05-SRJ-464-148

SIAZGA RESEARCH JOURNAL, 2024
Vol. 03 | No. 04 | December - 2024 | 44 – 55
DOI: https://doi.org/10.5281/zenodo.17440338

Original Article

Hybrid Machine Learning Model for Early Detection of Cucumber Leaf Curl Disease in Smart Agriculture

Imadullah[1], Ibadullah[1*], Shahzad Khan[1], Sadaf Mustafa[2] & Qaisar Aziz[1]

[1] *Department of Computing, Riphah International University Islamabad, Pakistan*
[2] *Harbin Institute of Technology, Shenzhen Campus, China*

Abstract

Efficient Accurate and timely detection of plant diseases plays a crucial role in maintaining crop health and ensuring sustainable agricultural productivity. Manual identification of leaf-based diseases is often labor-intensive, time-consuming, and subject to human inconsistency. To overcome these limitations, this study introduces a hybrid machine learning-based framework for the early detection of Cucumber Leaf Curl Disease (CLCuD) within a smart farming environment. The proposed system integrates Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Decision Tree (DT) classifiers as base learners to construct a robust ensemble model capable of high-precision classification. A real-time cucumber leaf image dataset was used for model training and evaluation. Experimental outcomes demonstrate that individual classifiers achieved accuracies of 96.37% (SVM), 97.00% (KNN), and 98.00% (DT), while the hybrid ensemble attained the highest accuracy of 98.29%. The results confirm that the proposed hybrid model offers superior detection accuracy, reliability, and efficiency compared to individual classifiers, providing a promising tool for early disease identification, yield preservation, and the advancement of precision agriculture through smart and sustainable farming practices.

Keywords: Cucumber Leaf Curl Disease, Disease Detection, Hybrid Machine Learning, IoT-based Agriculture, K-Nearest Neighbor, Precision Agriculture, Smart Farming, Support Vector Machine

How to cite:
Imadullah, Ibadullah, Khan, S., Mustafa, S., & Aziz, Q. (2024). Hybrid Machine Learning Model for Early Detection of Cucumber Leaf Curl Disease in Smart Agriculture. *Siazga Research Journal*, *3*(4), 44–55. https://doi.org/10.5281/zenodo.17440338

INTRODUCTION

In developing country, agriculture plays crucial role in the development of the countries and act as a back bone of the economy of the countries and it plays an important and major role for prosperity of the nation [1]. Plant diseases can occur frequently and are one of the most important factors/reason which greatly decrease the quality and yield any crops [2]. Among all productive food producing items, cucumber is one an important item of vegetables, However, cucumber cultivation is highly susceptible to various diseases that significantly diminish crop yield and productivity [3]. These yield-reducing diseases pose serious threats to cucumber production, often leading to substantial financial losses for farmers. Therefore, early and accurate identification of such diseases is essential to implement effective control measures, mitigate their spread, and ultimately enhance cucumber yield while minimizing the associated economic losses.

Early-stage detection and identification of crop disease plays an important roles on crop quality and its production capacity [4][5]. Manual detection of crop diseases is very time consuming and need more expertise in the relevant crop. Some crops disease are mostly visible on leaves and can be used as a parameters for the detection for the disease [6]. To meet the requirement, identification and recognition of different plant disease can be performed by using most advance Internet of Thing (IoT) based system along with machine learning algorithm. These IoT devices can be used to collect real time data and ML are widely used for the classification and identification of the disease types [7].

At present, cucumber curl leaf diseases

*Corresponding Author: Ibadullah | Department of Computing, Riphah International University Islamabad, Pakistan
ibad.ullah@riphah.edu.pk

identification depends on investigation of the cucumber field of the farmers and plant experts to identify the curl leaf diseases based on their experience of identification of leaf types or abnormal leaf structure [8]. However, this types of diagnosis and identification is often more laborious and time-consuming, which may also results in more financial loss. Meanwhile, the identification techniques for disease identification based on biosensors and other equipment's are hard and challenging to be implemented for each farmers due to high cost.

During recent years, IoTs along with ML has been plays crucial role in the early detection of crop disease [5]. In modern era, IoT offer online identification of plant disease during real time. For example, the images of the leaf disease are captured and upload it the servers for the diagnosis of diseases [4]. To this end, it is important to identify the disease timely and accurate way, in addition to the support of IoTs. The agriculture industry has benefited recently from a monitoring and automation system powered by the IoT, and it is anticipated that by 2050, the majority of nations will be utilizing this technology to boost food production and crop yields [9].

As a result of this forecast, IoT-based smart farming has gained popularity for monitoring the conditions in agricultural fields in real-time [10]. The environmental parameters of various crops, such as, temperature and moisture can be tracked using a smart farming paradigm that is IoT enabled. More easily and effectively, cucumber [11]. You can get information from farms about environmental factors like temperature, humidity, soil moisture, and light. Farmers can remotely access this data at their convenience. Additionally, the formers can use modern agriculture techniques to improve the crops quality by promoting most recent scientific cultivation methods, faraway monitoring, and better yield production methods [12].

However, achieving reliable system validation under diverse local environmental conditions remains a significant challenge for IoT-based systems. The concept of IoT has proven to be one of the newest technologies expected to impact all industries and businesses including agriculture and processing farm [13]. Global adoption of IoT technology is increasing, as around 43% of businesses worldwide use some form of IoT application and agriculture is no exception. IoT-based agriculture or smart farming, agricultural vehicle monitoring [14], livestock monitoring [15], warehousing maintenance, crop/plant care, etc. Smart agriculture offers greater productivity, less resources and products, automation, time and production information, farm and field surveys, and many more [16]. Using IoT to enable smart agriculture and farming is very efficient compared to traditional cultivation methods. IoT-based smart agriculture is typically network of smart objects that use different sensors (humidity, temperature and other parameters etc.) to monitor farms and collect some important data. Farmers can follow the status of their fields from anywhere.

