INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Check for updates

OPEN ACCESS

Received: 11-01-2024

Accepted: 27-02-2024

Published: 30-03-2024

Citation: Nirgude V, Rathi S (2024) i-PomDiagnoser: A Real-Time Pomegranate Disease Management System. Indian Journal of Science and Technology 17(14): 1391-1401. https://doi.org/10.17485/IJST/v17i14.57

* **Corresponding author**.

vaishali.nirgude@thakureducation.org

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

i-PomDiagnoser: A Real-Time Pomegranate Disease Management System

Vaishali Nirgude[1]*, Sheetal Rathi[2]

1 Research Scholar, Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India
2 Professor, Department of Computer Engineering, Thakur College of Engineering and Technology, Mumbai, Maharashtra, India

Abstract

Objectives: Designing and developing i-PomDiagnoser: a real-time pomegranate disease management system for disease detection, classification, prediction, recommending preventive measures, and analyzing abrupt climatic changes and their impact on pomegranates. **Methods**: A data collection framework has been designed and developed using an agriculture drone, sensors, camera, and other equipment to collect real field pomegranate images and micro-level parameters. Comprehensive Exploratory Data Analysis (EDA) and Feature Selection (FS) processes were carried out to improve the accuracy of disease classification and forecasting models. ML-based Binary, Multimodel, and Multilabel classifiers were implemented for disease classification. The models were trained on 11 years of historical data and tested on 5 months of actual field data. A hybrid pomegranate disease forecasting model has been developed for accurately forecasting micro-level parameters for the next 45 days to predict diseases. **Findings**: Micro-level (weather, soil, water) parameters specific to the agro-climatic zone were collected. The five most prominent distinct diseases are considered for experimentation namely Bacterial Blight (Telya), Anthracnose, Fruit spot, Fusarium Wilt, and Fruit borer. The proposed Improved Ensemble Multilabel Classifier (i-Ensemble-MLC) with a modified voting scheme has achieved a high classification accuracy of 95.82%, addressing model overfitting and data imbalance. Moreover, the hybrid pomegranate disease forecasting model, combining LSTM and i-Ensemble-MLC, demonstrated better performance with minimal error rates (MSE: 0.003, RMSE: 0.071, MAE: 0.048, R2: 0.7) compared to the existing model[1] (MSE:0.037, MAE:0.028). **Novelty**: The novelty lies in the creation of the all-in-one model, i-PomDiagnoser. This innovative system helps the farmers to correctly detect and predict the most prominent diseases of pomegranate.

Keywords: Pomegranate; Agriculture; Disease Forecasting; Machine Learning; Deep Learning

1 Introduction

The present state of agriculture in India, being the backbone of the country's economy, is facing significant challenges due to climate change, unpredictable weather patterns, and the increasing incidence of crop diseases [2]. These factors have resulted in a decrease in both the quantity and quality of agricultural products. In order to overcome these challenges and optimize agricultural production, it is crucial to make use of advanced technologies. There is a lot of research being done nowadays in the areas of agriculture and horticulture. The Indian government is also offering many opportunities to advance research and development in these areas. In India, agriculture is a significant source of income and the majority of the population is reliant on agricultural products. Agriculture contributes about 18.03 % to the total Gross Domestic Product (GDP) in the year 2022 and offers jobs to more than 50 % of the country's population [3]. Over the past decade, the climatic conditions are abruptly changing that lead to the occurrence of various pests and diseases which in turn affect the quality of fruits [4]. To address the challenges posed by the growing population and its impact on food supply, it is essential to focus on sustainable agricultural practices, efficient resource management and disease management [5]. Accurate early detection and prediction of fruit diseases benefit farmers, consumers, and the government to make crucial policy decisions about import-export, pricing, storing, and alleviating food shortages.

The motivation behind this research is to explore the application of Artificial Intelligence (AI) and Drone technology in fruit disease management, specifically focusing on detection and prediction models using real-field micro-level (weather/soil/water) datasets. By addressing the gaps in research and enhancing disease detection and prediction capabilities, small farmers in agro-climatic zones can directly benefit from more effective disease management practices, ultimately improving fruit quality, minimizing losses, and increasing overall agricultural output. Machine Learning (ML) and Deep Learning (DL) algorithms can be trained to analyze large amounts of data, including images and sensor readings, to identify early signs of disease in fruit crops. Drones equipped with cameras and sensors can capture detailed images and collect data from a bird's-eye view, allowing for early detection of diseases that may not be easily visible to the naked eye. This enables prompt intervention and targeted treatment, reducing the spread and impact of diseases. ML and DL algorithms can help identify patterns, predict disease outbreaks, and recommend appropriate preventive or control measures. By combining AI and drone technology with traditional disease management practices, farmers and agricultural experts can enhance their ability to detect, monitor, and predict fruit diseases efficiently.

We have conducted the farm survey (Sept 2019) and agriculture domain experts survey (Jan 2020) and realized further research work is necessary to address emerging diseases, optimize integrated disease management approaches, and analyze climate change and its impact on diseases. In this research work, we have chosen Pomegranate fruit for the study and experimentation of effective disease management. Pomegranate (*Punica granatum* L) known as the "fruits of the God" is one of the important cash fruit crops of arid and semiarid regions in India [6].Over the past ten years, India has become the leading pomegranate grower, producer, and exporter in the world. Among the eight major pomegranate-growing states, Maharashtra is the largest producer in India. However, Maharashtra's production has decreased recently due to highly contagious diseases like Bacterial Blight (Telya) [7].

