RESEARCH ARTICLE

SAMRIDDHI Volume 14, Issue 1, 2022 Print ISSN: 2229-7111 Online ISSN: 2454-5767

Ant Colony Optimization Algorithm for Disease Detection in Maize Leaf using Machine Learning Techniques

Alok Kumar[1]*, Manmohan Singh[2]

[1]Department of Computer Science and Engineering, Government College of Engineering, Dharmapuri, Tamil Nadu, India
[2]Department of Computer Science and Engineering, K L University, Guntur, Andhra Pradesh, India

Abstract

Plant diseases have affected the productivity of food in recent years. Because of this productivity loss, humans and animals are also affected, but the whole biodiversity would be affected. So, we should take preventive measures to stop this food destruction. Both humans and animals largely consume the maize crop. Due to some factors, Maize leaf is easily affected by some fungal or other diseases. Farmers could not find out the leaf diseases at the early stages. They need some advanced methods to detect these types of diseases. Early detection of leaf disease helps farmers to increase the Maize yield. In the proposed algorithm, we have used five supervised machine learning algorithms such as K-nearest neighbor (KNN), naive bayes (NB), support vector machine (SVM), random forest (RF), logistic regression (LR) and then these machine learning models have been implemented with the Ant colony optimization (ACO) Algorithm for optimizing the accuracy of disease detection in Maize Leaf. For leaf classification, color and texture features are extracted from an input dataset. Features of a leaf can be described by Hu moments, Haralick texture, and color histogram. After performing all Machine learning classifiers, we have analyzed that Random Forest with Ant Colony Optimization Algorithm gives the highest accuracy of 99.4% for disease detection in Maize Leaf.

Keywords: Colour Histogram, Haralick texture, Hu moments, Machine Learning Classifiers, Supervised Machine learning algorithms.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology (2022); DOI: 10.18090/samriddhi.v14i01.5

Introduction

Agriculture is one of the pillars of India and also for the world.[1] Food is essential for all living beings. If there is any change in the production of food quality, it will affect the whole world and the country's economic growth. Mostly, the crops are affected by leaf diseases. Identifying the early detection of the leaf disease was difficult for farmers. So, there is a need for farmers to identify the early detection of plant diseases.

Maize is one of the largest old crops in the world.[2] Zea mays are the scientific name of the Maize, which belongs to Poaceae. It is also known as corn and is originated from Mexico and Central America. It is also used to produce cereals and flour. It contains a high amount of starch. Both humans and animals can consume this. It gives the same food production quantity as same as wheat and rice. The reason for Maize leaf disease is due to the bacterial fungal and the deficiency of minerals in the soil, and the effect of this problem reflects in the leaf.

Corn Leaf Blight is caused by the fungus called Helminthosporium turcicum Pass. This fungus affects the crops at a younger age. This will affect the lower leaves first. At the starting stage, it will be like oval-shaped spots on leaves. Later, it becomes the cigar-shaped lesson, and it spreads all over the plant and leads the plant to die. Corn blight is shown in Figure 1.

Common Rust disease is caused due to the fungus named Puccina Sorghi. Symptoms of common rust are chlorotic flecks on the leaf surface. Pustules are oval-shaped and scattered through the leaf. Sometimes, these lessons spread over the leaf and cause the entire leaf to be dead. These pustules are in red at an early stage and then turn black. Common Rust is shown in Figure 2. Healthy leaf is free from any disease like blight and common rust. Healthy leaf is shown in shown in Figure 3.

Corresponding Author: Alok Kumar, Department of Computer Science and Engineering, Government College of Engineering, Dharmapuri, e-mail: gcedpialok@gmail.com

How to cite this article: Kumar, A., & Singh, M. (2022). Ant Colony Optimization Algorithm for Disease Detection in Maize Leaf using Machine Learning Techniques. *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology,* 14(1), 31-37.

Source of support: Nil

Conflict of interest: None

Figure 1: Blight

Figure 2: Common Rust

Figure 3: Healthy

This paper helps the farmers to detect the early-stage disease at the leaf. The image of tree leaf is the basic component of this paper, and all the steps are performed based on the leaf image. So, the image of the leaf should be clean, and this process should be done in pre-processing step. Here, color and texture features are extracted from the training step and classification process. Features of a leaf can be described by Hu moments, Haralick texture, and color histogram. In this paper, we use the different types of Machine learning model supervised learning classifiers such as Logistic Regression (LR), Naive Bayes (NB), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM), and Random Forest (RF). Random Forest Classifier gives the highest accuracy of all the ML algorithms after applying Ant Colony Optimization Algorithm. It gives the most precise result. Finally, the simulated result shows that the used networks classifier provides minimum error during training and better classification accuracy.