The proposed IoT-based architecture consists of three primary layers: the perception layer, the network layer, and the application layer.

- This is the foundational layer of the IoT framework, comprising various sensors and devices responsible for sensing and collecting environmental data from physical objects or field components.
- This layer facilitates the transmission of the collected data through communication networks and the internet, ensuring seamless connectivity between the perception layer and the higher-level application layer.
- Finally, the application layer is responsible for presenting the results to the end users. It also includes some important steps like data processing and applying ML algorithms to classify the data based on their behaviors.

The structure of this paper is organized as follows. The related work section reviews recent developments in smart irrigation and crop disease detection. The objectives and contributions section outlines the main goals and innovations of the study. The proposed IoT-based system section explains the system design and implementation of the real-time test-bed. The experimentation section describes dataset generation and applied machine learning algorithms. Finally, the results and conclusion present the model's performance evaluation and summarize key findings with directions for future research.

RELATED WORK

IoT applications can help farmers by keeping them up to date with crop information and weather to monitor condition of their fields [17]. With the development of IoT and ML, great efforts are made to develop new methods that can speed up the process of identifying diseases. So far, progress has been made in the identification of cucumber leaf curl and other leaf diseases based on classical ML such as SVM, kNN, AdaBoost, and PNN [18] [19]. Working on event-based disease detection in IoT systems, this article develops a new method for cucumber disease detection using a hybrid approach and an IoT-based system.

In open literature, IoT and ML can be widely used to identify different types of disease and this timely identification can greatly improve the production of the crops. For Instance, In their smart farming application, B. Lagun et al [20]. utilize IoT technology and sensor nodes placed in peach orchards to monitor the temperature and humidity. The system is designed to prevent frost damage during cold weather by sending a command to activate stoves, which warm up the orchard and control. The fungal disease prediction system was proposed by Truong et al [21] is an IoT-based system that sends real-time data to cloud server for storage. The system utilizes ML algorithms to predict fungal detection and prevention, enabling crop field managers and farmers to take preventive measures against fungal attacks. This system can forecast the occurrence of two plant diseases, which are critical threats to crops across the globe. With the aid of predictive analytics and real-time data, the IoT system helps to improve the prevention and monitoring of the spread of diseases. However, the proposed system requires specialized devices and applications, which may not be readily available to everyone.

Krishna et al. [22] developed an IoT-based system which aims to detect diseases and autonomously prevent them by spraying and sends alerts via SMS using NodeMCU modules. M. Nawaz et al [23] analyze and develop a plant disease detection system that uses IoT-equipped technology with the capability to observe and identify plant disease. The system incorporates several sensor nodes that collect environmental data which is processed via ML technology in the cloud. The system alerts the users via SMS or email instantaneously if a disease is identified. The advantage of the system is that it is economically viable and considerably efficient in detecting plant diseases and therefore assists farmers in reducing or eliminating crop loss. Throat et al [24] outlined a system that integrates the capabilities of IoT and ML, and offers features such as the identification of leaf diseases, humidity, and temperature, condition monitoring, and the sensing of soil moisture.

Rather than manual verification, sensors configured to detect temperature, humidity, and moisture were placed at various points of the farms. All the sensors are managed under one central controller dubbed Raspberry PI (RPI). The RPI connected camera could diagnose a certain type of leaf disease and also disseminate information about the farm on which the disease is, and other obstacles to crop growth, and sends this information over a WIFI server to the farmers through the RPI.

IoT solutions on rice farming and its integration with AI is presented by Wan et al. [25]. The first solution is a mobile application which allows farmers to diagnose diseases and pests infesting rice plants without any advice from an agronomist. The second solution is a monitoring system that remotely measures different environmental parameters in rice fields, thereby reducing intensive work for farmers. To develop the mobile app, the researchers collected six types of rice plant images for analysis and identification.

Furthermore, the researchers utilized Arduino Nano to control the solenoid valve. Finally, the classification of rice plant diseases was achieved using both Transfer Learning and DL models. [26] Proposed framework for smart crop monitoring involves four levels. The purpose is to examine soil nutrients such as soil moisture, NPK volume, and other factors using an Arduino-based sensor node. An IoT sensor node was hand-built and tested on an agricultural field. After data collection, the resulting data was analyzed to provide informative insights to the farmers for subsequent steps. Machine learning algorithms were employed at the cloud level to analyze the collected data, allowing farmers to monitor the soil nutrient index and manage nutrient requirements for improved crop yield. The system proposed in [27] utilizes an IoT-based cloud architecture, where sensor data are transmitted and stored on the Adafruit IO platform. A soil moisture sensor is integrated to capture real-time moisture levels from the agricultural field. The acquired data is represented in a percentage value ranging from 0 to 100 representing different moisture values. Based on the moisture content, an automatic message is sent to the user to turn on/off the irrigation. The user can control the irrigation process remotely. All these above IoT based solutions are utilized, to detect different types of disease and environmental parameters, which either cause the diseases or there is a chance of causing the disease in the corp.

