a single scalar value that summarizes the overall performance of a model, making it easy to compare different models or different settings of a single model (Géron 2019).

Using both the confusion matrix and ROC curve together provides a comprehensive evaluation of the model's performance. The confusion matrix provides detailed information on the model's errors, while the ROC curve provides a visual representation of the model's performance at different classification thresholds. AUC provides a single scalar value that summarizes the overall performance of the model.

## 3. Results and Discussion

The study collected a dataset of 600 images from the Plant Village dataset, which were categorized into three classes: healthy maize leaves (200 images), grey leaf spot (200 images), and common rust (200 images). K-means clustering was employed for image segmentation, and statistical features such as mean and standard deviation, along with GLCM texture features, were extracted after the segmentation process. The dataset was split into 75% for training and 25% for testing. Four conventional ML algorithms, DT, SVM, KNN, and ANN models, were employed to train and test the images. The algorithms performed well in classifying the diseased and healthy maize leaves, with accuracy rates ranging from 90% to 92.7%. SVM and ANN achieved the highest accuracy scores among the tested algorithms.

### *3.1 Decision tree classifier*

A DT classifier is a type of machine learning algorithm that can be used for classification problems like classifying diseased and healthy maize leaves. The algorithm works by dividing the data into smaller subsets based on certain attributes and making decisions about the class label of each record based on the values of these attributes. The DT classifier was able to achieve an accuracy of 90% in the test dataset. Using decision tree models to classify common rust, gray leaf spot and healthy maize leaves can be a highly effective approach in the field of plant pathology. The high accuracy of 90% achieved in this study suggests that the decision tree model was able to distinguish between the three different categories of maize leaves accurately. The confusion matrix and ROC curve for the DT classifier are shown in Figures 3 and 4.

