International Journal of Basic and Applied Sciences, 14 (5) (2025) 327-334

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/jk48hn96
Research paper

A Journey on The Exploration of Village Plant Dataset Using Machine Learning Models

K. Sujatha [1] *, Ganesh Khekare [2], Divya Midhunchakkaravarthy [3], M. Malathi [4], V. Srividhya [5], N.P.G. Bhavani [6]

[1] Department of Computer Science Engineering, Lincoln University College, Petaling Jaya, Malaysia
[2] School of Computer Science & Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India.
[3] Centre of Postgraduate Studies, Lincoln University College, Petaling Jaya, Malaysia.
[4] Department of ECE, Rajalakshmi Institute of Technology, Chennai, India
[5] Department of EEE, Meenakshi College of Engineering, Chennai, Tamil Nadu, India
[6] Department of ECE, Saveetha Institute of Medical and Technical Sciences, Chennai, India
**Corresponding author E-mail: pdf.Sujatha@lincoln.edu.my*

Received: July 30, 2025, Accepted: September 2, 2025, Published: September 10, 2025

Abstract

This article is coined for investigating the Village Plant dataset. Many researchers worldwide, carrying out their research in the domain of agriculture, are dependent on this open source dataset. A plant is vulnerable to several infirmities during its period of growth. Detection of the plant's ill health and monitoring the environmental parameters is the most challenging task in agriculture. Plant disease epidemic may have a significant effect on crop production, reducing the country's wealth. Early diagnosis of the occurrence of ill health in plants and the remedies are feasible using Artificial Intelligence (AI). Currently, methods like Deep Learning (DL) algorithms, machine vision techniques, and robotics play an important role in monitoring plant diseases and the growth status. This dataset contains multi-fold in-information about the plants. They include the normal and diseased images of plants like Bell Pepper, Tomato, Cucumber, and Potato. An Internet of Things (IoT) based plant data collection and integration system will provide data for this research, which optimizes the feature set through Ant Colony Optimization (ACO) for improving prediction in feature selection using deep learning models like DenseNet, ResNet 50, VGG 19, and Long Short-Term Memory (LSTM) networks, which in turn enhances plant productivity with advances in AI-driven agricultural diagnostics for plant stress prediction.

Keywords: *Plant Stress; Village Plant Dataset; Internet of Things; Ant Colony Optimization; Densenet; Resnet 50; VGG 19 and Long Short-Term Memory Networks.*

1. Introduction

A considerable section of the people in every country, particularly in the countryside, relies chiefly on agriculture as a source of revenue [1]. Plant ill-health is the prime factor that hinders plant growth. Because of this obstacle, crop yield and quality appreciably decrease, making an important contribution to crop cultivation. Consequently, early diagnosis and remedy for the plant's ill-health can conserve the entire crop. Agriculture has become more versatile and specific over the ancient times, with an attention on escalating the crop yield, improving the efficacy, and avoiding ecological outbreak. Universally, all the nations in the world create an enormous variety of crops based on their environment. These crops are then exported to different countries as per their demands. Therefore, it is essential to produce crops with the highest possible yield. Leaf, stem, root, and fruit are the most probable parts of the plant that are prone to diseases and multiply through them. The indications, reasons, and features for a particular plant disease are varied, including the atmospheric conditions based on the weather and locality. Nearly 20% to 40% of the world's crops are damaged due to the plant's ill-health, which is left unmonitored [2]. The main factor to be taken care of is to reduce the spread rate of the illness, to maximize the yield, which in turn increases the requirement of food for the nation's growing population. Hence, the most significant issue is that early plant disease detection and prevention of diseases can save the crop yield.

It is noted that the agriculturalists in the underdeveloped nations used a vision-based manual approach for plant disease identification [2], which is a delayed approach, and meanwhile, it was not possible to save the lives of the plants. Pathogens, like fungi, bacteria, viruses, protozoa, parasites, and insects, are the major disease-causing agents in plants. The growth of these organisms is favoured due to impulsive weather changes [3]. To increase the speed of the system, automated methods have been introduced to quickly and exactly recognize and analyse the plant diseases. But this is tremendously hard, expensive, and inaccurate. Methods like image processing and machine vision approaches are used for the detection and categorization of a plant's ill-health so that the crop productivity is increased. In the

Copyright ©K. Sujatha et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

earlier period, analysis is carried out using image processing and deep learning algorithms on a variety of plants. Initially, the study emphasizes collecting and integrating the plant data leveraging the Internet of Things (IoT).