However, to the best of our knowledge, there is very limited research work for the detection of disease in the cucumber vegetables [28] [29] [30]. These methods used huge amount of data of images generated from IoT based devices. For reader, we recommend the articles [31] and [32] which contains most of the recent research in the direction of disease detection using IoT and ML algorithm, In this approach, we used environmental parameters to detect the presence of curl leaf in the cucumber vegetable plant using hybrid approach.

In this study, a novel framework is introduced for the detection and identification of cucumber leaf curl disease utilizing data collected through IoT-enabled sensing devices. The experimental outcomes demonstrate that the proposed hybrid model effectively identifies the occurrence of curl leaf disease in cucumber plants. This technique, along with other similar ones, efficient tool for the early detection of the disease, thereby supporting timely intervention and improved crop management.

OBJECTIVES AND CONTRIBUTIONS

An emerging research focus is the application of IoT-based systems for the detection and identification of cucumber leaf diseases. The key contributions of this study are summarized as follows:

- A new IoT-based system, which is 3-layer architecture of IoT for detection of cucumber disease using environmental parameters.
- The proposed hybrid approach used different ML algorithms to find the best performing algorithm, which can be then used as a base learner, the output of the base learner is used as a input to the meta classifier, as a result the proposed algorithm can classify different classes of cucumber leaf disease.
- The real test-bed implementation to monitor the environmental conditions.

This article presents an explanation and analysis based on the application and evaluation of a ML classifier for the detection of curl disease in cucumber plants. In addition, this article discusses the importance of ML in agriculture, particularly in disease diagnosis, and explains how these ML algorithms can make predictions that are more accurate. Figure 1 presents the graphical representation of the proposed system. The primary objective of this study is to design and implement a hybrid machine learning prediction model for cucumber leaf curl disease, emphasizing both algorithmic accuracy and F1-score performance. In this article, we analyze the accuracy of the same data against different machine learning algorithms and compare the accuracy scores with better models. The ultimate goal is to deliver accurate and effective results in the detection of curl leaf disease.

PROPOSED IOT BASED SYSTEM.

The suggested structure for the smart farm monitoring system is given in Figure 1. The first layer is sensor layer, where sensor nodes are deployed in a real-time environment to gather data on the temperature, humidity, and soil moisture of a cucumber farm. In this system, the gateway node on the network layer is used to forward real-time data to the agriculture service layer, which serves as the data monitor. The second layer is the network layer, which provides transportation services. Presentation is the top layer because this sector of the economy offers services to farmers via SMS alert systems. Display environmental factors, disease incidence, water level detection, etc. In real time. The data is tracked, analyzed, and divided into various classes in this layer.

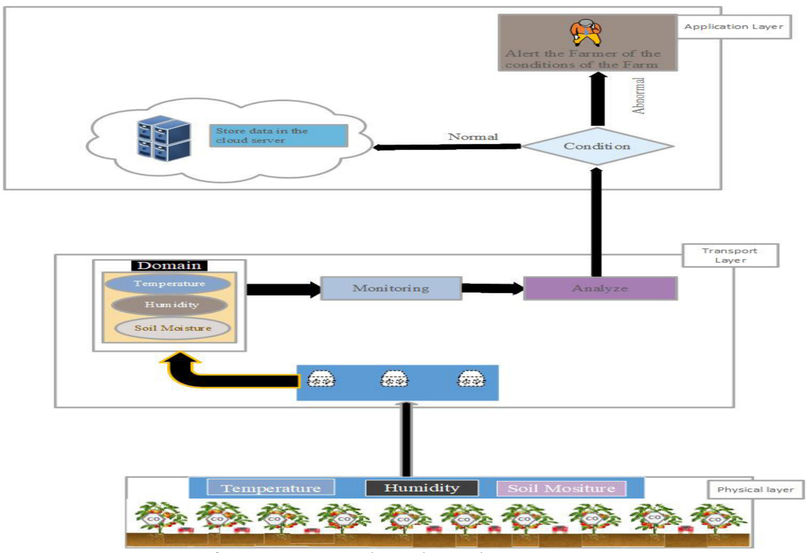

Fig. 1. Proposed IoT based System

SYSTEM DESIGN OVERVIEW OF PROPOSED IOT BASES SYSTEM

The proposed IoT system adopts a standard three-layer architecture for smart agriculture applications. At the first level of perception layer, the environmental sensors are scattered across the space of the cucumber farm to measure and capture the real-time values of temperature, air humidity, and soil moisture. These sensors continuously monitor field conditions, and the collected data are transmitted via the network layer to a remote cloud server for processing and decision-making. Specifically, the Thing Speak cloud platform is utilized to store sensor data, with a Wi-Fi module serving as the gateway for data transfer. In particular, the Thing Speak cloud is used to hold the information captured by the sensors, and the Wi-Fi module is the device used to communicate the information. The information is then processed and examined using several machine learning techniques which include SVM, KNN, and DT. The outcomes derived from this process are sent to the users through the application layer by means of different programming APIs. A detailed overview of each architectural layer is provided below, with Figure 1 illustrating the complete structure of the proposed system.

Physical Layer: In this layer, physical sensors are deployed to collect real time data. This layer is composed of three sensors, which collect humidity, temperature and soil moisture.

- Network Layer: Collected data are send to the server wirelessly to a remote cloud server
- Cloud Server: The collected data are collected at the server where these data are stored and further used for performing different types of operations. At the server, these data are classified using different machine learning algorithms according to its features and classify the data into normal and abnormal class.
- Presentation Layer: results are presented to the end user when there is abnormal data occur.

EXPERIMENTATION OF THE REAL TIME SYSTEM.