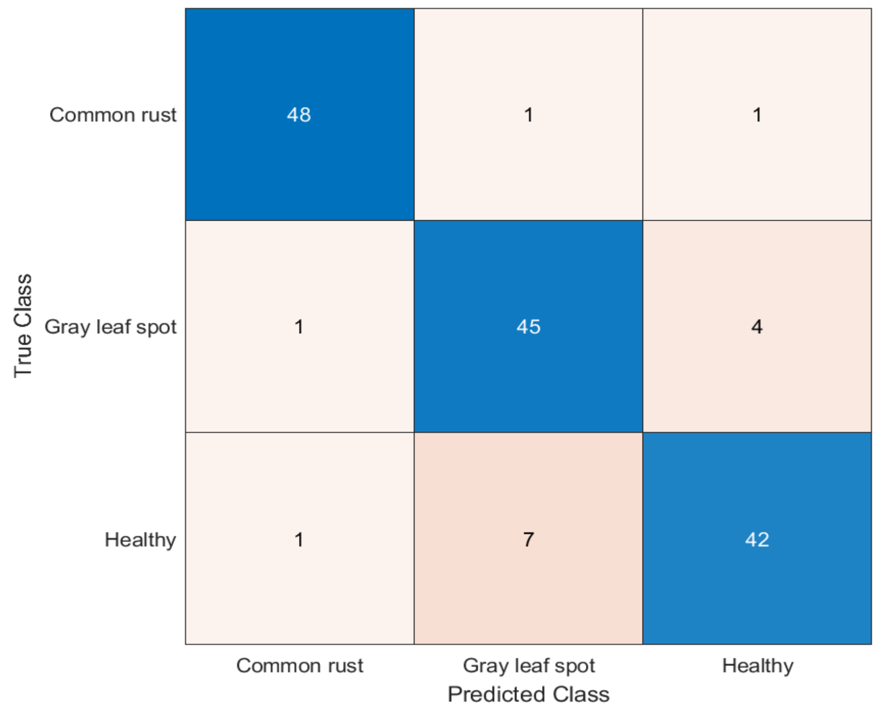



**Figure 3- Decision Tree confusion matrix**

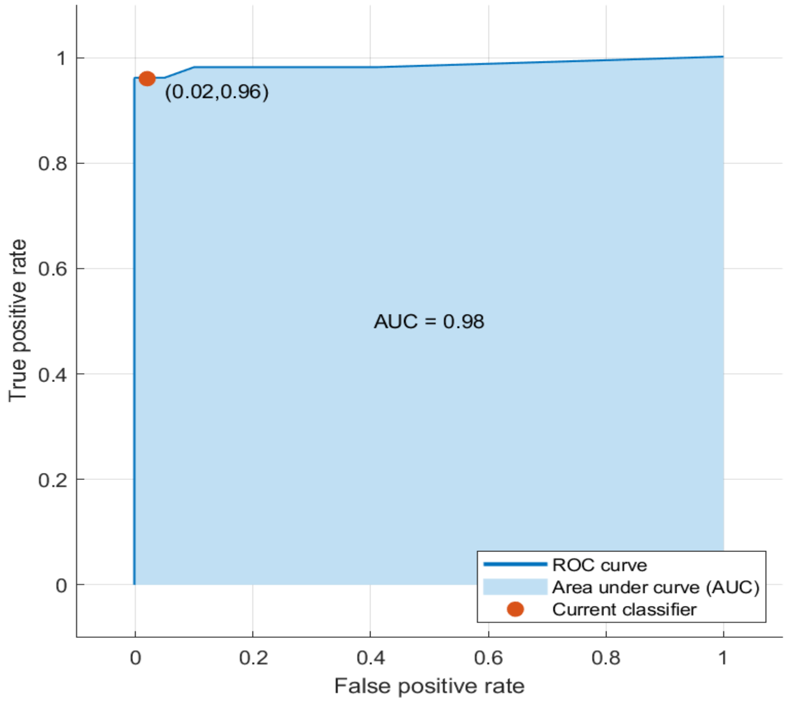



**Figure 4- ROC curve for Decision Tree model**

*3.2. Support vector machine classifier*

SVM is a popular machine learning algorithm that is used for classification problems. In the context of classifying maize leaves into common rust, gray spot, and healthy, SVM can be a viable solution. In this experiment, an SVM classifier was used to classify diseased maize leaves (common rust, gray leaf spot), and healthy maize leaves. The dataset was divided into a training and testing set, with the training set being used to train the model and the test set used to evaluate its accuracy. The results of the model show that it had 91.3% accuracy rate on the test dataset, indicating that it was able to classify diseased and healthy leaves correctly with a high degree of accuracy. The confusion matrix and ROC curve for the SVM classifier are shown in Figures 5 and 6.