1.1 Research Contributions

This research work intends to study and analyze various diseases on pomegranates, design and develop a disease management system for disease detection, classification, prediction and recommend disease preventive measures on real field datasets, and analyze sudden climatic changes and their impact on pomegranate diseases. This work can be useful for the farmers, gram-panchayat, and agro-industry.

Contribution 1: The research outcome relies heavily on the data quality used in the processing. Proper validation of the data before analysis is essential. To ensure accuracy and quality, we have designed and developed a data collection framework using agricultural drones, sensors, cameras, and other equipment. This framework has been used to collect real field stage wise disease development image datasets from two distinct seasons (Ambie and Hast Bahar), along with micro-level weather-soil-water parameters specific to the agro-climatic zone. As far as we know, this is the first attempt made to collect data from the Pathare Village, located in Nashik District, Maharashtra State. Both the image and weather-soil-water datasets have been thoroughly validated by the agriculture domain expert. In Section 2.1, present a concise data collection procedure and dataset description

Contribution 2: There are various diseases occur on pomegranate due to sudden climatic changes. Most of the researchers have considered only one or two major diseases for pomegranate disease detection and prediction. However, our approach takes a broader perspective. We have considered the five most prominent occurring diseases on pomegranate fruits: Bacterial Blight (Telya), Anthracnose, Fruit Spot, Fusarium Wilt, and Fruit Borer. Hence, we have designed and developed the all-in-one model, i-PomDiagnoser.

Contribution 3: We have proposed an Improved Ensemble Multilabel Classifier (i-Ensemble-MLC) with a modified voting scheme for solving model overfitting, and data imbalance problems, making the model more generalized and improving Multilabel disease classification accuracy. In Section 2.4, elaborate on the i-Ensemble-MLC algorithm.

Contribution 4: We have designed and developed a hybrid forecasting model by combining Long-Short Term Memory (LSTM) and i-Ensemble-MLC (classifier) on micro-level parameters (Time-series data). This model has been used to forecast micro level parameters for next 45 days in order to predict Multilabel diseases. Section 3.2, illustrates the experimental results.

A literature survey was carried out to comprehensively understand the current state of research and identify any gaps that need to be addressed (Table 1). Specifically, we focused on studying deep learning methods for managing crop diseases, encompassing the identification, classification, and prediction of diseases in crops and fruits based on a data-driven approach as well as image data sets. Additionally, we have provided an in-depth literature review on pomegranate cultivation, including information on pests and diseases affecting pomegranate crops, management strategies, techniques for collecting real-field data, and feature selection methods.

Table 1. Gaps and Findings

Agri Area Application	Methodology	Dataset & Size	Performance Metrics	Gaps/Future Scope
Fruit disease detection, Classification, and Visualization (Image-Based Approach)	CNN-LSTM 2021 [8]	Pomegranate Fruit Images Size: 6519	Accuracy: 98.17 %	• Only major diseases are considered for disease detection. • Automatic fruit disease detection and classification is difficult because symptoms are not well defined. • Real-field fruit disease detection is difficult due to the presence of complex background. • Need to deal with Deep Learning 'Black Box' problem.
	CNN 2023 [9]	Pomegranate Leaf Images Size: 1844	Accuracy: 98.38%	
	ResNet Model Grad-CAM 2023 [10]	Pomegranate Fruit Images Size: 1221	Accuracy: 98.55%	
	Inception-CNN 2020 [11]	Grape Leaf Images Size: 3646	Accuracy: 97.22%	
	CNN Model 2019 [12]	Guava Leaf Images Size: 87,000	Accuracy: 95.61%	
	VGGNet 2019 [13]	Sugarcane Images Size: 2100	Accuracy: 95.4%	
	Deep CNN 2020 [14]	Kiwifruit Leaf Images Size: 11322	Accuracy: 98.54%	
	CNN-Inception V3 Model 2020 [15]	Plum fruit Images Size:87,848	Accuracy: 92%	
	CNN Model Grad-CAM 2021 [16]	Coffee Leaf Images Size: 1560	Accuracy: 98%	
	AlexNet Model 2023 [17]	Pomegranate Leaf Images Size: 1245	Accuracy: 97.60%	
Fruit/Crop disease forecasting (Data-Driven Approach)	LSTM 2019 [18]	Cotton: Weather parameters	AUC: 0.97	• Need more research on different agro-climatic zones. • All agro-climatic parameters (weather-soil-water) are not considered for prediction. • Need to analyze sudden climatic changes and impact on diseases. • Fruit Disease Prediction/Forecasting has not been explored in depth.
	RF 2019 [19]	Mango: Weather parameters	MSE: 0.007	
	LSTM 2020 [20]	Rice: Weather parameters	MSE: 0.135	
	LSTM 2022 [21]	Weather Forecasting	MSE: 0.08	
	Hybrid ARIMA-Bi-LSTM2022 [1]	Rice: Weather parameters	MSE: 0.037	

2 Methodology

The proposed system architecture of a real-time pomegranate disease management system: i-PomDiagnoser for disease detection, visualization, classification, prediction, and recommended disease preventive measures is presented in Figure 1.