DATASET GENERATION

In our experimentation, Arduino board-based sensor nodes are deployed and collects real-time data from a smart cucumber farm. For a two-month period, we have used these prototypes in a smart farm. We use deployed sensor nodes in this study to gather real-time data from a cucumber farm. The dataset, which is stored on a distant server for future processing, consists of information about temperature, humidity, and soil moisture. Using the proposed system, data are recorded and these data can be downloaded in CSV format which can analyzed with the using different ML algorithm to check the classification performance in order to find out the ML algorithm. According to recent studies, environmental parameters such as temperature, humidity, and soil moisture play a crucial role in identifying the presence of cucumber leaf curl disease [33]. The optimal conditions for the onset of this disease are typically observed at temperatures ranging between 30°C and 40°C, relative humidity levels of 50% to 90%, and soil moisture within the range of 20 mm to 50 mm. Based on these parameters, the collected dataset was categorized into three distinct classes. In Class 1, leaf curl disease is present when the temperature ranges from 30°C to 40°C, relative humidity lies between 50% and 90%, and soil moisture varies from 6 mm to 50 mm. Class 2 represents partial disease occurrence under similar temperature and humidity ranges but with soil moisture maintained between 20 mm and 50 mm. The following ranges in class three data indicate that disease does not occur: 30 to 40 °C for temperature, 60 to 75 % for humidity, and 20 to 50 mm for soil moisture. We have classified the generated dataset based on these mentioned parameters.

PROPOSED HYBRID MODEL

We are developing a method that combines different types of ML algorithms to improve the classification performance of ML algorithms. The proposed ML hybrid model has been implemented in Python [34] using pandas, Matplotlib, sklearn, Plotly and other simple libraries. We use the generated data in our experiments using real generated data. There are three types of classes of data, i.e., label 1 represents the normal class, label 2 represents the file that may contain the curl virus, and tag 2 the curl virus. The most accurate ML algorithm is selected for analysis and application to achieve better results. The hybrid model consists of SVM, KNN and DT as base classifier [35]. In our experiment, we chose three ML algorithms based on their performance, as they have the highest accuracy compared to other classifications. We choose a tree-based ML algorithm to achieve the highest accuracy while being able to overcome overfitting. The algorithm with the best performance is selected and other algorithms with low performance are eliminated. In this process, multiple classifiers are combined as the base model and then trained on the same data using met classifiers [36] using different machine learning methods. Classifiers are discussed below.

Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a robust and widely adopted machine learning algorithm primarily used for classification tasks [37]. The central idea entails finding an optimal hyperplane that discriminately separates the distinct classes of data points. Such hyperplanes are formed to maximise the distance between the separating boundary and the close data samples of each class. SVM employs various kernel functions to deal with linear and non-linear distributions of data. These functions map the input features to a higher dimensional space to enable more accurate and precise decision boundary formation. The training of the model is formulated as an optimization problem that minimizes the hinge loss function, while a regularization parameter (C) is employed to control the trade-off between maximizing the margin and minimizing classification errors, ensuring improved generalization and stability.

The stepwise workflow of the SVM algorithm is summarized as follows:

- The input data are first transformed into a higher-dimensional feature space using an appropriate kernel function.
- An optimal hyperplane is then determined by solving an optimization problem that maximizes the margin between the two classes.
- The model parameters are refined by minimizing the objective (loss) function subject to the regularization parameter C.
- The resulting classifier is subsequently validated on unseen data to assess its predictive capability.

The SVM framework remains one of the most powerful and versatile approaches for classification, offering robust performance across both linear and non-linear datasets, and has therefore become a preferred method in numerous machine learning applications.

K NEAREST NEIGHBOR CLASSIFIER

K-Nearest Neighbors (KNN) is one of the most established and popular methods in machine learning and is applied for classification and regression tasks. Its basic tenet revolves around the assumption that the feature space is populated by clusters of similar points. For a query instance, the algorithm computes the k surrounding members of the training dataset, classified according to a chosen distance, for instance, Euclidean or Manhattan. Output is then generated from the k nearest neighbors: for the classification problem, the instance is classified to the class corresponding to the predominant class of the k neighbors, while for regression, the predicted value is calculated from the outputs corresponding to the k neighbors. The distance between the training dataset and the query for the k hyperparameter is critical to model performance, and is usually obtained by trial and error, or through cross-validation, depending on the dataset characteristics.

DECISION TREE

A Decision Tree is one of the supervised learning methods which is used for both classification and regression problems. The algorithm works by progressively dividing the dataset into smaller, more uniform subsets according to specific feature values. During this constructive stage, the most strategic attribute is chosen at any location based on information gain or the Gini index. The attribute value is used to configure a recursive tree structure where parts of the tree correspond to internal nodes. From the vantage of the nodes all the way to the tip, features are the outcomes of a series of classification interrogatives, branches are outcomes and the nodes on the far end, or leaves, provide the end classification or the expected value.

The decision tree learning process can be outlined as follows:

- The entire dataset is initially treated as the root node.
- For every attribute, the algorithm computes a splitting criterion such as information gain or Gini index to evaluate the purity of the resulting subsets.
- The attribute yielding the maximum information gain or the minimum impurity is chosen as the optimal feature for division.
- The dataset is then partitioned into branches according to the selected feature's values.
- This splitting procedure continues recursively

until a stopping condition is met, such as reaching the predefined tree depth or the minimum number of samples required in a terminal node.

- Finally, each terminal (leaf) node is assigned a class label or a predicted output based on the dominant class or the mean of the target values within that node.

The decision tree model can be mathematically formulated as follows.