This paper is organized as follows. Part 2 represents related work of Maize Leaf Disease Detection and Classification. Section 3 represents the detailed explanation of Machine Learning Supervised Model classifiers. Section 4 presents the working of a proposed method for Maize leaf disease detection. Section 5 describes the Experimental Result Analysis. Finally, section 6 provides the conclusion and future scope of the proposed method.

Literature Review

In this section, we give the details about the related papers which describes how to detect the leaf disease and also the various image processing techniques.

M. Haggag *et al.*, (2019) described learning vector quantization (LVQ) to classify the images of the leaf.[3] X. Ye and Q. Zhu, (2019) used an incremental learning algorithm and support vector data description. For feature extraction, they use the CNN and AM-SoftMax loss function.[4] P. Jiang *et al.*, (2019) focuses on a deep learning approach that is based on improved convolutional neural networks. In this INAR-SSD model is trained and developed.[5]

P. Revathi and M. Hemalatha, (2018) used edge detection algorithms on the pre-processed image to detect the diseased spot area edge. Image segmentation can be done by K-means clustering algorithm and features are extracted from the segmented image. SVM classifier are used to differentiate the diseased leaf from the healthy leaf.[6]

K. P. Ferentinos, (2018) used convolutional neural network models to detect the plant disease. Different deep learning methodologies were used. Dataset contains 25 different plants of about 87,848 images with a set of separate classes.[7] M. A. Khan *et al.*, (2018) proposed Correlation coefficient-based segmentation method which separates the infected areas from the background. VGG16 and Caffe AlexNet are used for feature extraction. Support Vector Machine algorithm are used to detect the diseased leaf.[8]

H. Al-Hiary *et al.*, (2017) proposed the new method consists of four steps. First, they identify the most green colored pixels. In second step, pixels are masked based on the specific threshold values. Last, pixels with zero, red, green and blue values are completely removed. At last disease was identified.[9]

J. Sun *et al.*, (2017) proposed a traditional convolutional neural network and also combined with squeeze and excitation (SE) module is used to identify the disease. With the help inception structure, the characteristic of the disease leaf dataset can be identified easily.[10]

Machine Learning Classification Algorithms

Classification is the important step that identifies the image which belongs to which group. This classifies the image based on some supervised machine learning models. The classification method comprises of two-part such as the training part and the testing part. In the training part, the dataset is trained with Machine learning Algorithms to give an accurate output. Then, the testing part is used to evaluate the model's performance using some performance metrics. The testing part mainly focuses on how fast and how much time it will take to give the precise output. The description of different types of classification models such as LR, NB, KNN, SVM, and RF are used for analysis is described below.

Logistic Regresssion (LR)

Logistic regression is a supervised learning technique that classifies the values.[2] In this, we use the 'S-shaped logistic function called sigmoid function, which predicts the two maximum values (0 or 1). It is the mathematical function used to predict the output values. It also uses the threshold value concept. The values above the threshold value go to 1, and the values below the threshold value go to 0. Based on the observations, it can be divided into three types. These are:

Binomial logistic regression: The output value could be only two possible dependent variables such as 0 or 1.

Multinomial logistic regression: The output value maybe 3 or more than 3 unordered types of dependent variables. eg: Determining the shapes, square, rectangle, and triangle values.

Ordinal logistic regression: The output value maybe 3 or more than 3 ordered types of dependent variables. eg: Determining the sizes, the values should be low, medium, and height.

Naive Bayes (NB)

Naive Bayes classifier is a supervised machine learning algorithm based on Bayes theorem, which is a variant of a probabilistic classifier.[1] It predicts the output value based on probability. It determines the probability of a hypothesis with prior knowledge, which depends on conditional probability. Bayes' Theorem is stated as:

P (h | d) = (P (d | h) * P(h)) / P(d)

here,

P (h | d) is Posterior probability

P (d | h) is Likelihood probability

P (h) is Prior Probability

P (d) is Marginal Probability

K-nearest Neighbors (KNN)

It is a supervised machine learning technique and it assumes the similarity between the training data and the test data. [1] It puts the test data into the correct category. It is the simplest method for both classification and regression. It does not make any assumption on underlying data, which is also known as a non-parametric algorithm. For example, if we have a picture of a living being, the system will find whether the picture belongs to a dog or cat. First, it will find the similar characters between those pictures. Then, it will classify the pictures.