2.1 Data Collection

Pomegranate fruit disease classification and prediction are based on a data-driven approach. A framework of data collection has been designed using agriculture drone, sensors, camera, and other equipment to collect micro-level (weather, soil, water) parameters. Emphasized quality data collection through different sources such as historical data, on-site measurements, ground

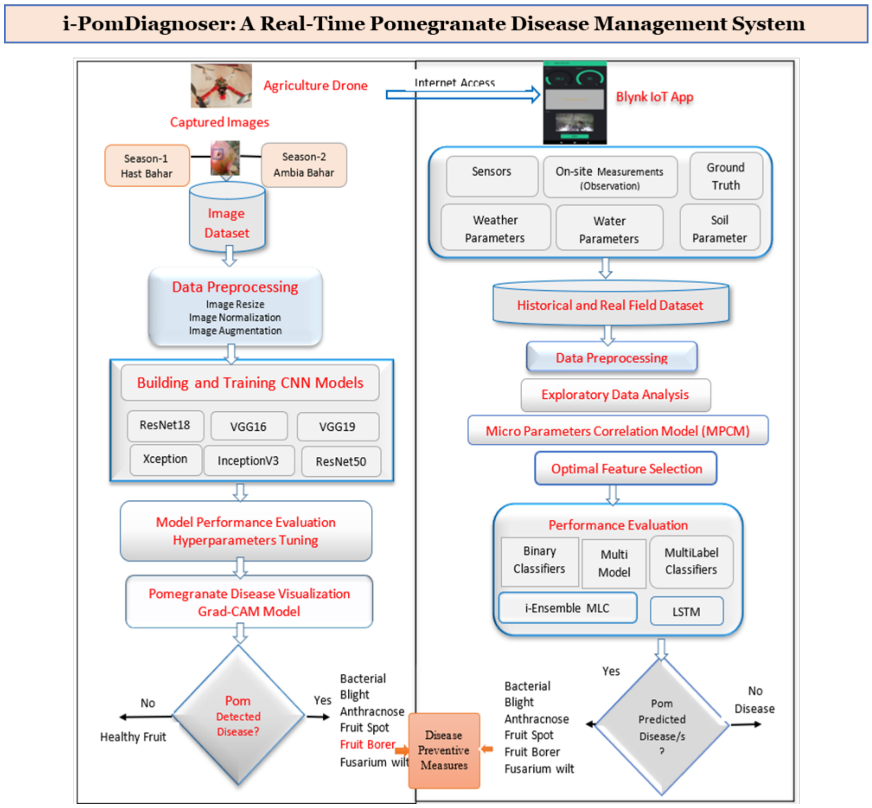

Fig 1. Overall System Architecture of i-PomDiagnoser

truth, and via Agriculture drone, sensors, cameras, and other equipment. We have configured an Agriculture drone with the required specifications. Weather parameters were collected by placing the sensors on the drone assembly. Soil Moisture-pH combo has been used to collect Soil pH, Soil Moisture, and Water pH parameters. TDS-3 Meter has been used to collect water TDS value. Pomegrante grows well in loamy, well drained and moderately acid to slightly alkaline soil. Hence, required pH is 5.5 to 7.2 and water TDS should be less than 2000 mg/L for maintaining the optimal health of pomegranate plants. Healthy and unhealthy (Infected) fruit images were collected using a Raspberry camera mounted on the drone. The data has been collected by moving the drone across the line of the different trees. Then the collected data is sent to the Blynk IoT server and displayed on the Blynk IoT app. We have collected real field micro-level parameters from 1st Oct 2021 to 28th Feb 2022 (Total 151 Records and 32 Parameters) to test the model performance. We have also collected the historical weather-soil-water parameters from 1st January 2010 to 1st Dec 2021 (A total of 4227 Records and 32 Parameters) to train the ML and DL models for enhancing fruit classification and prediction accuracy. Details about micro-level parameters are shown in Table 2.

Table 2. Description of real field collected micro-level parameters

Parameter	Description
Fruit/Crop	Pomegranate
Disease/Pest	Bacterial Blight, Anthracnose, Blight Blora, Fruit Spot/ Rot, Fusarium Wilt
Location (State / Dist. / Village)	Maharashtra / Nashik / Pathare
Observation (DD-MM-YY)	Period: 1st Oct 2021 – 28th Feb 2022, (No. of Records:151, No. of Paremters: 32)
Weather Parameters	DHT-11 sensors are used to measure Temperature(^{0}C) and Humidity (%). Temperature in (^{0}C) was recorded 3 times a day. (T_M: 7:00 am-8:00 am, T_A= 1:00pm - 2:00pm, T_E=6:00 pm to 7:00 pm, T_{avg}) Humidity in (%) was recorded 3 times a day. (H_M: 7:00 am-8:00 am, H_A= 1:00pm - 2:00pm, H_E=6:00 pm to 7:00 pm, H_{avg}) WS= Windspeed (km/hour), P=Pressure(mb), PP=Precipitation (mm), WD=Weather Description (Heavy rainfall, Light Rainfall, Patchy Rainfall, Cloudy, Sunny/Clear), SH= Sunshine Hours (hrs), Data collection source: https://www.weather-forecast.com/locations/Nashik