Consider a dataset D = {(x1, y1), (x2, y2), …, (xn, yn)}, where xi represents the feature vector of the ith observation and yi denotes its corresponding class label or regression target. The algorithm begins by initializing the root node with the entire dataset D. A feature f is then selected that provides the best partition of the data into subsets D1 and D2, based on a splitting criterion such as information gain or Gini index.

A decision node is created according to the chosen feature f, and the process is recursively repeated for each subset until a stopping condition such as a maximum depth or minimum number of samples per node is reached. Each terminal (leaf) node is then assigned to a class label or regression output according to the dominant category or the mean target value of the samples it contains.

The decision tree is conceptually straightforward and capable of handling both categorical and continuous variables. Moreover, it can tolerate missing values and noisy data. However, when the tree grows excessively deep, it tends to overfit the training data. To mitigate this issue, strategies such as pruning, regularization, or ensemble learning techniques (e.g., bagging or boosting) are typically employed.

In this research, a hybrid methodology is proposed by integrating several high-performing machine learning algorithms to improve predictive accuracy. The overall workflow of the developed model is depicted in Figure 2, and the corresponding execution process is outlined as follows:

- Collect the data of curl dataset, which was generated using IoT devices.
- Preprocessing step: In this step, we have implemented SMOTE algorithm to make the data set more reliable and informative. We have three different classes in our data with various number of records which seem very imbalanced. To fix the issue, we performed these steps.
- The dataset is divided into training and testing data.
- Train the base models SVM, KNN and DT respectively
- Then these models are used to generate a hybrid model.
- Once the hybrid model is trained, it gives the prediction on the basis of testing dataset.

In the proposed methodology, a single generated dataset was employed and subsequently divided into separate training and testing subsets. Consistent with standard experimental practice, 80% of the data was used to train the hybrid model, while the remaining 20% was reserved for performance evaluation. The model integrates Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Decision Tree (DT) algorithms to predict the occurrence of cucumber leaf curl disease using the test samples. The resulting outputs were analyzed, visualized, and comparatively assessed across the different classifiers. The proposed framework provides the following key advantages:

- Implementation of three machine learning algorithms that achieve high accuracy on the dataset.
- Evaluation of classification accuracy across different machine learning algorithms to identify the optimal model.
- Development of a novel hybrid model with enhanced accuracy compared to individual machine learning algorithms.

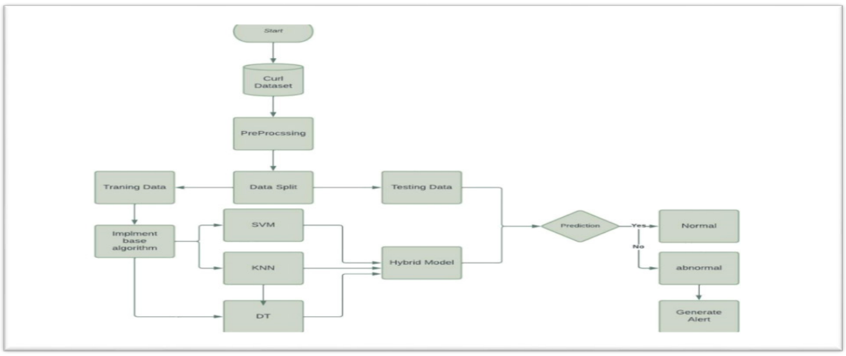

Fig. 2. Proposed Hybrid Model

HYBRID METHOD

In this work, we used a technique called stacking (because it achieve high accuracy as compare to voting and bagging in our experimentation) to build our hybrid model. Stacking is a type of ensemble technique where multiple classification models are combined with a meta-classifier to make predictions. The models are arranged in multiple layers, with the lower layers taking attributes as input from the original data and passing their predictions to the uppermost layer, which makes the final prediction based on a combination of the different models. In this study, this technique was employed to process the original dataset using multiple independent machine learning models as base learners. The predictions generated by these models were then combined and provided as input to a meta-classifier, which produced the final prediction. Only the algorithms that demonstrated superior performance were retained, while the less effective ones were excluded. By integrating multiple classifiers trained on the same dataset through diverse learning algorithms, the proposed hybrid model achieved improved overall performance. The graphical representation of the developed hybrid framework is illustrated in Figure 2.

RESULTS OF THE PROPOSED SYSTEM

It can be analyzed from recent studies [40], accuracy is one of the widely adopted performance metric and a lots of above 70% scientific research this to compare their proposed method. To evaluate the classification performance of the hybrid model, we conduct several experiments using a leaf image database (which is generated using IoT devices) for cucumber diseases. We compare with three different ML algorithm: SVM, KNN and DT ML algorithm and used these algorithms in our proposed hybrid algorithm. We have tested different ML algorithms and we have compared only three for competitive analysis. We have trained each algorithm on a percentile split of 80-20. All the experiments are carried out on Inter core i7 computer with 8 GB RAM and for classification and development of the algorithm, we have used Python. To check the performance of the proposed system, we have used different classification parameters. We have used commonly used parameters include [41] [42]:

To assess the performance of the proposed hybrid machine learning model for cucumber leaf curl disease detection, several standard evaluation metrics were employed.