The working of the KNN algorithm:

- choose the number of neighbors K
- calculate the Euclidean distance from the number of neighbors K
- Take the K nearest neighbor data based on the Euclidean distance
- Count the amount of data in each category and assign the new category based on the number of neighbors is maximum

Euclidean distance formula:

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Support Vector Machine

It is a supervised learning method and is best for classification and regression.[1] It is also known as SVM. In this, the best line creates a boundary between the classes in the n-dimensional plane. The decision boundary line is called a hyperplane. The dimensions of the hyperplane depend upon the features of the dataset. The hyperplane should always have a maximum margin, meaning the maximum distance between the data points. Suppose if we have two features, then the hyperplane will be a straight line. More than two features, it will be a 2-dimension plane. Support vectors help to build the SVM classifier. Support vectors mean data points closer to the hyperplane and influence the orientation of the hyperplane. There are two types of SVM classifiers. They are Linear SVM and Non-linear SVM.

Random Forest

A random forest algorithm is a supervised machine learning algorithm.[2] The algorithm's name means it creates the forest and produces the output based on the random picking data. The final output can be obtained by majority voting. The only difference between the random forest and decision tree is finding root rode and splitting the feature nodes randomly.[11] It contains mainly two steps. First, it should develop a random forest. Next, predict the values from the random forest. It is ensemble learning, combining multiple classifiers and predicting the output value. It contains the number of decision trees because instead of predicting the output value based on one decision tree, it predicts the output value from the majority of the decision tree. How large the tree will develop; the more accuracy will gain, and the overfitting problem can be avoided.[12]

The working steps of a Random Forest:

- Select the K random data points from the dataset.
- Build the decision tree from those datasets.
- Choose the N number of decision trees that you want to build
- Repeat the first two steps until the maximum number of trees was built
- The new data points will classify according to the trees by majority voting.

Working on a Proposed System

This section describes the working of a proposed system. The proposed system contains Image procurement, image pre-processing, Image segmentation, Feature extraction, and classification. In classification, we have used the supervised machine learning algorithms such as K-Nearest neighbor (KNN), Naive Bayes (NB), Support vector machine (SVM), Logistic regression (LR), Random forest (RF). Then these machine learning models have been implemented with the Ant Colony Optimization (ACO) Algorithm for optimizing the accuracy of disease detection in Maize Leaf. The description of each step is given below:

Image Procurement

Image Procurement is the primary step for image classification. It is the first step in image processing.[13] In our paper, the image dataset was retrieved from the Kaggle database. We take 3200 data in the form of images, and three class labels such as Common Rust, Leaf Blight, and Healthy, having 800,800 and 1600 images, respectively.

Image Pre-processing

The reason for using this is to get a better image standard, which does not mean increasing the image content. It only deletes the unwanted noises in the image. The unwanted noises are due to the dust few drops of water. The image sizes of all the leaves can be resized. It is used to convert the raw images into the correct sized image, which means images can be processed with a suitable Machine learning model.[14] RGB images are converted into HSV, and then HSV is converted into the grayscale format.

Image Segmentation

There are two classifications of Image segmentation i.e. Semantic Segmentation and Instance Segmentation. In Semantic Segmentation, we segment the image pixels into their respective classes. In Instance Segmentation, we deal with multiple objects. Image segmentation divides the images into multiple segments, also known as pixels, for further processing. Segmentation algorithms work on the base of two types as Similarities and Discontinuity.