Continued on next page

Table 2 continued

Soil Parameters	Soil pH-moisture combo tester kit is used to measure Soil pH, Soil Moisture, and Water pH. SM= Soil moisture (%), SpH= Soil pH, WpH= Water pH
Water Parameters	TDS-3 meter is used to measure water TDS. $W_{TDS} =$ Water TDS (mg/l)

2.2 Procedure of data collection

1. Agriculture drone with 2.4 GHz programmable transmitter which consists of 6 channels for drone direction and speed control. The minimum flight time is 15 mins which carries a 1-1.5 kg load. It flies up to 6-7 ft.
2. Agriculture drone assembly platform on which different components are placed. Such as the DHT-11 sensor which captures temp and humidity. Also, a Raspberry Pi camera- model : SKU-128023 is mounted on a drone to capture images of pomegranate fruits and leaves.
3. NodeMCU ESP-32 i.e., Wi-Fi module along with flight controller that provides Wi-Fi connectivity, Image capturing, and Sensor readings.
4. To provide a power supply to the drone a battery of 11.1 V is used in a single package. There are 4 propellers with an Electronic Speed Controller and a Brushless DC Electric Motor (BLDC) for high speed.
5. The remote control Transmitter sends signals to the Flight controller. DHT-11 sensor and Camera connected to ESP32 MCU module which transmits temp, humidity, and captured images to the Blynk server via internet access and displays that information on the Blynk IoT app.
6. A soil pH-Moisture combo tester kit is used to measure soil moisture, water pH, and soil pH Values.
7. TDS-3 Meter is used to measure water Total Dissolve Solid (TDS).

2.3 Data Pre-processing

Data pre-processing is necessary to ensure data quality, improve compatibility with machine learning algorithms, enhance feature representation, reduce noise, prevent overfitting, and optimize computational efficiency. By addressing these aspects, data pre-processing helps set a real foundation for effective data analysis and more accurate predictions. We have collected a total of 32 micro-level weather, soil, and water parameters. The performance of the prediction model is dependent on data quality and selected optimal features, Therefore, we have performed detailed Exploratory Data Analysis (EDA), designed and developed a correlation model using Pearson correlation, and applied various feature selection techniques to select important features [(22)]. We have performed data pre-processing to handle categorical data such as weather description, scaling numerical features, and handling missing data by averaging. Pre-processing large datasets helped to optimize the use of computational resources and reduce training time.

2.4 Design of Pomegranate Disease Classification Model

We have designed and developed Binary, Multimodel, and Multilabel disease classification models by using Machine Learning techniques. A binary classifier is designed to classify instances into one of two exclusive classes or categories. A multimodel classifier combines predictions from multiple individual classifiers to make a final prediction. A multilabel classifier is capable of assigning multiple labels or categories to each instance. Binary classifiers predicts the likelihood of diseases. ML classifiers such the Logistic Regression, K-Nearest Neighbour, Naive Bayes, Decision Tree, Random Forest, Gradient Boosting, and Ada Boosting algorithms have been developed for effective prediction of diseases. Experimental results proved that Random Forests gave the minimum loss and the maximum accuracy for binary classification. Multimodel classifiers accurately predicted the diseases using micro-level parameters. Whereas, Multilable classifiers are used to perform multilabel classifications. i.e. there are chances of multiple diseases occurring at the same time based on the condition of weather-soil-water parameters. Experimental results illustrate that Random forest gave the highest accuracy in all the groups. Comparative analysis of all ML-based multilabel classifiers have provided in Section 3.1 and Table 3. Improved Ensemble Multilable Classifier (i-Ensemble-MLC) was designed in order to further increase the accuracy of multilable classification, address data imbalances, and strengthen the model. The following section goes into detail about the i-Ensemble-MLC algorithm.

Input: {Training Dataset D_{Tr} (X_train, y_train), Testing Dataset D_T (X_test, y_test), classifiers ($C_1, C_2, C_3, C_4, \ldots, C_n$)}
Output: { MultiLabelModel_result, i.e., MultiLabel model files and Result file path }

1. Initializing with multiple classifiers $C_1, C_2, C_3, C_4, \ldots, C_n$.
2. Develop machine learning models $M_1, M_2, M_3, M_4, \ldots, M_n$.

3. Monitor classification report {Accuracy, Loss, Zero_one_loss, Hamming_loss} $CR_1, CR_2, CR_3, CR_4, \ldots, CR_n$ on training $P_{iTr} = M_i(D_{Tr})$ and testing $P_{iT} = M_i(D_T)$ dataset for MultiLabel model.
4. Save classification report $CR_1, CR_2, CR_3, CR_4, \ldots, CR_n$ and machine learning model $M_1, M_2, M_3, M_4, \ldots, M_n$.
5. Select top 3 model ($M_{1opt.}, M_{2opt.}, M_{3opt.}$) with best optimal accuracy score.
6. Fit MultiLabel classifier model MLC = ($M_{1opt.}, M_{2opt.}, M_{3opt.}$).
7. Evaluate final predictions using voting method: $P_{final} = \frac{AND\,(MLC_{pred})}{OR\,(MLC_{pred})} = \frac{M_{1pred} \wedge M_{2pred} \wedge M_{3pred}}{M_{1pred} \vee M_{2pred} \vee M_{3pred}}$