Accuracy measures the proportion of correctly classified samples compared to the total number of samples in the dataset and is defined as:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (1)$$

Precision indicates the fraction of correctly identified positive cases among all samples predicted as positive and is computed as:

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

Recall, also referred to as **sensitivity**, evaluates the ability of the model to correctly detect all actual positive instances and is expressed as:

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

Specificity represents the proportion of correctly predicted negative cases among all actual negatives and is formulated as:

$$Specificity = \frac{TN}{(TN + FP)} \quad (4)$$

Finally, the F1-score serves as a balanced measure between precision and recall, representing their harmonic mean. A higher F1-score denotes superior model performance and is given by:

$$F1\,Score = 2 * \frac{(precision * recall)}{(precision + recall)} \quad (5)$$

Here, TP, FP, TN, and FN refer to true positives, false positives, true negatives, and false negatives, respectively. These metrics collectively provide a comprehensive evaluation of the classification accuracy and reliability of the proposed hybrid model in detecting cucumber leaf curl disease.

Classification Performance Different Machine Learning Algorithm

The proposed hybrid model utilises ML techniques such as Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and Decision Tree (DT) for the classification and prediction of cucumber leaf curl disease. These classifiers are integrated into the hybrid model, as illustrated in Table 1, which summarizes the classification results of the SVM classifier on three cucumber disease datasets. The hybrid model captures the essence of SVM, KNN, and DT to further improve detection precision [24–27]. Table 1 shows that the hybrid model attained an accuracy of 98.29%, substantially eclipsing the individual algorithm accuracies (SVM at 97.86% and KNN & DT at 97.56%). This confirms that the hybrid approach sustains detection precision while reducing overfitting.

Table 1

Classification performance of the Hybrid approach

S/no	Name	Accuracy	F1 Score	Simulation Time
1	SVM	97.86	.96	0.81s
2	KNN	97.56	.97	0.97s
3	DT	97.56	.97	796ms
4	Proposed Hybrid model	98.29	.98	1.3 S

Table 2 shows the confusion matrix of our proposed algorithm. It can be analyzed, that the proposed hybrid correctly classified most of the data in their corresponding class. The proposed hybrid system correctly classify all the instances correctly and no data is miss classified as abnormal class. Similarly, for class 2, only one entry is miss classified as they belong from normal class and two entries are classified as they belong to the entries of the third class. For class 3, no data is miss classified and all the data are correctly classified.

Table 2

Confusion Matrix of the proposed hybrid algorithm

	Normal Class	Class 2	Class 3
Normal Class	65	0	0
Class 2	1	42	2
Class 3	0	0	65

Table 3 presents the classification performance of the proposed hybrid model. The results demonstrate that the model achieved consistently high accuracy across all three classes. The overall classification accuracy reached 98.29%, indicating that the model correctly predicted the class labels for 98.29% of all instances in the dataset. Among the individual classes, Class 1 attained the highest precision score of 1.0, reflecting that all its positive predictions were accurate. Similarly, Class 2 achieved the highest recall value of 1.0, confirming that the model successfully identified all true positive samples within this class. The overall F1-score of 0.98 further highlights the model's balanced performance in terms of both precision and recall, demonstrating its robustness and reliability in detecting cucumber leaf curl disease.

Table 3

Classification Performance of the proposed hybrid model

Classes	Accuracy	Precision	Recall	F1 Score
Class 1	99.43%	1.0	0.98	0.99
Class 2	98.29%	0.93	1.0	0.97
Class 2	98.86%	1.0	0.97	0.98
Average	98.29%	97.66	98.33	0.98

In order to compare the classification performance of the proposed hybrid model, we have compared our results with three high performing ML algorithms. Table 3 shows the classification performance of our proposed model with other methods. It can be analyzed from table that the proposed model perform well in term of accuracy and archives highest accuracy as compared to other ML algorithm. Similarly, the proposed system also have highest F1 score which show its effectiveness in the detection of curl leaf disease identification.

Figure 3 and 4 shows the comparison of out hybrid approach with other algorithm. We have compared our approach with other machine-learning algorithm in term of accuracy and F1 score. From the figures, it can be shown that the proposed hybrid algorithm archives high accuracy and F1 score as compared to other SVM, KNN and DT.