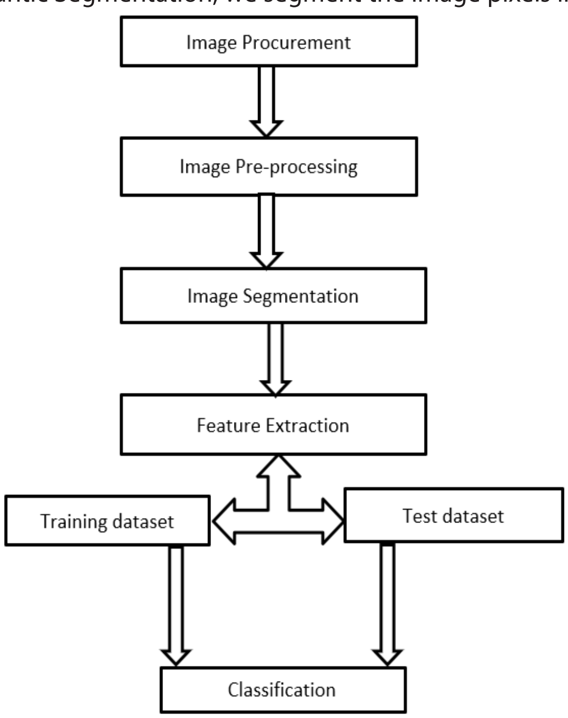

Figure 4: Diagram for the classification process

Feature Extraction

In the feature extraction process, an initial set of raw data is reduced to more manageable groups for processing. Feature extraction brings out the characteristics of the leaf, which is used to classify the images. Feature of the leaf can be obtained by Haralick texture, Hu Moments, and Color Histogram. Feature of the leaf based on shape, color, and texture. The following paragraph describes the feature extraction methods.

Haralick Texture

Haralick texture features are used to describe the "texture" of an image. Haralick texture features are used to distinguish between rough and smooth surfaces. Haralick texture features are calculated from the Gray Level Co-occurrence Matrix. It consists of two steps. In the first step, the GLCM is computed, and in the second step, features are calculated. It is a square matrix that counts the number of Gray levels in the image.

HU MOMENTS

Hu Moments are an image descriptor utilized to characterize the shape of an object in an image. The shape to be described can either be a segmented binary image or the boundary of the object. Hu Moments are used to describe the shape of the leaf. Hu Moments can be calculated from the OpenCV built-in function. There are seven moments in Hu Moments.

Color Histogram

Color histogram is a method to represent the color feature of an image by counting how many values of each color occur in the image and forming a representing histogram. Color Histogram is used to extract the color feature of an image. It can define the color of the leaf. Counting the number of Pixels in each of the bins, the Color Histogram of the leaf image can be calculated.

Classification of Leaf

Based on Feature extraction, the leaf can be classified into diseased leaf or not. In the Classification process, two phases are performed, the training and testing phase. For performing the classification of the images, we used various Supervised Machine Learning classifiers, and then these machine learning models have implemented with the Ant Colony Optimization (ACO) Algorithm for optimizing the accuracy of disease detection in Maize Leaf. In Figure 4, we have shown the classification process.

Ant Colony Optimization Algorithm

As mentioned before ACO algorithm is a meta-heuristic that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. In general, meta-heuristic doesn't guarantee an optimal solution is ever found.

Pseudo Code:

u = 0 /* Initialization of the iteration counter */
Q = Initialize population () /* Generation of q individuals */ Repeat until end condition
τ = Update pheromone matrix (Q)
T = Construct solutions (τ)
T_{best} = best solution of T
Q = Update population (T_{best})
u = u + 1 /* Increment of the iteration counter */

Experimental Result Analysis

We executed ACO - ML algorithms on image datasets to analyze the quality of the proposed system. All the datasets are executed on the platform of Intel Core (TM) i7, 4.50 GHz processor having 8GB main memory with 64-bit based, using a 64-bit operating system.