2.5 Design of Pomegranate Disease Forecasting Model

We have designed and developed a hybrid forecasting model by combining Long-Short Term Memory (LSTM) and i-Ensemble-MLC (classifier) on micro-level parameters (Time-series data). This model has been used to forecast micro-level parameters for the next 45 days in order to predict Multilabel diseases. LSTM architecture with mathematical model discussed in section 2.5.1. Initially, we fed the optimal 8 features (Temperature (T), Relative Humidity (RH), Pressure(P), Precipitation (PP), Weather Description (WDesc), Windspeed (WS), Sunshine Hours (SH), and Soil Moisture (SM)) to LSTM model for forecasting micro level weather-soil-water parameters. Further, these predicted 8 features are fed to the i-Ensemble-MLC model to predict Multilabel diseases as shown in Figure 2. The performance of the pomegranate disease forecasting model has been evaluated by using MSE, RMSE, MAE, R^2 evaluation metrics. Further, recommended Pomegranate disease preventive measures.

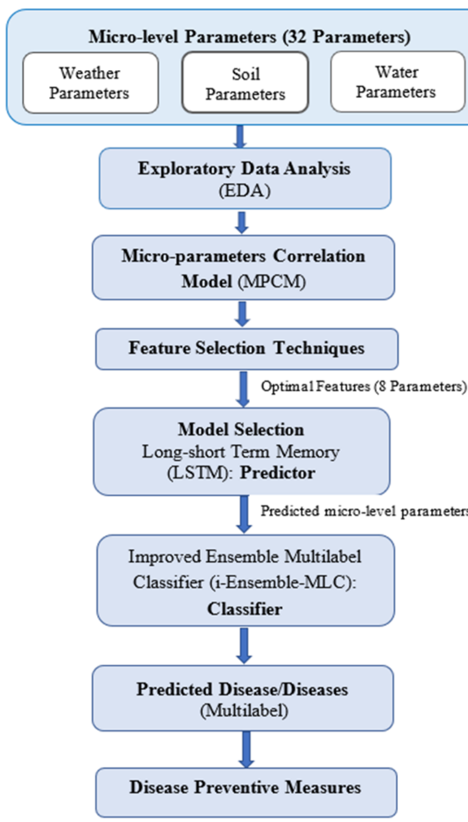

Fig 2. Proposed hybrid pomegranate disease forecasting model

2.5.1 LSTM Mathematical model

Long Short-Term Memory (LSTM) is a recurrent neural network designed for long-term dependencies, suitable for time series data. It tackles the vanishing gradient problem by adding short-term memory using memory cells (C_t) and three gates: input (i_t), forget (f_t), and output gates (O_t) [(21)].

Mathematical model of LSTM

Step 1. Forget irrelevant parts of the previous state, determined by sigmoid function which checks the previous state (h_{t-1}), the current input x_t and computes the function f_t.

$$f_t = \sigma\left(w_f.\left[h_{t-1}, x_t\right] + b_f\right) \tag{i}$$

Step 2: Selectively update cell state values using sigmoid function and tanh function, where i_t is determined by the sigmoid function and C_t is determined by the tanh function.

$$i_t = \sigma\left(w_i.\left[h_{t-1}, x_t\right] + b_i\right) \tag{ii}$$

$$C_t = tanh\left(w_C.\left[h_{t-1}, x_t\right] + b_C\right) \tag{iii}$$

Step 3: Output certain parts of cell state.

$$O_t = \sigma\left(w_O.\left[h_{t-1}, x_t\right] + b_O\right) \tag{iv}$$

$$h_t = o_{t*} tanh\left(c_t\right) \tag{v}$$

3 Results and Discussion

3.1 Experimental Results and Performance Measures of Disease Classification Model

Multilabel classification is a scenario where a data point can be assigned to more than one class, and there are many classes available. In our study, we implemented several machine learning-based models, including LR, KNN, NB, DT, RF, GB, and AB, to act as multilabel classifiers. Among these classifiers, Gradient Boosting (GB) and Random Forest (RF) showed outstanding performance with test accuracies of 97.01% and 96.77%, respectively. They also demonstrated low hamming loss, specifically 0.65% and 0.69%. However, it is important to note that the high training accuracy of DT (100%) and RF (100%) suggest overfitting of the models. Experimental results have presented in Table 3. Further investigation was necessary to address the issue of overfitting and ensure the model's robustness. Therefore, we have proposed an Improved Ensemble Multilabel Classifier (i-Ensemble-MLC) model with a modified voting method for solving model overfitting and data imbalance problems, making the model more generalized and improving disease classification accuracy. A detailed explanation of i-Ensemble-MLC is given in section 3.4. The Hypothesis "If informative features are selected for model development, then predictive performance can be improved" was verified by using i-Ensemble-MLC. Before feature selection (Total number of features: 32) model accuracy was 57.94% whereas after optimal feature selection (Selected Features: 08, Features: Tavg, RH, SH, SM, PP, WD, WS, P) accuracy improved to 95.82% (Table 4). Multilabel classifier's performance has been measured by using evaluation metrics Accuracy, Precision, Recall, F-score, and Hamming Loss.