2. Background

Deep Learning, a subset of artificial intelligence (AI), has made an unprecedented transformation in agriculture by providing tools for automated image analysis. One of the most popular deep learning architectures used for image classification and disease detection is Convolutional Neural Networks (CNNs). CNNs are very useful for automatically learning hierarchical features from images, which makes them well-suited for imaging [6]–[8]. Given the numerous CNN models that exist, VGG 19, ResNet 50, and DenseNet are some of the most promising in detecting cancer. VGG 19 is a deeper model and is able to capture finer details quickly, but it may be computationally expensive. ResNet 50 uses well-known residual connections to address the vanishing gradient problem and improve the training of deeper networks, resulting in better performance on feature extraction and classification. Dense connectivity improves feature reuse and gradient flow, thus increasing image classification accuracy. All the researchers have demonstrated a remarkable ability to enhance the diagnostic performance for plant disease detection, but many issues still need to be solved, and the barriers that are related mainly to overfitting or computational complexity must be overcome [9], [10].

The Ant Colony Optimization (ACO) method is a metaheuristic algorithm with which the foraging behaviour of ants has been simulated. Feature selection in machine learning is a type of optimization problem where ACO has been successfully applied. To reduce data dimensionality, improve model performance, and minimize computational expenses, feature selection is one of the significant steps. The ACO aims to mimic how ants search for the shortest path between food sources and their nests while automating which features are most important in a dataset [11]–[13]. This feature selection algorithm works the best when one weights and rates features appropriately, then ants also converge on a better subset of pheromones considering such feature factorization. Several previous studies have shown the utility of ACO in improving machine learning models in different applications, including some related to agriculture. ACO is proposed to optimize the selection of features from the leaf images, so that there is a potential impact on enhancing feature selection approaches for more accurate classification models with high performance [14], [15].

Combining ACO and deep learning models provides a promising path toward improving model performance, more specifically in plant disease detection. Some earlier researchers have investigated the synergy of ACO and deep learning frameworks since they can provide better outcomes than using a single approach. ACO improves the selection of features through which the relevant attributes from complex data sets are selected, and these are sent to deep learning models for training as well as prediction. By integrating dotted lesions into the dataset (label), these networks increase accuracy in detecting subtle patterns of diseases from the leaf surface. The features of the leaf images are optimized by ACO in VGG 19, ResNet50, and DenseNet at the feature level to improve their performance on plant disease detection. Many studies have shown that improving diagnostic accuracy and reducing the risk of overfitting this approach was beneficial by improving model generalization capabilities[16], [17].

LSTM (Long Short-Term Memory) networks are a class of RNNs designed to recognize sequential data from hyperspectral images. Since LSTMs can capture dependencies across time and have a powerful ability to predict sequence outcomes, they help model plant diseases based on the history of the open-source datasets. LSTMs are a valuable addition to deep learning models, providing another perspective when interpreting hyperspectral data, such as the leaf colour variation, pesticide treatment history, and its response[18]–[20]. Combining CNN models with LSTMs allows a holistic analysis of the convolutional features learned from imaging and analysis of hyperspectral patterns on a large scale. This combination of data sources significantly improves the diagnostic accuracy globally and provides complete information about the plant health status. There has been much research on the efficiency of using LSTMs for agricultural applications, including plant disease detection, which has been shown to increase prediction and classification performances.

Although new techniques of integrating ACO, deep learning, and LSTM networks for plant disease diagnosis have been beneficial in yielding better results, the current state-of-the-art research has several gaps. Although integrating these techniques has shown more excellent diagnostic as well as methodological accuracy, challenges remain with respect to how best they can be applied [21], [22]. Failures related to problems such as feature selection, model complexity balancing, and integration of different data sources should be systematically analyzed. Furthermore, more inclusive studies to investigate integrated approaches for different types and stages of plant diseases are also required. Frequent research in this area will help to fill these gaps, and the results can provide potential pathways for integrating swarm intelligence with deep learning for more qualitative diagnostics for plant disease detection.

3. Data Collection and Fusion of Plant Diseases Based on IoT

The swift evolution of IoT technology offers robust backing for achieving this objective. An effective and real-time system for collecting and integrating plant-related data can be established by amalgamating diverse sensors, intelligent devices, and data processing systems. This system enables comprehensive monitoring and analysis of plant disease and plant growth-related conditions. The IoT-based plant data collection and integration system primarily comprises three components: the perception layer, network layer, and application layer. The IoT architecture is depicted in Figure 1. The perception layer primarily gathers pertinent data of the IoT system, employing various data acquisition and sensing technologies to facilitate dynamic connectivity and information gathering. The network layer is mainly a variety of communication systems, which is the key part of the IoT and provides a more efficient communication protocol to realize the IoT. The application layer is mainly the interface between the IoT and users. It can obtain corresponding data according to different processes, classify, screen, and process the data according to the need, and finally present the results to users. The realization of this part mainly depends on various databases and professional software.