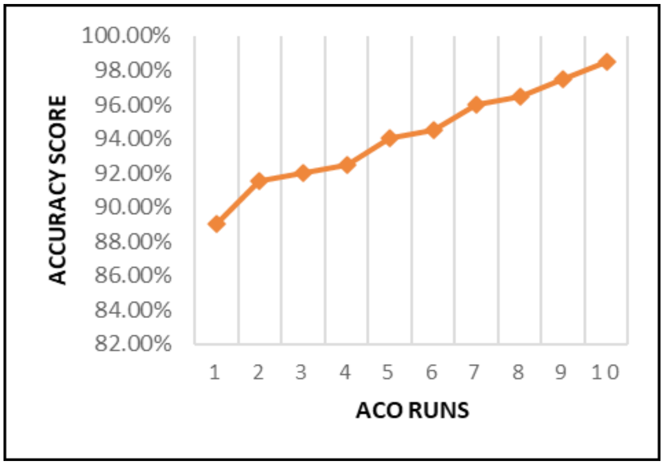

Figure 5: Graphical Representation of Accuracy Using KNN Algorithm

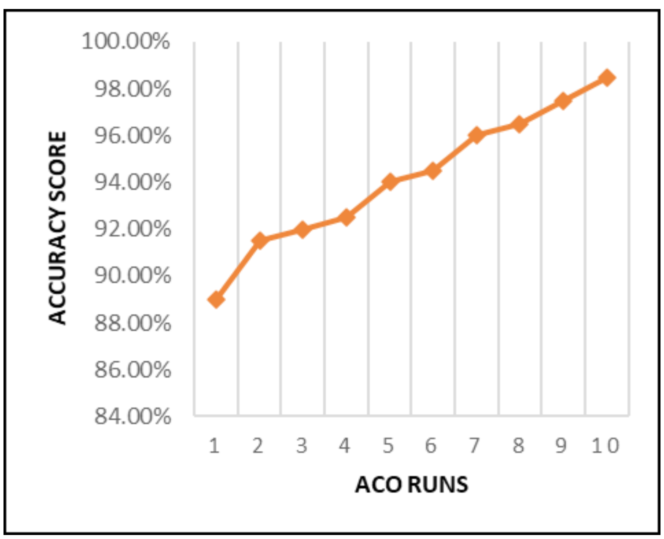

Figure 6: Graphical Representation of accuracy using NB Algorithm

In Table 1, we have compared the accuracy score of KNN through various runs of ACO Algorithm. In Figure 5, we have plotted the Table 1 data; on the x-axis, we have kept ACO runs, and on the y-axis, we have kept the accuracy score so that easily anyone can understand. In Table 2, we have compared the accuracy score of NB through various runs of ACO Algorithm. In Figure 6, we plotted the Table 2 data with a 10 runs accuracy score. In Table 3, we have compared the accuracy score of SVM through various runs of ACO Algorithm. In Figure 7, we have plotted the Table 3 data. In Table 4, we have compared the accuracy score of LR through various runs of ACO Algorithm. In Figure 8, we have plotted the Table 4 data. In Table 5, we have compared the accuracy score of RF through various runs of ACO Algorithm. In Figure 9, we have plotted the Table 5 data. In Table 6, we have compared the best accuracy score of all different 5 machine learning algorithms and analyzed that RF algorithm has given the best accuracy score with ACO algorithm. In Figure 10, we have plotted the Table 6 data, we have Graphically represented the accuracy score of all different 5 machine learning algorithms. From the above graph and accuracy table, we concluded that