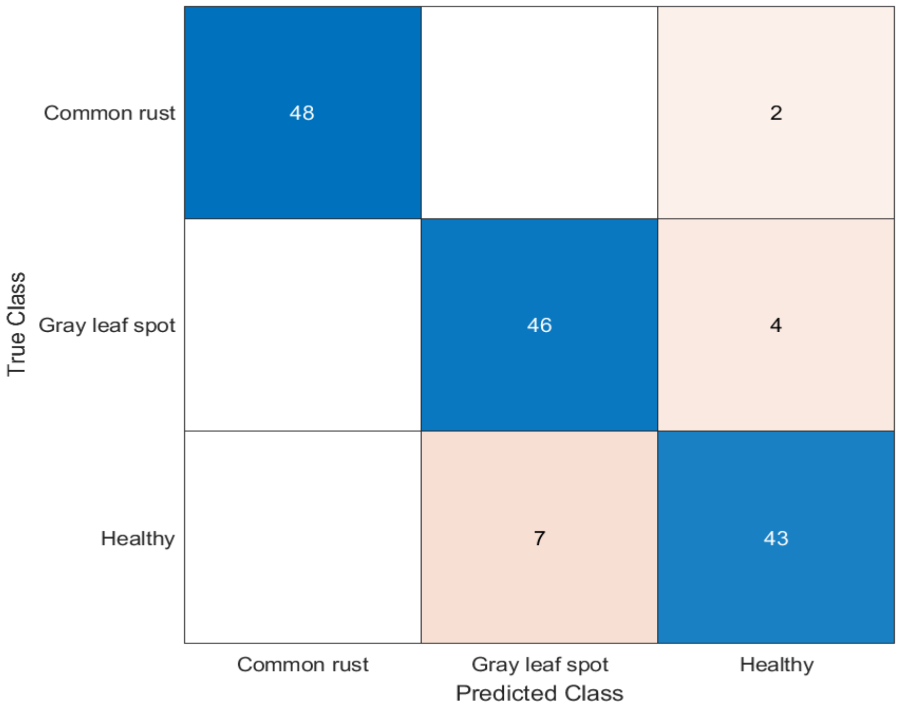



**Figure 5- SVM confusion matrix**

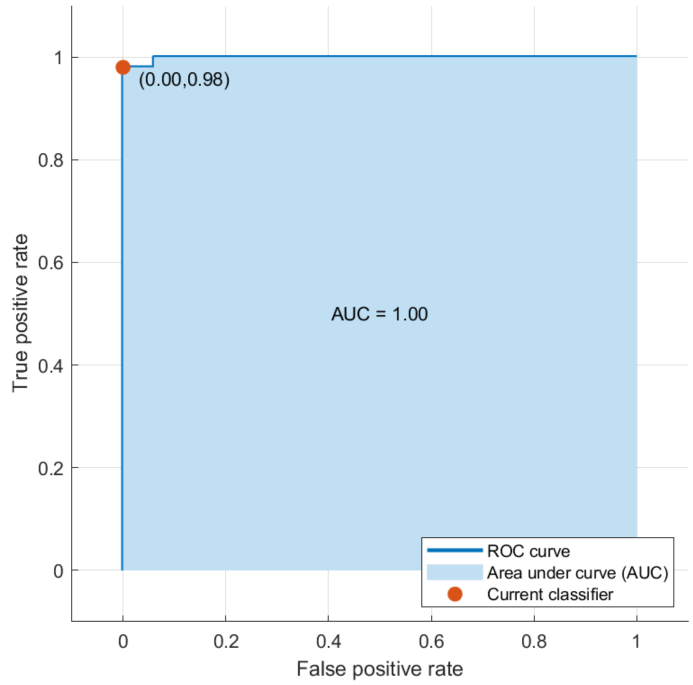



**Figure 6- ROC curve for Support Vector Machine model**

*3.3 K- Nearest neighbor classifier*

The KNN model is a machine learning algorithm that can be used to detect diseased leaves (common rust and gray leaf spot) and healthy maize leaves in images. The model works by classifying a test image based on the closest match to the images in the training dataset. The algorithm is simple and efficient, making it a popular choice for image classification tasks. To implement the KNN model for maize leaf detection, a dataset of images of diseased maize leaves and healthy images was used to train and test the model. The result showed that the KNN model was able to achieve an accuracy of 90% in the test dataset, indicating that it was able to classify diseased and healthy leaves correctly with a high degree of accuracy. The confusion matrix and ROC curve, shown in Figures 7 and 8, provide further insight into the performance of the KNN classifier.