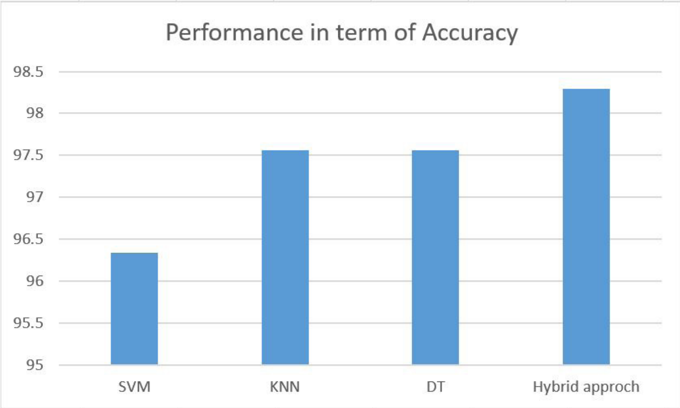

Fig. 3. Performance Comparison with Algorithms

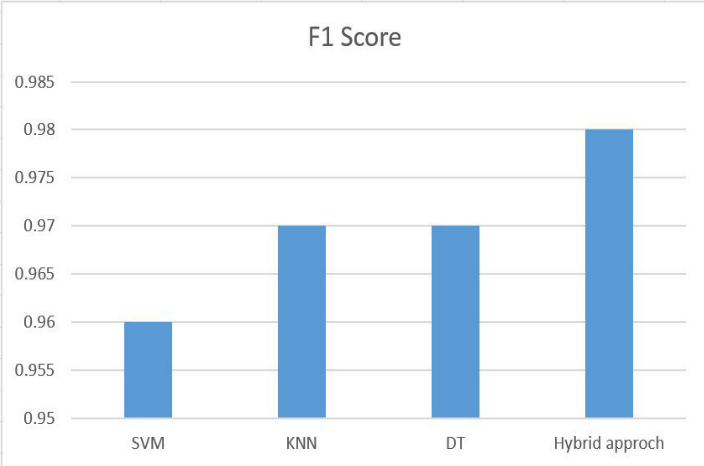

Fig. 4. F1-Score Comparison

CONCLUSION

Rapid and automated cucumber disease detection for curl leaf's plays an important role in crop disease detections. This study developed a new method to detect cucumber leaf diseases based on IoT and ML algorithms. Experimental results show that the proposed method can identify cucumber leaf diseases. This study explores possible ways to use the strategically important Internet of Things to identify foliar diseases in agricultural areas in real time. We propose the identification of new diseases based on cucumber leaves. The experimental evaluation of the proposed model on the cucumber leaf curl disease dataset demonstrates that the hybrid approach is both effective and practically feasible. The model achieved an average accuracy of 98.23% and attained the highest F1-score of 0.98, outperforming the other machine learning algorithms used for comparison. These results confirm the robustness and efficiency of the proposed hybrid model in accurately detecting curl cucumber leaf disease, thereby validating its potential applicability in real-world smart farming environments.

Future Work

Future research will further explore the proposed system by improving model generalizability with the addition of diverse cucumber varieties and environmental parameters from multiple regions in the dataset. Model refinement for disease detection could result from augmented datasets comprising soil nutrient levels or images of diseased cucumber leaves. The proposed scope of this research aims to augment the current hybrid machine learning framework with the integrated application of Convolutional Neural Networks and other advanced deep learning architectures. The proposed framework will also be applicable for resource-constrained smallholder farmers through the precise optimization of the IoT architecture for low-cost and energy-efficient sensors. The focus demonstrated real-world application will include the deployment of the framework to larger farms and the integration with mobile devices to provide real-time alerts to farmers.

Competing Interests

The authors did not declare any competing interest.

References

[1]. L. Raufer, J. Wiedey, M. Mueller, P. Penava, and R. Buettner, "A deep learning-based approach for the detection of cucumber diseases," *PLoS ONE*, vol. 20, no. 4, pp. e0320764, Apr. 2025.

[2]. A. L. Duguma and X. Bai, "How the Internet of Things technology improves agricultural efficiency," *Artificial Intelligence Review*, vol. 58, no. 2, Dec. 2024.

[3]. S. M. Omer, K. Z. Ghafoor, and S. Askar, "An intelligent system for cucumber leaf disease diagnosis based on the tuned convolutional neural network algorithm," *Mobile Information Systems*, vol. 2022, pp. 1–16, Dec. 2022.

[4]. Z. Bi *et al.*, "Apple leaf disease severity grading based on deep learning and the DRL-Watershed algorithm," *Scientific Reports*, vol. 15, no. 1, Aug. 2025.

[5]. S. Zhang, W. Huang, and C. Zhang, 'Three-channel convolutional neural networks for vegetable leaf disease recognition', *Cognitive Systems Research*, vol. 53, pp. 31–41, 2019.

[6]. T. Rahim, Ibadullah, A. Nazir, M. S. Tanveer, and M. R. Qureshi, "A deep learning approach to PCOS diagnosis: Two-stream CNN with transformer attention mechanism," *SES*, vol. 3, no. 7, pp. 1–20, Jul. 2025.

[7]. B. S. Sukhadeo, Y. D. Sinkar, S. D. Dhurgude, and S. V. Athawale, "Plant disease detection using machine learning techniques based on Internet of Things (IoT) sensor network," *Internet Technology Letters*, vol. 8, no. 2, Jul. 2024.

[8]. A. Johannes *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.

[9]. Z. Zhang *et al.*, "Enhancing wheat disease diagnosis in a greenhouse using image deep features and parallel feature fusion," *Frontiers in Plant Science*, vol. 13, Mar. 2022.

[10]. P. Kartikeyan and G. Shrivastava, 'Review on emerging trends in detection of plant diseases using image processing with machine learning', *International Journal of Computer Applications*, vol. 975, no. 8887, 2021.

[11]. N. P. Martono, T. Sawada, T. Uchino, and H. Ohwada, "Deep learning-based indoor localization using wireless sensor network: An efficient approach for livestock monitoring," *Vietnam Journal of Computer Science*, vol. 11, no. 3, pp. 447–463, Apr. 2024.

[12]. R. Benameur, A. Dahane, B. Kechar, and A. E. H. Benyamina, "An innovative smart and sustainable low-cost irrigation system for anomaly detection using deep learning," *Sensors*, vol. 24, no. 4, p. 1162, Jan. 2024.

[13]. Ibadullah*, M. S. Tanveer, M. Khan, and H. ur Rahman, "Sperm morphology classification using Xception-CBAM: A deep learning approach on the SMIDS dataset," *Annual Methodological Archive Research Review*, vol. 3, no. 7, pp. 149–181, Jul. 2025.

[14]. Shakir Ullah, Imad Ullah, Ibad Ullah, Naseer Ullah, and Muhammad Taufiq, "A NOVEL MODIFIED RELATIVE DISCRIMINATION CRITERION FOR FEATURE RANKING IN TEXT CLASSIFICATION", *SES*, vol. 3, no. 10, pp. 349–367, Oct. 2025.

[15]. J. Arshad *et al.*, 'Deployment of wireless sensor network and iot platform to implement an intelligent animal monitoring system', *Sustainability*, vol. 14, no. 10, p. 6249, 2022.