The hardware components within the perception layer serve as the foundational elements of the system, encompassing a range of sensors and intelligent devices. The wearable devices related to plant stress are a temperature sensor, a moisture sensor, a pH sensor, a light intensity sensor, and a leaf sensor to gather real-time data on the plant stress and plant diseases. Additionally, vision sensors can be installed to monitor the plant disease conditions. In the network layer, the data transmission module conveys the data collected by the hardware layer to the data processing module. This is typically accomplished through low-power wide-area network technologies (e.g., LoRa, NB-IoT) or wireless communication technologies like Wi-Fi and Bluetooth. The data processing module is the heart of the system, responsible for receiving, cleaning, storing, and analyzing the data originating from the hardware layer. The data cleaning process includes removing invalid data and correcting erroneous data to ensure the accuracy and consistency of the data. Data storage relies on a high-performance database system for fast data retrieval and analysis. Data analysis involves advanced mining and machine learning technologies to extract valuable information and insights. The application layer is the interface between the system and users, including data visualization tools, data analysis report generation tools, training guidance systems, etc. Through the application layer, the coaching team can intuitively check the real-time data, historical data, and data analysis results of players to formulate more scientific training plans and tactical strategies.

Data mining technology in the IoT system mainly depends on four basic technologies: database, AI, mathematical statistics, and visualization. The algorithm input of data mining technology is a database, the algorithm output is knowledge or pattern extraction and discovery, and the algorithm processing is the concrete design of the search method. The description or explanation of algorithm design is mainly divided into input, output, and processing. Data mining algorithm mainly involves three aspects: mining objects, mining tasks, and mining methods.

The methodologies employed in data mining technology can be categorized into four groups: statistical methods, machine learning approaches, neural network techniques, and database methods. Statistical methods can be further subdivided into regression analysis and discriminant analysis. Data mining technology encompasses many targets, including relational, spatial, and text databases. Neural network techniques can be classified as feedforward or self-organizing neural networks. Machine learning within data mining technology primarily refers to the genetic algorithm. Data mining technology primarily relies on multidimensional data analysis. Data mining is a lengthy process with specific steps, primarily from knowledge discovery. Its primary characteristic lies in its ability to extract, transform, analyze, and model large datasets, thereby isolating key data.

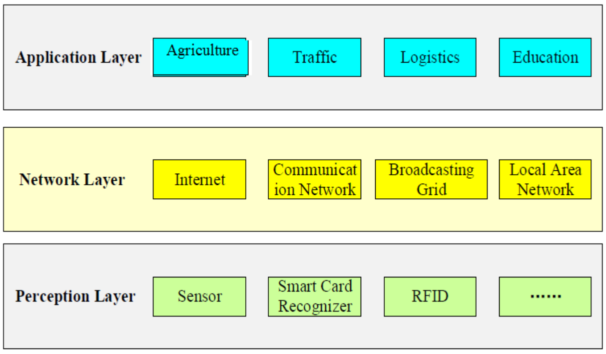

Fig. 1: Pictorial Representation of IoT Architecture.

For healthy potato plants, a typical chlorophyll level is often measured using a SPAD meter, with readings considered optimal between 35 to 45 SPAD units, which correlates closely with a good nitrogen status in the leaves, indicating that a healthy potato plant will have a relatively high chlorophyll content due to its dependence on nitrogen for chlorophyll synthesis; however, the exact values can vary based on potato variety and growing conditions.

The leaf images of the potato plant for healthy condition, Early blight, and Late blight diseases, with the corresponding variation in chlorophyll content pertaining to three stages (Low Level-LL, Medium Level-ML, and High Level-HL), are taken from the kaggle database and listed in Tables 1,2, and 3.

Table 1: Various Stages of Healthy Potato Plants from the Kaggle database

S. No	Name of the Plant	Image of the leaves	Chlorophyll	Status
1.	Potato		Low Level (LL)	Non-diseased/healthy
2.	Potato		Low Level (LL)	Non-diseased/healthy

Table 2: Various Stages of Late Blight in Potato Plants from Kaggle Database

S. No	Name of the Plant	Image of the leaves	Chlorophyll	Status
1.	Potato		Low Level (LL)	Diseased/Late Blight
2.	Potato		Low Level (LL)	Diseased/Late Blight

Table 3: Various Stages of Early Blight in Potato Plants from Kaggle Database

S. No	Name of the Plant	Image of the leaves	Chlorophyll	Status
1.	Potato		Low Level (LL)	Diseased/Early Blight
2.	Potato		Low Level (LL)	Diseased/Early Blight

4. Materials and Methods

4.1. Deep Learning Models

In this research, deep learning models are central to the task of diagnosing plant disease, mainly through the analysis of plant stress and plant diseases. The primary models employed include Convolutional Neural Networks (CNNs), such as ResNet 50, DenseNet, and VGG 19, alongside Long Short-Term Memory (LSTM) networks, which are used for analysing sequential data, including the plant disease data and its history. Each of these models brings unique strengths to the research, enabling precise and robust predictions that significantly contribute to early diagnosis and effective treatment planning.