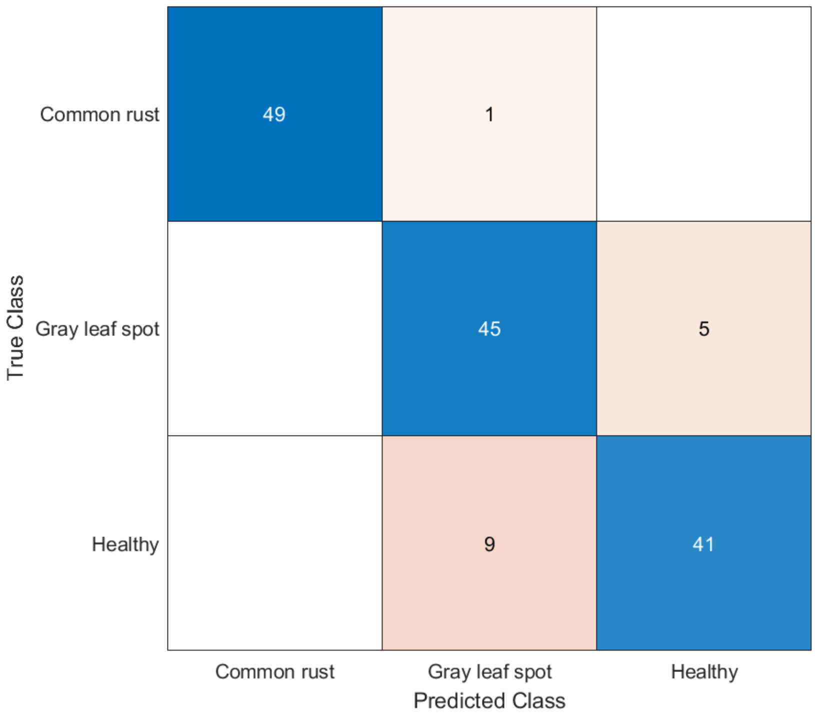



**Figure 7- KNN confusion matrix**

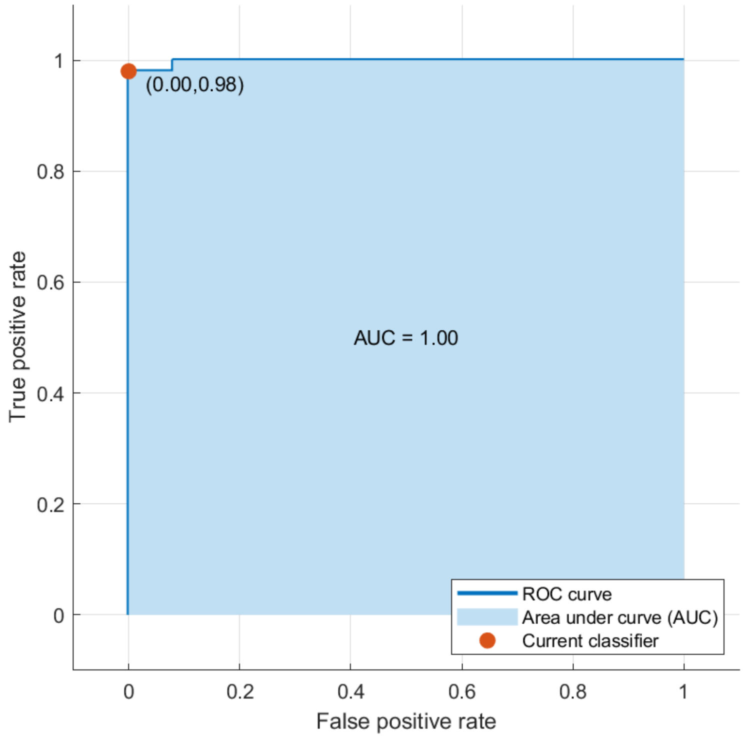



**Figure 8- ROC curve for K-Nearest Neighbors model**

*3.4 Artificial neural network classifier*

The ANN model is a machine learning technique that can be used to detect diseased leaves (common rust and gray leaf spot) and healthy maize leaves in images. The ANN model works by learning a mapping between inputs and outputs based on a training dataset. The mapping is represented by a network of interconnected nodes that process information and make predictions. The results indicated that the model achieved an accuracy rate of 92.7% for the test dataset, demonstrating its ability to accurately distinguish between healthy and diseased maize leaves. The confusion matrix and ROC curve for the ANN classifier are shown in Figures 9 and 10.