Table 1: Comparing Accuracy of KNN Through Various Runs Of ACO

ACO RUNS	*ACCURACY SCORE*
1	89.0%
2	91.5%
3	92.0%
4	92.5%
5	94.0%
6	94.5%
7	96.0%
8	96.5%
9	97.5%
10	98.5%

Table 2: Comparing Accuracy of NB Through Various Runs Of ACO

ACO RUNS	*ACCURACY SCORE*
1	89.00%
2	90.50%
3	91.20%
4	91.78%
5	92.56%
6	93.90%
7	94.60%
8	95.90%
9	96.50%
10	98.20%

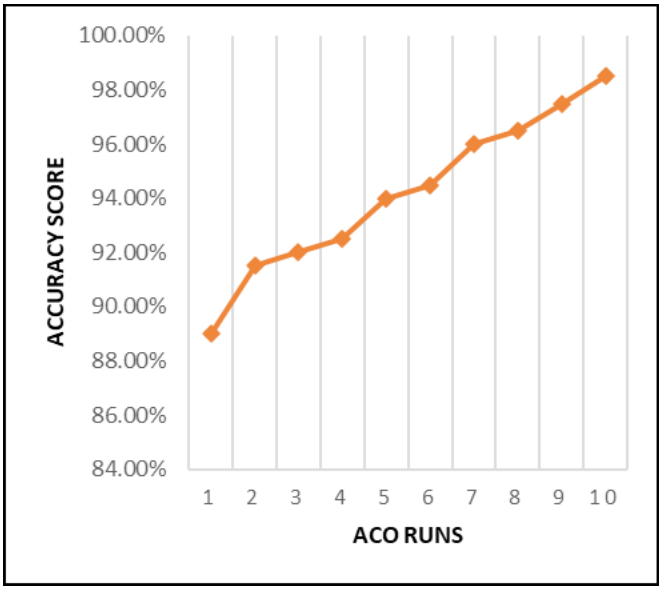

Figure 7: Graphical representation of accuracy using SVM Algorithm

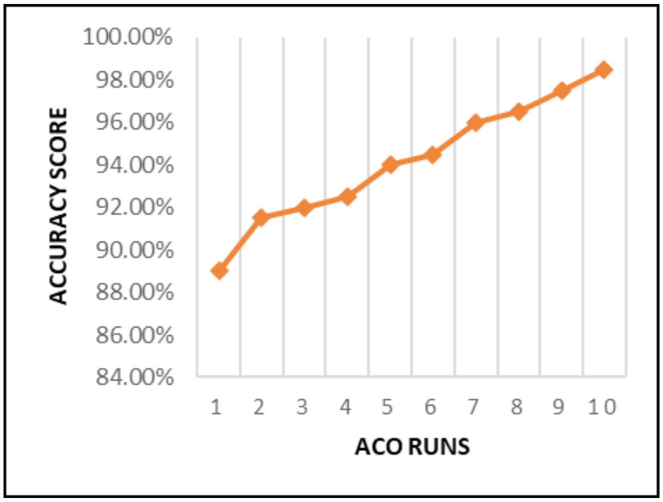

Figure 8: Graphical representation of accuracy using LR algorithm

Table 3: Comparing the accuracy of SVM through various runs of ACO

ACO RUNS	*ACCURACY SCORE*
1	87.24%
2	88.50%
3	90.50%
4	92.30%
5	94.50%
6	96.34%
7	97.50%
8	97.80%
9	98.30%
10	98.40%

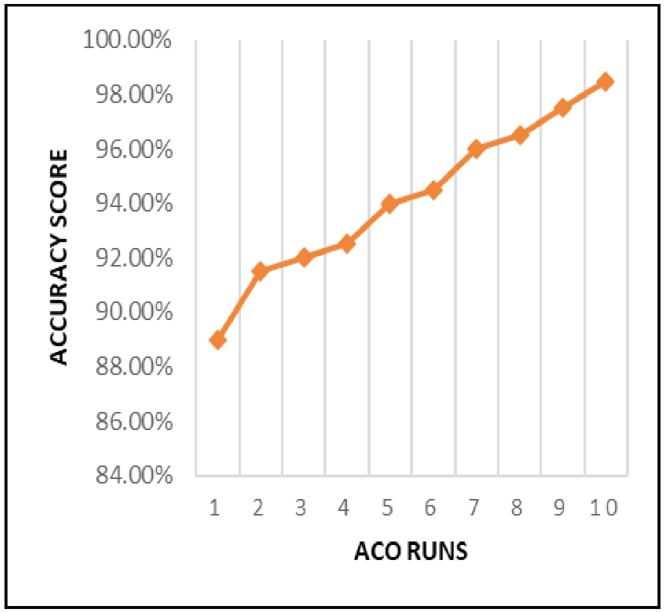

Figure 9: Graphical representation of accuracy using RF Algorithm

Table 4: Comparing the accuracy of LR through various runs of ACO

ACO RUNS	*ACCURACY SCORE*
1	88.24%
2	89.50%
3	90.80%
4	91.37%
5	93.50%
6	95.64%
7	97.23%
8	97.56%
9	97.30%
10	98.90%

Table 5: Comparing the accuracy of RF through various runs of ACO

ACO RUNS	*ACCURACY SCORE*
1	89.24%
2	90.25%
3	91.68%
4	93.47%
5	95.50%
6	96.44%
7	97.63%
8	98.26%
9	98.90%
10	99.40%

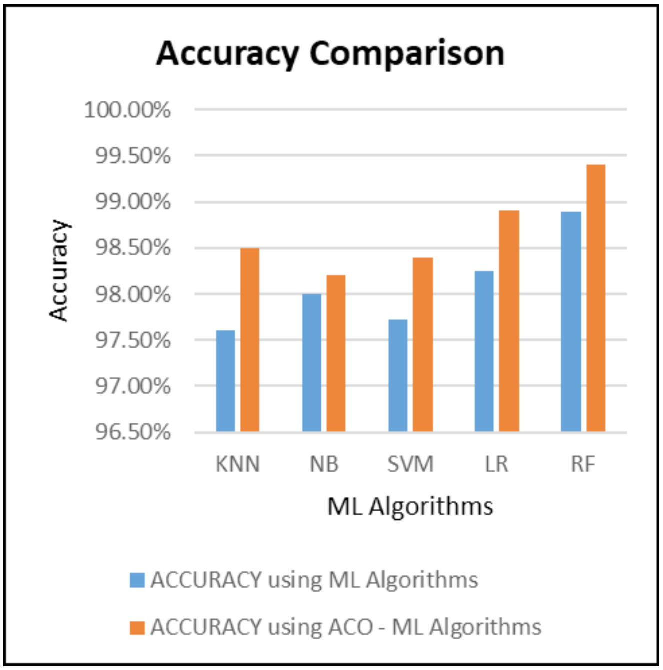

Figure 10: Graphical representation of overall accuracy comparison

Table 6: Prediction of overall accuracy for each ML model

ML Algorithms	*ACCURACY using ML algorithms*	*ACCURACY using ACO - ML algorithms*
KNN	97.60%	98.50%
NB	98.00%	98.20%
SVM	97.72%	98.40%
LR	98.25%	98.90%
RF	98.89%	99.40%

the highest accuracy of all is 99.40% for both common rust and blight, which Random Forest Classifier gave.

Conclusions & Future Direction

This paper correctly identifies and implements the various Machine learning models along with Ant Colony Optimization algorithm. We used 3200 leaf images in order to predict the maize leaf disease detection. Based on the results, we have concluded that this paper helps to detect Maize leaf diseases at the early stage. This paper also executes the different algorithms to get the highest disease prediction classifier. With the help of a Random Forest classifier with ACO Algorithm, early leaf disease detection can be done, and we can avoid productivity loss.

Finally, we concluded that the ACO Algorithm was successfully implemented and worked well for both common rust and blight datasets. ACO Algorithm worked better for all the algorithmic approaches. Especially, Random Forest (RF) With the ACO Algorithm provides the overall best accuracy of 99.40%. Random Forest machine algorithm with ACO creates more impact for maize leaf disease detection. This paper can be further extended by sending leaf disease messages to the farmers by using the wireless network. They can also suggest which pesticide they can use to stop the plant disease.