The CNN models worked well on the leaf images from the Kaggle data set (https://www.kaggle.com/datasets/emmarex/plantdisease) gathered from online sources. The images, which portray different plant diseases, are employed to teach the CNN models for determining if the illness is deficient and the corresponding stage. Like backpropagation, CNNs are designed to automatically and adaptively learn spatial hierarchies of features through the process called convolution. Types of Layers in CNN: The initial layers are the Convolutional layer, the Pooling Layer, and the Fully connected layers. The convolutional layers apply a set of filters to the input image to create feature maps, which learn various features such as edges, textures, and essential details in an image. The feature maps are then downsampled using pooling layers that reduce the spatial dimensions but keep only the necessary information. The last layers are fully connected, meaning they deal with feature maps and classify the image into one of several labels.

ResNet 50 is one of the principal architectures in this research. It is designed to allow intense networks from the train and not suffer from vanishing gradients associated with deeply placed weights in a backpropagation algorithm. The major innovation in ResNet 50 is the introduction of residual blocks, which permits fitting for adding identified with input and fitting out of unrestricted functions. Mathematically, the expression of the function is given in Equation 1.

$$y = F(x, \{Wi\}) + x \tag{1}$$

Where x is the input to the residual block, $F(x, \{Wi\})$ denotes the item to be learnt as residual mapping, and y represents the output. Where the output x of one or more layers is directly added back to it, together with a special skip connection called an identity function, which retains information from the previous layer and enables easy training thanks to Backward Propagation, ensuring that gradients remain intact, the deep architecture of ResNet 50 in this research works well to capture complex patterns available on the images and helps identify the affected regions accurately.

Another model used in this research is DenseNet, which connects each layer using a feed-forward method for the vanishing gradient problem. Unlike those above architectures, which only take the last output as input, each layer in a DenseNet will get its feed-forward from all layers before it and provide features to all subsequent layers. Symbolically, this dense connectivity pattern is written as in Equation 2.

$$x\ell = H\ell([x0, x1, \ldots, x\ell - 1]) \tag{2}$$

In the above definitions, $x\ell$ are layer $l\ell^{th}$ outputs, and $H\ell$ denotes a series of operations that resembles batch normalization followed by ReLU before convolution, etc., while $[x_0, x_1 \ldots, x_{(xl-1)}]$ is carrying concatenation over feature maps from layers 0 to $\ell-1$. This rich interconnectivity not only facilitates the passage of information and gradients throughout the network but also results in more effective use of parameters. Therefore, DenseNet has the potential for high accuracy with fewer parameters, which is well-suited for deployment tasks such as image analysis (the precise structures on images are essential due to differential diagnosis).

Another CNN model, VGG 19 (a variant of the network), is used in this research. VGG 19 — simple and consequently more effective, this network has only tiny (3x3) convolutional filters with depths increasing further in the early sections of its architecture. With a straightforward architecture of VGG (each convolution layer followed by ReLU and max-pooling), it excels at acquiring high-level features in images. The final Convolutional Layer output is flattened and runs through several Fully Connected layers before classifying the input. When dealing with leaf images, subtle changes in the texture of tissue or border characteristics can be very important to discriminate different stages of plant diseases, so having longer deep models like VGG 19 helps us extract better representation from our input data.

This research uses CNNs as well as Long Short-Term Memory (LSTM) networks to process sequential data like health history records. LSTMs are an RNN model trained to learn sequence dependencies. The LSTM network is the heart of this algorithm, called the memory cell, which allows the recording of information across multiple periods. LSTM uses gates (input gate, forget gate, and output gate) to control the information flow into and out of a memory cell, enabling the network to remember or discard previous states. The equations for each of these gates are represented in Equations 3-7.

$$ft = \sigma(Wf \cdot [ht - 1, xt] + bf) \tag{3}$$

$$\mathrm{it} = \sigma(\mathrm{Wi} \cdot [\mathrm{ht}-1, \mathrm{xt}] + \mathrm{bi}) \quad (4)$$

$$\mathrm{ot} = \sigma(\mathrm{Wo} \cdot [\mathrm{ht}-1, \mathrm{xt}] + \mathrm{bo}) \quad (5)$$

$$\mathrm{Ct} = \mathrm{ft} * \mathrm{Ct} - 1 + \mathrm{it} * \tanh(\mathrm{WC} \cdot [\mathrm{ht}-1, \mathrm{xt}] + \mathrm{bC}) \quad (6)$$

$$\mathrm{ht} = \mathrm{ot} * \tanh(\mathrm{Ct}) \quad (7)$$

In Equation 6, 'ft' represents the forget gate, input gate, and output gates, respectively, where itis the cell state and ht is LSTM's hidden states at step, respectively, with σ(t) denoting the sigmoid function and tanh(x) representing the hyperbolic tangent. These gates allow the LSTM to learn which bits of information. By joining the LSM model with the CNN models, they make use of both spatial feature extractions provided by a multi-channel convolutional layer and temporal predictions accomplished via long-term, short-term memory networks, generating better data analysis.

These deep learning models are then integrated with an Ant Colony Optimization (ACO) based feature selection technique to improve their diagnostic and prognostic accuracy of plant disease. This ensures that deep learning models' predictive capacity, which is enhanced as they avoid being confused by redundant and irrelevant images and agricultural records.