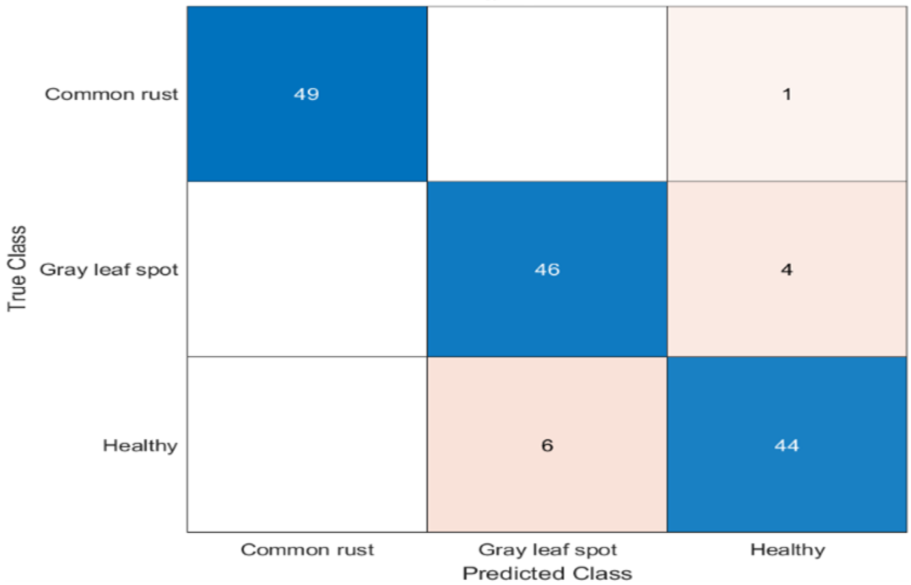



**Figure 9- ANN confusion matrix**

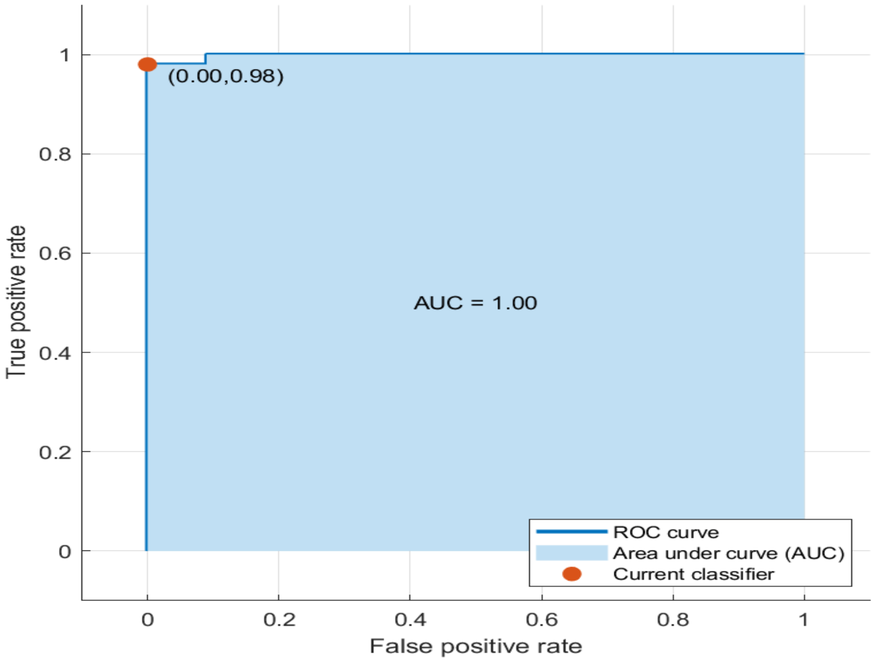



**Figure 10- ROC curve for Artificial Neural Network model**

## 4. Discussion

Many studies have been carried out to enhance the accuracy of detecting maize diseases by applying machine learning and deep learning algorithms. Table 1 presents the studies focused on classifying diseased and healthy maize, rice, grape, and tomato leaves. Table 1 also gives the diseases that were classified, the models used, and their accuracy rates.

Ramesh & Vydeki (2019) proposed a machine learning-based classification methodology that included KNN and ANN. Statistical features, including mean and standard deviation along with GLCM texture features, were extracted to detect blast disease in Indian rice. Test accuracies of 63 and 88% for normal images were achieved using KNN and ANN algorithms, respectively, and test accuracies of 79 and 90% were achieved for blast-affected images using KNN and ANN, respectively.

**Table 1- Comparison of accuracy rates of classifying diseased and healthy maize, rice, grape, and tomato leaves using traditional machine learning and deep learning models**

| *Author(s)* | *Model architecture* | *Detected diseases* | *Accuracy* |
|---|---|---|---|
| Ramesh & Vydeki (2019) | KNN and ANN | Healthy images, blast affected images | 63% and 88%.<br>79% and 90%. |
| Panigrahi et al. (2020) | SVM, NB, KNN, DT, and RF | Northern leaf blight, common rust, and Cercospora leaf spot | 77.56%, 77.46%, 76.16%, 74.35%, and 79.23%. |
| Padol & Yadav (2016) | Linear SVM | Grape leaf images | 93.33% |
| Chokey & Jain (2019) | Linear SVM, quadratic SVM, and cubic SVM | Common smut, common rust | 79.5%, 73.1%, and 83.3%. |
| Mukhopadhyay et al. (2021) | SVM | Healthy or infected tomato leaves. | 99.83% |
| Waheed et al. (2020) | XceptionNet, EfficientNet, VGG19Net, NASNet, and Optimized DenseNet | Common rust, Cercospora leaf spot, gray leaf spot, northern leaf blight, and healthy crop. | 93.52%, 99.84%, 96.36%, 91.9%, and 98.06%. |
| Subramanian et al. (2022) | VGG16, ResNet50, InceptionV3, and Xception | Blight, common rust, grey leaf spot, healthy. | 95.04%, 79.44%, 96.9%, and 95.45%. |
| Yang et al. (2023) | VGG11, EfficientNet, Inception-v3, MobileNet, and ResNet50.<br>ViT, Improved ViT, and multi-path context feature aggregation network. | Leaf blight, gray leaf, healthy leaf, and leaf rust | 97.9%, 91.6%, 97.2%, 90.2% and 96.6%.<br>93.9%, 98.7%, and 99.5% |
| Xu et al. (2021) | VGGNet-16, DenseNet, ResNet, AlexNet, and TCI-ALEXN. | Northern leaf blight, common rust, healthy, and gray leaf spot. | 86.17, 88.8, 90.48, 90.86, and 93.28. |
| Our study | DT, SVM, KNN, and ANN. | Grey leaf spot, common rust, and healthy leaves. | 90%, 91.3%, 90%, and 92.7%. |