Table 3. Performace analysis of ML-based Multilabel Classifiers

Multilabel Classifiers	Training Accuracy (%)	Test Accuracy (%)	Precision (%)	Recall (%)	F-score (%)	HammingLoss (%)
Logistic Regression (LR)	71.33	71.71	80.8	69.8	73	7.88
K-NN	88.84	85.97	93	84	88	3.42
Naïve Bayes (NB)	62.41	63.59	50	90	62	13.76
Support Vector Machine (SVM)	97.43	78.64	78	51	61	6.19
Decision Tree (DT)	100	95.04	97	95	96	1.07
Random Forest (RF)	100	96.77	99	96	97	0.69
Gradient Boosting (GB)	97.84	97.01	99	96	97.6	0.65
Ada Boosting (AB)	96.45	94.85	98.5	95.8	97.2	1.14

Table 4. Informative features selection (Iterative) for improving the model's performance

i-Ensemble-MLC Performance Measure Values in (%)	Before Feature Selection	After Feature Selection (FS)					
	No. of Parameters:32	No. of Parameters: 28	No. of Parameters: 24	No. of Parameters: 20	No. of Parameters: 16	No. of Parameters: 10	No. of OptimalSelected Parameters: 8 (FS techniques+ Domain Expert Inputs)
Accuracy	57.94	69.58	71.55	73.52	75.1	94.56	95.82
Precision	79.8	79.8	79.8	79.8	79.8	98	99
Recall	1.8	31.2	34.2	37.8	41	94.2	95.4
F-score	3.6	43.6	47	50.4	53.2	96.2	97.6
Zero one loss	42.06	30.42	28.45	26.48	24.9	5.54	4.18
Hamming Loss	14.75	9.6	9.03	8.38	7.8	1.18	0.96

3.2 Experimental Results and Performance Measures of Disease Forecasting Model

Further, the proposed i-Ensemble-MLC has been used for developing a hybrid pomegranate disease forecasting model. We have designed and developed a hybrid forecasting model by combining Long-Short Term Memory (LSTM) and i-Ensemble-MLC (classifier) on micro-level parameters (Time-series data). This model has been used to forecast micro-level parameters for the next 45 days to predict Multilabel diseases. Initially, we fed the optimal 8 features (Tavg, RH, SH, SM, PP, WD, WS, P) to the LSTM model for forecasting micro-level weather-soil-water parameters. Further, these predicted 8 features are fed to the i-Ensemble-MLC model to predict multilabel diseases. The performance of the pomegranate disease forecasting model has been evaluated by using various evaluation metrics. The second hypothesis the agriculture decision-making process depends on weather, soil, and water parameters, and fruits are affected due to changes in all these parameters. If all the agro-climatic parameters are considered for fruit disease prediction, it will improve prediction accuracy" tested on the forecasting model. Experimental results showed that the hybrid forecasting model obtained a minimum error rate i.e., MSE is 0.003, RMSE is 0.071, MAE is 0.048 and R2 is 0.7 over the existing models[1,23]. Based on the above prediction, disease-preventative measures are recommended for farmers. Figure 3 illustrates the sample weather parameter predictions on the test and validation datasets. X-axis represents number of samples of temperature and humidity parameters and Y-axis represents accuracy. It has been observed that the actual and predicted weather parameters are similar with minimal error rates. Table 5 provides a performance comparison of all 8 micro-level parameters using evaluation metrics.

Table 5. Performance comparison of micro-level parameters using evaluation metrics

Sr. No.	Micro-level Parameter	MSE	RMSE	MAE	R^2
1	T=Temperature (°C)	0.0	0.06	0.04	0.89
2	RH=Relative Humidity in (%)	0.0	0.06	0.04	0.47
3	WS= Wind speed (km/hour)	0.0	0.06	0.05	0.87
4	P=Pressure (mb)	0.0	0.05	0.04	0.92
5	PP=Precipitation (mm)	0.0	0.06	0.05	0.79
6	WD=Weather Description	0.0	0.05	0.02	0.2
7	SH=Sunshine Hours (hrs)	0.03	0.18	0.13	0.5
8	SM=Soil Moisture (%)	0.0	0.05	0.02	0.98
	Average	**0.003**	**0.071**	**0.048**	**0.7**