4.2. Ant Colony Optimization

This study uses a heuristic algorithm known as Ant Colony Optimization (ACO) to solve this feature selection optimization in image analysis, including agricultural history records. Within this framework, ACO is utilized to improve the performance of deep learning models by feature selection, further refining diagnosis and prognosis accuracy for the specific disease type, such as plant disease in our study.

ACO is based on the actual ant colony foraging behavior, where ants find a path between their colony and food linearly by leaving traces (pheromones) from which others can follow. The goal is to find a subset of the features that best balances between model complexity and predictive accuracy.

It starts with the feature extraction from plant images using deep learning models such as ResNet 50, DenseNet, etc. These are features such as patterns, textures, and other image characteristics that the ACO algorithm takes as input. Attributes present in the images are also extracted and introduced into the ACO procedure.

In ACO, the solution space (for this problem in terms of possibilities of different feature subsets) is explored by a set of artificial ants. This one shows an ant as the potential candidate solution that is possibly selected as a feature. Where the probability p_{ij}^{k} of an ant 'k' to move from the node 'I' (a particular feature) towards the j^{th} one is regulated by a pheromone trail τ_{ij} and heuristic information η_{ij} expressed in Equation 8.

$$p^{K}{}_{ij} = \frac{[\tau_{ij}]^{\alpha}[\eta_{ij}]^{\beta}}{\sum \mathrm{lEN}_{j}^{k}[\tau_{ij}]^{\alpha}[\eta_{ij}]^{\beta}} \quad (8)$$

Where τ_{ij} is the intensity of pheromone on the path from feature 'I' to 'j' and η_{ij} is heuristic information, often related to the importance or relevance degree that a given characteristic has in the classification task. The importance of the pheromone over that of heuristic information is determined by parameters α and β, respectively.

While ants test various feature subsets, they rank them according to how well these would perform in the classification of CT scan images or disease progression prediction by training the deep models on images and studying their generalization capabilities. The accuracy of these models is used in updating pheromone trails. Pheromone update rule denoted in Equation 9.

$$\tau_{ij} \leftarrow (1-\rho) \cdot \tau_{ij} + \sum_{k=1}^{m} \Delta\tau_{ij}{}^{k} \quad (9)$$

Where the ρ is called as evaporation, which helps to prevent a rapid convergence of algorithm on a suboptimal solution by decreasing pheromone intensity with every time, this is denoted by $\Delta\tau_{ij}^{k}$, which indicates the total pheromone deposited on that particular path between feature 'i' and 'j' of ant kg, which is dependent with respect to performance(accuracy) achieved through the Deep Learning model trained upon the subset selected.

ACO progressively updates the selection of features by exploring new paths (feature subsets) and updating pheromone deposits based on the path quality scores achieved. The algorithm repeats this process until it reaches an optimal feature subset that can best predict the deep learning models.

The best feature subset identified by ACO in this research is then applied to the CNN model (ResNet 50, DenseNet, VGG 19) for plant image analysis and LSTM for analysis. ACO dramatically improves the performance of these models by emphasizing only the most pertinent features, resulting in a better understanding of plant disease stages from the images and more science-based predictions for disease progression.

4.3. Preprocessing of The Dataset

Dataset preprocessing is an essential factor in this research, through which clean and standard input data are provided to the deep learning models for accurate diagnosis and prognosis of plant diseases. The datasets used in this work include leaf images from different stages, detailed as normal/non-diseased and diseased/infected leaves with sign indicators, lab results, and background information. The image data and textual records both have to go through multiple preprocessing steps in order for them to be effectively trained on and analyzed by deep learning models.

The raw data are gathered from online resources and present different resolution levels as well as contrast/noise. Many preprocessing steps are applied to standardize these images. Initially, all images are resized to a specific size because of the entering proportions for Convolutional Neural Networks (CNNs) like ResNet 50, DenseNet, and VGG19. This is an essential step because CNNs require that the input images be a fixed size being passed to it. The images are usually rescaled to 224x224 pixels, which is a standard input size for many deep-learning models.

After that, a couple of contrast enhancement techniques are applied to enhance the visibility of important structures in the images (like boundaries or abnormal changes in leaf colour). There are various techniques for improving contrast in an image, and histogram equalization is one method of doing so. It spreads the intensities of the pixels across different levels, thus accentuating features otherwise difficult to see. Even in imaging the plant-related data, it is essential to consider slight variations of colour changes on the surface of the leaf, as they can imply the presence of disease.

Another crucial preprocessing step is noise reduction since the plant images captured are highly prone to presenting noise that can blur out significant visual information. The images are then generally smoothed using different filters, like Gaussian or median filtering, to remove the noise. This step is to identify those artefacts to minimize the interference with the learning process of relevant features from images.