In this study, we focused on using traditional machine learning to detect healthy and diseased maize leaves. In our initial experiment, extracting statistical and GLCM texture features without processing resulted in a test classification accuracy ranging between 66.7% and 72.7%. Attempting to enhance the accuracy, initially color features were extracted, yet this did not yield improvement due to the close similarity in shape between the diseased classes. To address this challenge, K-means clustering was implemented to separate the diseased leaf portions. After applying K-means clustering, the test classification accuracy significantly improved from 66.7, 70.7, 70.7, and 72.7% to 90, 90, 91.3, and 92.7%, using DT, SVM, KNN, and ANN respectively.

From the previous studies presented in Table 1, few studies focused on detecting plant diseases using traditional machine learning techniques, especially extracting GLCM texture features along with statistical features. The highest accuracy (99.83%) was achieved in a study conducted by Mukhopadhyay et al. (2021) to classify healthy and infected tomato leaves. They obtained this high accuracy because they focused only on classifying tomato leaves as healthy or infected, so that the differences between these two classes were clear compared to our study, in which we focused on classifying two different diseases and healthy maize leaves. Another study was conducted by Ramesh & Vydeki (2019) to detect healthy and blast-affected images. GLCM texture features, along with statistical features, were extracted from the Indian rice images. They achieved test accuracies of 63%, 88%, 79%, and 90% for healthy and blast-affected rice images, which is less than the highest accuracy achieved in our study of 92% using the ANN model. This highlights the efficiency of our proposed traditional machine learning model in detecting different types of maize diseases effectively and with high accuracy. Also, from previous studies, the use of deep learning models to detect healthy and diseased maize leaves achieved very high accuracy, ranging from 79.74% to 99.84%, but requires a large amount of data. Instead of just focusing on one approach, it is necessary to improve the accuracy of existing traditional machine learning and see how they perform.

## 4. Conclusions

This study aimed to identify healthy and diseased maize leaves using traditional machine learning algorithms. For this purpose, six hundred images from the Plant Village dataset were used, which were categorized into three groups: healthy leaves (200 images), grey leaf spot (200 images), and common rust (200 images). K-means clustering was employed for image segmentation, and statistical features such as mean and standard deviation, along with GLCM texture features, were extracted after the segmentation process. Then the dataset was split into training and testing sets, where 75% of the images were used for training and 25% for testing. Four conventional ML algorithms, DT, SVM, KNN, and ANN models, were employed to train and test the images.

The study findings showed that all four algorithms performed well in accurately detecting diseased and healthy maize leaves, with accuracy rates ranging from 90% to 92.7%. Support vector machines and artificial neural networks achieved the highest accuracy scores, which suggests that these algorithms are more effective in identifying diseased maize leaves.

## References


Chokey T & Jain S (2019). Quality Assessment of Crops using Machine Learning Techniques. *2019 Amity International Conference on Artificial Intelligence (AICAI)* pp. 259–263. https://doi.org/10.1109/AICAI.2019.8701294

Dhingra G, Kumar V & Joshi H D (2019). A novel computer vision based neutrosophic approach for leaf disease identification and classification. *Measurement* 135: 782–794. https://doi.org/10.1016/J.MEASUREMENT.2018.12.027

Donatelli M, Magarey R D, Bregaglio S, Willocquet L, Whish J P M & Savary S (2017). Modelling the impacts of pests and diseases on agricultural systems. *Agricultural Systems* 155: 213–224. https://doi.org/10.1016/J.AGSY.2017.01.019

Fulari U, Shastri R & Fulari A (2020). Leaf Disease Detection Using Machine Learning. *Journal of Seybold Report* 15(9): 1828–1832

Géron A (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (R. Roumeliotis & N. Tache, Eds.; 2nd ed.). O'Reilly Media, Inc.