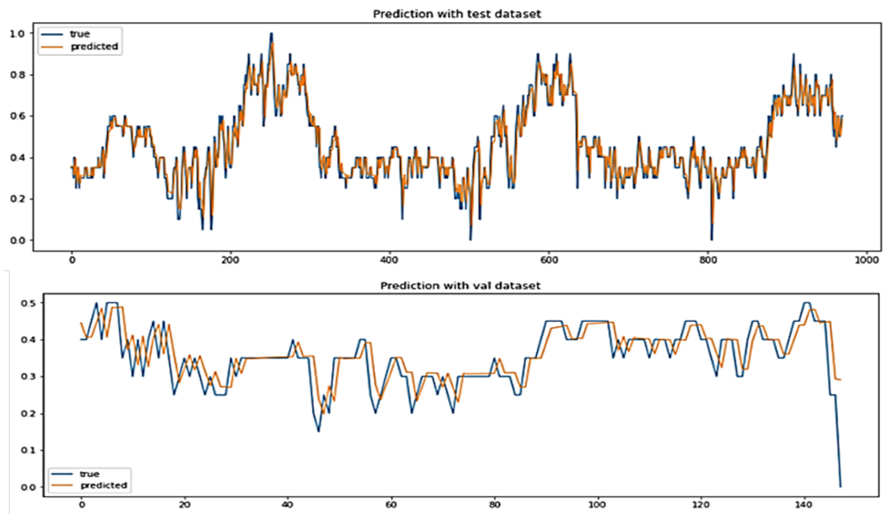

Fig 3. Sample weather parameter prediction on test and validation datasets

4 Discussion

Real-time micro-level weather, soil, and water parameters were collected by using a data collection framework. We aimed to find the relationships between micro-level weather, soil, and water parameters and disease occurrences. We have done extensive Exploratory Data Analysis and also developed the Micro Parameter Correlation Model to discover correlations between weather-soil-water parameters with diseases[22]. Optimal 8 features have been selected by applying various Feature Selection (FS) techniques. We have created a pomegranate disease multilabel classification model using i-Ensemble-MLC, and a forecasting model was developed using a hybrid of LSTM and i-Ensemble-MLC models. A further recommendation of disease prevention measures. We have compared the existing prediction models with their strength and weakness (Table 6).

Our pomegranate disease prediction model surpasses existing models, which typically focus on one or two diseases and rely solely on weather parameters[23]. Recognizing that agricultural decision-making depends on weather, soil, and water parameters, we incorporated all relevant agro-climatic parameters in our model, resulting in enhanced prediction accuracy. Details are shown in Table 7. Experimental results showed that hybrid forecasting model obtained minimum error rate i.e., MSE is 0.003, RMSE is 0.071, MAE is 0.048 and R2 is 0.7 over the existing models[1].

Table 6. Comparision of existing prediction models with strength and weakness

Forecasting / Prediction Models		Strength	Weakness
Traditional Model	Autoregressive Integrated Moving Average (ARIMA)[1]	Simple and interpretable Linear model for short term forecasting.	Not handle non-linear (Non-stationary) and complex data.
	Autoregressive Fractionally Integrated Moving Average (ARFIMA)	Capture both short term and long term dependencies and handle both stationary and non-stationary data.	Computationally complex for large datasets and get overfitted for noisy dataset.
	Seasonal Autoregressive Integrated Moving Average (SARIMA)	Designed for time series with seasonal repeating patterns and handle non-stationary data with seasonality.	Computationally complex and overfitted.
ML Models	Linear Regression[19]	Simple and suitable for linear data	Not handle non-linear data
	Decision Tree[19]	Capture non-linear pattern	Prone to overfitting for deep tree
	Random Forest[22]	Robust and capture complex relationship	Computationally expensive
	Gradient Boosting[22]	High predictive accuracy	Computationally expensive
	Support Vector Machines[19]	Capture complex patterns	Sensitive to the choice of kernel and hyper-parameters
	Artificial Neural Network (ANN)	Handle complex and long-term dependencies	Computationally expensive

Continued on next page

Table 6 continued

DL Models	Recurrent Neural Network (RNN) [20]	Suitable for sequential data processing	Suffer from the vanishing and exploding gradient problem, Short term memory
	Long Short-Term Memory (LSTM) [21]	Solve vanishing gradient problem and suitable for long-term dependencies	Requires large dataset and more training times
	Gated Recurrent Unit (GRU)	Computationally less expensive than LSTM	Not capture long-term dependencies

Table 7. Comparison of the proposed disease prediction model with existing models

Reference	**Name of Fruit / Plant**	**Type of Disease**	**Prediction Models**	**MSE**	**RMSE**	**MAE**	**R^2**
Jawade et. al2019 [19]	Mango	Thrips	RF	0.007	0.0849	0.0472	-
Wahyono et.al2020 [20]	Rice	Pest Attack	LSTM	0.135	0.368	-	-
Suleman & Shridevi2022 [21]	Weather Forecasting	-	LSTM	0.08	-	0.231	0.9
Kumar et.al.2022 [1]	Rice	Rice Blast	Hybrid ARIMA-BiLSTM	0.037	-	0.028	-
Proposed Prediction Model	Pomegranate	Bacterial Blight, Anthracnose, Fruit Borer, Fruit Rot, Wilt	Hybrid LSTM + i-Ensemble-MLC	**0.003**	**0.071**	**0.048**	**0.7**

5 Conclusion

This research work focuses primarily on area-specific research within the agro-climatic zone. We have developed the i-PomDiagnoser: A real-time pomegranate disease management system that encompasses disease detection, classification, prediction, recommendation of preventive measures, and analysis of the sudden climatic changes and their impact on diseases. Our experimentation revolved around the most prevalent diseases affecting pomegranates, namely Anthracnose, Bacterial Blight (Telya), Blight Borer, Fruit Rot, and Fusarium wilt. To ensure accurate and high-quality data collection, we have designed and developed a data collection framework using agricultural drones, sensors, camera, and other equipment to capture images and micro-level (weather-soil-water) parameters.