After enhancement and denoising of the images, they are normalized. This step is crucial because it helps to standardize the pixel intensity values at a higher level of 0 and n-1 (generally between 0–255), which eventually leads to more stable deep learning models. This ensures that the input features are on a similar scale, which, in turn, can help speed up the convergence of our model during training.

Also, the plant disease data, along with the indicators, lab results, and the previous plant-related data, are cleared up to be input within a Long Short-Term Memory (LSTM) network. Several steps should be performed in order for these textual and numerical files to act as input and pre-processing.

Missing values are often present in data, so first of all, we clean the data to manage these missing values. It uses several imputation techniques, which depend on the mechanism of missing data. Then, the mean of the data is used to fill in missing numerical values, and the most frequent value among categories for categorical data, or a 0/1 Placeholder Value. This step ensures that the data needed for a complete dataset has been satisfied and that the LSTM network can be trained on all relevant plant information. Then, the categories in plant diseases are converted into numeric shapes via one-hot encoding, such as being or not having a particular symptom. This is required because deep learning models are based on numerical data, and converting the categorical data into such a format helps provide an efficient way to represent the variable.

Also, the numerical data in a record is brought on feature scaling and normalised in a range similar to the image data, the numerical values. This is of particular importance in the context of LSTMs because it will prevent any one feature from having a higher weighted sum than others just because they have larger values; therefore, each feature (assuming same-scale numerical data) becomes equally relevant to model prediction.

Finally, structuring the data in a temporal sequence allows the retrieval of the sequential nature of plant data. Some information is time-dependent, such as lab results or symptoms evolving. It requires pre-postprocessing of the data to frame it over time and map dependencies for an LSTM sequence network, which is essential in predicting the disease progression as well as the outcome.

5. Results and Discussion

All models were trained and rigorously tested to objectively evaluate their accuracy at predicting plant disease from images of the leaves and the relevant information. The results are shown in Figure 2. After evaluation, ACO+LSTM with DenseNet was the most accurate among all, obtaining an accuracy of 97.876%. The high precision may be explained by the advantage of DenseNet over the residual network to capture detailed features in leaf images and its further optimization through ACO, as well as complementary analysis from plant history with LSTM, which leads to identifying the plant diseases. The ResNet 50 model paired with LSTM and optimized using ACO also resulted in a vital accuracy of 95.6%. While DenseNet -Agro achieved the best result among all others, ResNet 50 has yet to be shown to be a powerful feature extractor and obtained almost similar performance as its counterpart.

The VGG 19 model, which was equipped with ACO and LSTM, achieved an accuracy of 90.23%. Even though VGG 19 is very good, it provides evidence that the model might only be able to capture some aspects of variance in data, as well as DenseNet and ResNet 50. These results together provide pieces of evidence that the DenseNet model performs better in all cases than other models and is, therefore, the best option for predicting the various stages of plant diseases.

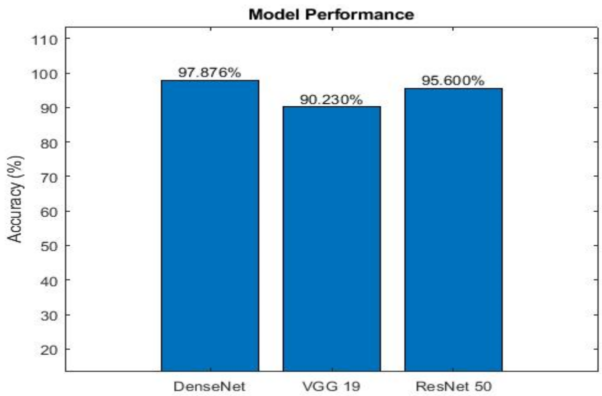

Fig.2: Accuracy of Each Model.

The performance score shown in Figure 3 and Table 5 compares three deep learning models (DenseNet, ResNet 50, and VGG19) with Ant Colony Optimization and the Long Short-Term Memory network in terms of predicting plant leaf diseases, along with the plant growth parameter records. The DenseNet model, achieving a precision and recall of 0.98, exhibits a balanced and highly accurate performance, as reflected in its F1-score of 0.98. The AUC-ROC value of 0.99 further highlights DenseNet's exceptional capability in distinguishing between diseased and healthy plant leaves, making it the most reliable model in this study.

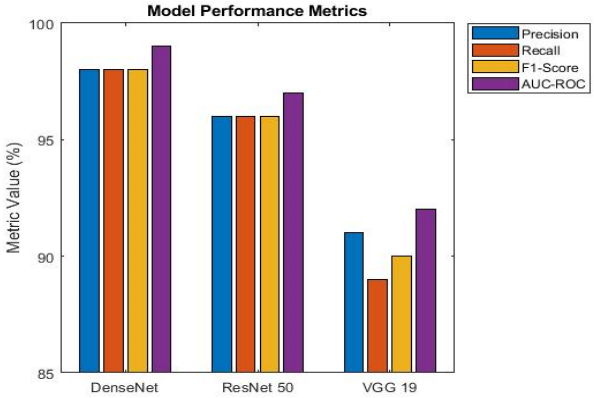

Fig: 3: Performance Score of Each Model.