References

[1] Ganatra, N., & Patel, A. (2020). Deep Learning Methods and Applications for Precision Agriculture. *Machine Learning for Predictive Analysis,* 515-527.

[2] Sun, J., Yang, Y., Xiaofei, H., & Xiaohong, W. (2020). Northern Maize Leaf Blight Detection Under Complex Field Environment Based on Deep Learning. *IEEE Access*, 8, 33679 – 33688.

[3] Haggag, M., Abdelhay, S., Mecheter, A., Gowid, S., Musharavati, F., & Ghani, S. (2019). An intelligent hybrid experimental-based deep learning algorithm for tomato-sorting controllers. *IEEE Access*, 7, 106890–106898.

[4] Ye, X., & Zhu, Q. (2019). Class-incremental learning based on feature extraction of CNN with optimized softmax and one-class classifiers. *IEEE Access*, 7, 42024–42031.

[5] Jiang, P., Chen, Y., Liu, B., He, D., & Liang, C. (2019). Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access*, 7, 59069–59080.

[6] Revathi, & Hemalatha, M. (2018). Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques. *IEEE*, 169-173.

[7] Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture*, 145, 311–318.

[8] Khan, M. A., Akram, T., Sharif, M., Awais, M., Javed, K., Ali, H., & Saba, T. (2018). CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features. *Computers and Electronics in Agriculture*, 155, 220–236.

[9] Al-Hiary, H., Bani-Ahmad, S., Reyalat, M., Braik, M., & ALRahamneh, Z. (2017). Fast and Accurate Detection and Classification of Plant Diseases. *IJCA*, 17(1), 31-38.

[10] Sun, J., Tan, W. J., & Mao, H. P. (2017). Identification of plant leaf diseases based on improved convolutional neural network. *Transactions of the Chinese Society of Agricultural Engineering*, 33, 151–162.

[11] Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A. Y., Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture*, 155, 257-282.

[12] Bai, X., Cao, Z., Zhao, L., Zhang, J., Lv, C., Li, C., & Xie, J. (2018). Rice heading stage automatic observation by multi-classifier cascade-based rice spike detection method. *Agricultural and Forest Meteorology*, 259, 260–270.

[13] Patricio, D. I., & Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review, *Computers and Electronics in Agriculture*, 153, 69–81.

[14] Lin, Z., Mu, S., Shi, A., Pang, C., & Sun, X. (2018). A novel method of maize leaf disease image identification based on a multichannel convolutional neural network, *Transactions of the ASABE (American Society of Agricultural and Biological Engineers)*, 61(5), 1461–1474.