Haque M, Marwaha S, Deb C, Nigam S & Arora A (2023). Recognition of diseases of maize crop using deep learning models. *Neural Computing and Applications* 35(10): 7407–7421. https://doi.org/10.1007/s00521-022-08003-9

Jasrotia S, Yadav J, Rajpal N, Arora M & Chaudhary J (2023). Convolutional Neural Network Based Maize Plant Disease Identification. *Procedia Computer Science* 218(1): 1712–1721. https://doi.org/10.1016/J.PROCS.2023.01.149

Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H & Wang Y (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology* 2. https://doi.org/10.1136/svn-2017-000101

Mokhtar U, El-Bendary N, Hassenian A, Emary E, Mahmood M, Hefny H & Tolba M (2015). SVM-Based Detection of Tomato Leaves Diseases. In D. Filev & J (Eds.), *Advances in Intelligent Systems and Computing* (Vol. 323, pp. 641–652). Springer, Cham. https://doi.org/10.1007/978-3-319-11310-4_55

Mukhopadhyay Dr. S, Paul M, Pal R & De D (2021). Tea leaf disease detection using multi-objective image segmentation. *Multimedia Tools and Applications*, 80(1). https://doi.org/10.1007/s11042-020-09567-1

Oerke E C & Dehne H W (2004). Safeguarding production—losses in major crops and the role of crop protection. *Crop Protection* 23(4): 275–285. https://doi.org/10.1016/J.CROPRO.2003.10.001

Padol P B & Yadav A A (2016). SVM classifier based grape leaf disease detection. *2016 Conference on Advances in Signal Processing (CASP)* pp. 175–179. https://doi.org/10.1109/CASP.2016.7746160

Panigrahi K P, Sahoo A K & Das H (2020). A CNN Approach for Corn Leaves Disease Detection to support Digital Agricultural System. *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)* pp. 678–683. https://doi.org/10.1109/ICOEI48184.2020.9142871

Ramakrishnan M & Sahaya A N A (2015). Groundnut leaf disease detection and classification by using back probagation algorithm. *2015 International Conference on Communications and Signal Processing (ICCSP)* pp. 964–968. https://doi.org/10.1109/ICCSP.2015.7322641

Ramesh S & Vydeki D (2019). Application of machine learning in detection of blast disease in South Indian rice crops. *Journal of Phytology* 11(1): 31–37. https://doi.org/10.25081/jp.2019.v11.5476

Sibiya M & Sumbwanyambe M (2021). Automatic Fuzzy Logic-Based Maize Common Rust Disease Severity Predictions with Thresholding and Deep Learning. *Pathogens* 10(2): 131. https://doi.org/10.3390/pathogens10020131

Singh V, Varsha & Misra A K (2015). Detection of unhealthy region of plant leaves using image processing and genetic algorithm. *2015 International Conference on Advances in Computer Engineering and Applications* pp. 1028–1032. https://doi.org/10.1109/ICACEA.2015.7164858

Subramanian M, Shanmugavadivel K & Nandhini P S (2022). On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves. *Neural Computing and Applications* 34: 13951–13968. https://doi.org/10.1007/s00521-022-07246-w

Vasavi P, Punitha A & Rao V N T (2022). Crop leaf disease detection and classification using machine learning and deep learning algorithms by visual symptoms: A review. *International Journal of Electrical and Computer Engineering* 12(2): 2079–2086. https://doi.org/10.11591/ijece.v12i2.pp2079-2086

Waheed A, Goyal M, Gupta D, Khanna A, Hassanien A E & Pandey H M (2020). An optimized dense convolutional neural network model for disease recognition and classification in corn leaf. *Computers and Electronics in Agriculture* 175(12): 105456. https://doi.org/10.1016/j.compag.2020.105456

Xu Y, Zhao B, Zhai Y, Chen Q & Zhou Y (2021). Maize Diseases Identification Method Based on Multi-Scale Convolutional Global Pooling Neural Network. *IEEE Access* 9: 27959–27970. https://doi.org/10.1109/ACCESS.2021.3058267

Yang W, Shen P, Ye Z, Zhu Z, Xu C, Liu Y & Mei L (2023). Adversarial Training Collaborating Multi-Path Context Feature Aggregation Network for Maize Disease Density Prediction. *Processes* 11(4): 1132. https://doi.org/10.3390/pr11041132

Zhang X, Qiao Y, Meng F, Fan C & Zhang M (2018). Identification of Maize Leaf Diseases Using Improved Deep Convolutional Neural Networks. *IEEE Access* 6: 30370–30377. https://doi.org/10.1109/ACCESS.2018.2844405