ResNet 50, with a precision, recall, and F1 score of 0.96, is also predictive but slightly behind DenseNet. It has an AUC-ROC value of 0.97, implying it is still an effective model for plant leaf disease diagnosis, but not among the best-performing models available today. VGG 19 model shows slightly lower performance of precision=0.91, recall=0.89, and F1-score=0.90. The AUC-ROC value of 0.92 is considered to be good but not exceptional discriminative ability. Conclusively, Table 5 depicts that DenseNet gives the best result with better reliability and accuracy performance over all these models for this application.

Table5: Classification Methods and Performance Measures

S.No	Models used	Accuracy (%)	Precision(%)	Recall(%)	F1-score(%)	AUC-ROC(%)
1.	DenseNet	97	97	97	97	97
2.	ResNet 50	96.15	96	96	96	97
3.	VGG 19	90.23	91	89	90	92
4.	ACO+LSTM with DenseNet	98.5	98.3	98.3	98.3	98
5.	ACO+LSTM with ResNet 50	96.2	97	97	97	97
6.	ACO+LSTM with VGG 19	93	93.4	93.2	94	94

This indicates that the proposed ACO-based DenseNet with LSTM exceeds all other models with an accuracy of 98.5%, which suggests the great potential of its ability in the management of complex medical imaging and temporal data. ResNet 50, another model optimized by ACO, has an accuracy of 96.2%, which reassures robustness and efficiency. However, although VGG 19 is showing marked increases in subsequent training epochs, its accuracy is 90.23%, which is less compared to other models.

6. Conclusion

In this research, Ant Colony Optimization (ACO) anomaly-based learning framework integration with improved deep models, including DenseNet, ResNet 50, and VGG19, along with LSTM Networks, can predict the stage of the Blight disease with the chlorophyll content from the leaf images for a potato plant more accurately. Given the strong performance of an ACO in this research, feature selection is crucial for deep learning models by capturing essential features that are helpful in disease classification. The validation of this approach by swarm-based deep learning in medical fields suggests that similar investigations can be conducted with other disease diagnoses and prognoses. The results highlight the significance of integrating robust optimization algorithms with cutting-edge deep learning architectures for fostering enhanced accuracy, reliability, and generalizability. In the future, AI-mediated healthcare solutions will be responsive to enriching patient outcomes and streamlining medical proceedings.

Funding Information

The authors state no funding is involved.

Conflict of Interest Statement

The authors state no conflict of interest.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Author 1	✓	✓				✓			✓					
Author 2					✓									
Author 3					✓									
Author 4									✓	✓	✓			
Author 5									✓	✓	✓			

C: Conceptualization	I: Investigation	Vi: Visualization
M: Methodology	R: Resources	Su: Supervision
So: Software	D: Data Curation	P: Project administration
Va : Validation	O: Writing - Original Draft	Fu: Funding acquisition
Fo: Formal analysis	E: Writing - Review &Editing	

References

[1] Rei Sonobe, Yuhei Hirono and Ayako Oi, “Non-Destructive Detection of Tea Leaf Chlorophyll Content Using Hyperspectral Reflectance and Machine Learning Algorithms”, *Plants,* 9, 368;2020. https://doi.org/10.3390/plants9030368.

[2] Kowshik Kumar Saha, Cornelia Weltzien, Bodo Bookhagen, Manuela Zude-Sasse, “Chlorophyll content estimation and ripeness detection in tomato fruit based on NDVI from dual wavelength LiDAR point cloud data”, *Journal of Food Engineering,* 383, 112218, 2024. https://doi.org/10.1016/j.jfoodeng.2024.112218.

[3] Amar Kumar Deya, Manisha Sharmaa, M.R.Meshram, “An Analysis of Leaf Chlorophyll Measurement Method using Chlorophyll Meter and Image Processing Technique”, *International Conference on Computational Modeling and Security, Procedia Computer Science,* 85,pp: 286 – 292,2016. https://doi.org/10.1016/j.procs.2016.05.235.

[4] R. Kavitha Lakshmi, Nickolas Savarimuthu, “A Novel Transfer Learning Ensemble based Deep Neural Network for Plant Disease Detection”,International Conference on Computational Performance Evaluation, *IEEE Xplore*,2021. https://doi.org/10.1109/ComPE53109.2021.9751910.

[5] Kishor Chandra Kandpaland Amit Kumar, “Migrating from Invasive to Non-invasive Techniques for Enhanced Leaf Chlorophyll Content Estimations Efficiency”,*Critical Reviews in Analytical Chemistry*,pp: 2583-2598, 2023. https://doi.org/10.1080/10408347.2023.2188425.

[6] Utkarsha N. Fulari, Rajveer K. Shastri, Anuj N. Fulari, “Leaf Disease Detection Using Machine Learning”, *Journal of Seybold Report* , VOLUME 15, ISSUE 9, Page: 1828, 2020. https://www.researchgate.net/publication/344282301.