[16]. G. Idoje, T. Dagiuklas, and M. Iqbal, 'Survey for smart farming technologies: Challenges and issues', *Computers & Electrical Engineering*, vol. 92, p. 107104, 2021.

[17]. H. Orchi, M. Sadik, and M. Khaldoun, 'On using artificial intelligence and the internet of things for crop disease detection: A contemporary survey', *Agriculture*, vol. 12, no. 1, p. 9, 2022.

[18]. S. Zhang, S. Zhang, C. Zhang, X. Wang, and Y. Shi, 'Cucumber leaf disease identification with global pooling dilated convolutional neural network', *Computers and Electronics in Agriculture*, vol. 162, pp. 422–430, 2019.

[19]. D. Pixia, W. Xiangdong, and others, 'Recognition of greenhouse cucumber disease based on image processing technology', *Open Journal of Applied Sciences*, vol. 3, no. 01, pp. 27–31, 2013.

[20]. K. Brun-Laguna *et al.*, 'A demo of the PEACH IoT-based frost event prediction system for precision agriculture', in *2016 13th annual IEEE international conference on sensing, communication, and networking (SECON)*, 2016, pp. 1–3.

[21]. T. Truong, A. Dinh, and K. Wahid, 'An IoT environmental data collection system for fungal detection in crop fields', in *2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE)*, 2017, pp. 1–4.

[22]. M. Krishna, S. Sulthana, V. Sireesha, Y. Prasanna, and V. Sucharitha, 'Plant disease detection and pesticide spraying using dip and IoT', *J. Emerg. Technol. Innov. Res*, vol. 6, pp. 54–58, 2019.

[23]. M. A. Nawaz *et al.*, 'Plant disease detection using internet of thing (IoT)', *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 1, 2020.

[24]. A. Thorat, S. Kumari, and N. D. Valakunde, 'An IoT based smart solution for leaf disease detection', in *2017 international conference on big data, IoT and data science (BID)*, 2017, pp. 193–198.

[25]. T. T. Win, 'AI and IoT methods for plant disease detection in Myanmar', *Kobe Institute of Computing: Kobe, Tokyo*, 2018.

[26]. K. K. Ghanshala, R. Chauhan, and R. Joshi, 'A novel framework for smart crop monitoring using internet of things (IOT)', in *2018 first international conference on secure cyber computing and communication (ICSCCC)*, 2018, pp. 62–67.

[27]. K. N.-A. Siddiquee *et al.*, 'Development of algorithms for an iot-based smart agriculture monitoring system', *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–16, 2022.

[28]. B. Qu, X. Jing, X. Wang, Y. Li, and Y. Liang, 'Design on cucumber traceability system based on the internet of things', in *Computer and computing technologies in agriculture V: 5th IFIP TC 5/SIG 5.1 conference, CCTA 2011, beijing, china, october 29-31, 2011, proceedings, part I 5*, Springer, 2012, pp. 199–208.

[29]. R. Rayhana, G. Xiao, and Z. Liu, 'Internet of things empowered smart greenhouse farming', *IEEE Journal of Radio Frequency Identification*, vol. 4, no. 3, pp. 195–211, 2020.

[30]. M. Halder, A. Sarkar, and H. Bahar, 'Plant disease detection by image pro-cessing: a literature review', *image*, vol. 1, p. 3, 2019.

[31]. P. Kolhe, K. Kalbande, and A. Deshmukh, 'Internet of thing and machine learning approach for agricultural application: A review', in *2022 10th international conference on emerging trends in engineering and technology-signal and information processing (ICETET-SIP-22)*, IEEE, 2022, pp. 1–6.

[32]. T. Saranya, C. Deisy, S. Sridevi, and K. S. M. Anbananthen, 'A comparative study of deep learning and Internet of Things for precision agriculture', *Engineering Applications of Artificial Intelligence*, vol. 122, p. 106034, 2023.

[33]. S. Zhang, X. Wu, Z. You, and L. Zhang, 'Leaf image based cucumber disease recognition using sparse representation classification', *Computers and electronics in agriculture*, vol. 134, pp. 135–141, 2017.

[34]. R. Vallat, 'Pingouin: statistics in python.', *J. Open Source Softw.*, vol. 3, no. 31, p. 1026, 2018.

[35]. X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, 'A survey on ensemble learning', *Frontiers of Computer Science*, vol. 14, pp. 241–258, 2020.

[36]. C. B. C. Latha and S. C. Jeeva, 'Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques', *Informatics in Medicine Unlocked*, vol. 16, p. 100203, 2019.

[37]. S. Ibrahim, D. N. Khan Marwat, N. Ullah, K. S. Nisar, and Kamran, "Flow and heat transfer in a meandering channel," *Frontiers in Materials*, vol. 10, Jul. 2023.

[38]. D. T. Larose and C. D. Larose, 'k-nearest neighbor algorithm', 2014.

[39]. P. Pawar, V. Turkar, and P. Patil, 'Cucumber disease detection using artificial neural network', in *2016 international conference on inventive computation technologies (ICICT)*, 2016, pp. 1–5.

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

[41]. F. Salo, M. Injadat, A. B. Nassif, A. Shami, and A. Essex, 'Data mining techniques in intrusion detection systems: A systematic literature review', *IEEE Access*, vol. 6, pp. 56046–56058, 2018.

[42]. H. Li *et al.*, "Classification of electrocardiogram signals with waveform morphological analysis and support vector machines," *Medical & Biological Engineering & Computing*, vol. 60, no. 1, pp. 109–119, Jan. 2022.