[7] Lukas Wiku Kuswidiyanto, Hyun-Ho Noh and Xiongzhe Han, “Plant Disease Diagnosis Using Deep Learning Based on Aerial Hyperspectral Images: A Review”, *Remote Sensing*, 14, 6031, 2022, https://doi.org/10.3390/rs14236031.

[8] Ziran Ye, Xiangfeng Tan, Mengdi Dai, Xuting Chen, Yuanxiang Zhong, Yi Zhang, Yunjie Ruan and Dedong Kong, “A hyperspectral deep learning attention model for predicting lettuce chlorophyll content”,*Plant Methods,* pp: *20:22,* 2024.https://doi.org/10.1186/s13007-024-01148-9..

[9] Seyed Mohamad Javidan , Ahmad Banakar, Kamran Rahnama, Keyvan Asefpour Vakilian, Yiannis Ampatzidis, “Feature engineering to identify plant diseases using image processing and artificial intelligence: A comprehensive review”, *Smart Agricultural Technology*, Volume 8, 100480, 2024. https://doi.org/10.1016/j.atech.2024.100480.

[10] Juan Ignacio Arribas, Gonzalo V. Sánchez-Ferrero , Gonzalo Ruiz-Ruiz, Jaime Gómez-Gil, “Leaf classification in sunflower crops by computer vision and neural networks”, *Computers and Electronics inAgriculture,* Volume 78, Issue 1, pp: 9-18, 2011. https://doi.org/10.1016/j.compag.2011.05.007.

[11] R. N. Strange and P. R. Scott, "Plant disease: a threat to global food security",*Annu. Rev. Phytopathol*, vol. 43, pp. 83-116, 2005. Vol. 43:83-116https://doi.org/10.1146/annurev.phyto.43.113004.133839.

[12] J. G. A. Barbedo, "Digital image processing techniques for detecting quantifying and classifying plant diseases",SpringerPlus, vol. 2, no. 1, pp. 1-12, 2013. https://doi.org/10.1186/2193-1801-2-660.

[13] A. Macedo-Cruz, G. Pajares, M. Santos and I. Villegas-Romero, "Digital image sensor-based assessment of the status of oat (avena sativa l.) crops after frost damage", *Sensors*, vol. 11, no. 6, pp. 6015-6036, 2011. https://doi.org/10.3390/s110606015.

[14] C. Bock, G. Poole, P. Parker and T. Gottwald, "Plant disease severity estimated visually by digital photography and image analysis and by hyper-spectral imaging", *Critical reviews in plant sciences*, vol. 29, no. 2, pp. 59-107, 2010. https://doi.org/10.1080/07352681003617285.

[15] K. G. Liakos, P. Busato, D. Moshou, S. Pearson and D. Bochtis, "Machine learning in agriculture: A review", *Sensors,* vol. 18, no. 8, pp. 2674, 2018. https://doi.org/10.1080/07352681003617285.

[16] T. Rumpf, A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne and L. Plümer, "Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance",*Computers and electronics in agriculture*, vol. 74, no. 1, pp. 91-99, 2010. https://doi.org/10.1016/j.compag.2010.06.009.

[17] X. E. Pantazi, D. Moshou, A. A. Tamouridou and S. Kasderidis, "Leaf disease recognition in vine plants based on local binary patterns and one class support vector machines", *IFIP International Conference on Artificial Intelligence Applications and Innovations*, pp. 319-327, 2016. https://doi.org/10.1007/978-3-319-44944-9_27.

[18] A. Johannes, A. Picon, A. Alvarez-Gila, J. Echazarra, S. Rodriguez-Vaamonde, A. D. Navajas, et al., "Automatic plant disease diagnosis using mobile capture devices applied on a wheat use case", *Computers and electronics in agriculture*, vol. 138, pp. 200-209, 2017. https://doi.org/10.1016/j.compag.2017.04.013.

[19] PoornimaSingh Thakur, Pritee Khanna, Tanuja Sheorey, Aparajita Ojha, “Trends in vision-based machine learning techniques for plant disease identification: A systematic review”, *Expert Systems with Applications,* Volume 208,118117, 2022. https://doi.org/10.1016/j.eswa.2022.118117.

[20] M. H. Saleem, J. Potgieter and K. M. Arif, "Plant disease detection and classification by deep learning", *Plants,* vol. 8, no. 11, pp. 468, 2019. https://doi.org/10.3390/plants8110468.

[21] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis", *Computers and Electronics in Agriculture*, vol. 145, pp. 311-318, 2018. https://doi.org/10.1016/j.compag.2018.01.009.

[22] J. Boulent, S. Foucher, J. Théau and P.-L. St-Charles, "Convolutional neural networks for the automatic identification of plant diseases", *Frontiers in plant science*, vol. 10, pp. 941, 2019. https://doi.org/10.1016/j.compag.2018.01.009.