Pomegranate disease classification and forecasting models have been designed and developed using ML and DL techniques on micro-level weather-soil-water parameters. We have performed detailed exploratory data analysis and feature selection to enhance disease prediction and classification accuracy. The Micro Parameter Correlation Model (MPCM) has been developed to identify correlations between micro-level parameters and diseases. We have proposed an Improved Ensemble Multilabel Classifier (i-Ensemble-MLC) with a modified voting system, achieving a high test accuracy of up to 95.82% and a low hamming loss of 0.96 % in Pomegranate Multilabel disease classification.

Furthermore, for disease forecasting, we have designed a hybrid model using LSTM and i-Ensemble-MLC, achieving a Mean Squared Error (MSE) of 0.003%. This represents an improvement compared to the existing technique [1], which had an MSE of 0.037%. Additionally, recommendations for pomegranate disease treatments and preventative measures to ensure fruit quality and avoid losses. We expect that the outcomes of this research will contribute to the attainment of the ICAR Vision-2050 objectives for pomegranate research and development.

Future Scope

The designed models, tested initially on pomegranates, may also be applicable to other fruits like grapes, mangoes, guavas, etc., as they are susceptible to similar diseases. This research work can be extended to diagnose any infection or diseases occurring inside the fruits and predict the fruit yield.

References

1) Varsha M, Poornima B, Kumar MP, Basavarajappa S. Novel Hybrid ARIMA-BiLSTM Model for Forecasting of Rice Blast Disease Outbreaks for Sustainable Rice Production. Research Square Platform LLC. 2022. Available from: https://doi.org/10.21203/rs.3.rs-2103058/v1.
2) Ankit P, Rashmi C, Krishan K. Analysis of agriculture input consumption by Indian Farmers. *International Journal of Economic Plants*. 2;2020(7):86–90. Available from: http://dx.doi.org/10.23910/2/2020.0369.
3) Joshi PK, Varshney D. Agricultural Technologies in India: A Review. National Bank for Agriculture and Rural Development. 2022. Available from: https://www.nabard.org/auth/writereaddata/tender/1507223612Paper-5-Agricultural-Tech-in-India-Dr.Joshi-&-Varshney.pdf.
4) Shekhar M, Singh N. The Impact of Climate Change on Changing Pattern of Maize Diseases in Indian Subcontinent: A Review. In: El-Esawi MA, editor. Maize Genetic Resources - Breeding Strategies and Recent Advances. IntechOpen. 2021. Available from: https://www.intechopen.com/chapters/79453.
5) Liliane TN, Charles MS. Factors Affecting Yield of Crops. In: Amanullah, editor. Agronomy - Climate Change & Food Security. IntechOpen. 2020. Available from: https://www.intechopen.com/chapters/70658.
6) Phadke M, Karandikar B, Gulati A. Grapes and Pomegranate Value Chains. In: Agricultural Value Chains in India. India Studies in Business and Economics;Singapore. Springer. 2022;p. 145–193. Available from: https://doi.org/10.1007/978-981-33-4268-2_5.
7) Yadav DB, Sanap DJ, Pokharkar VG. Total Factor Productivity Growth in Pomegranate Crop of Maharashtra. *Indian Journal of Agriculture Business*. 2020;6(1):9–14. Available from: https://journals.indexcopernicus.com/api/file/viewByFileId/1073776.pdf.
8) Vasumathi MT, Kamarasan M. An Effective Pomegranate Fruit Classification Based On CNN-LSTM Deep Learning Models. *Indian Journal of Science and Technology*. 2021;14(16):1310–1319. Available from: https://doi.org/10.17485/IJST/v14i16.432.
9) Wakhare PB, Neduncheliyan S. Using Image Processing and Deep Learning Techniques Detect and Identify Pomegranate Leaf Diseases. *Indian Journal of Science and Technology*. 2023;16(18):1323–1331. Available from: https://doi.org/10.17485/IJST/v16i18.768.
10) Nirgude V, Rathi S. Improving the accuracy of real field pomegranate fruit diseases detection and visualization using convolution neural networks and grad-CAM. *International Journal of Data Analysis Techniques and Strategies*. 2023;15(1-2):57–75. Available from: https://doi.org/10.1504/IJDATS.2023.132562.
11) Liu B, Ding Z, Tian L, He D, Li S, Wang H. Grape Leaf Disease Identification Using Improved Deep Convolutional Neural Networks. *Frontiers in Plant Science*. 2020;11:1–14. Available from: https://doi.org/10.3389/fpls.2020.01082.
12) Haque ASMFA, Islam GMR, Hafiz R, Hakim MA. A Computer Vision System for Guava Disease Detection and Recommend Curative Solution Using Deep Learning Approach. In: 22nd International Conference on Computer and Information Technology (ICCIT). IEEE. 2020;p. 1–6. Available from: https://doi.org/10.1109/ICCIT48885.2019.9038598.
13) Militante SV, Gerardo BD. Detecting Sugarcane Diseases through Adaptive Deep Learning Models of Convolutional Neural Network. In: 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS). IEEE. 2020;p. 1–5. Available from: https://doi.org/10.1109/ICETAS48360.2019.9117332.
14) Liu B, Ding Z, Zhang Y, He D, He J. Kiwifruit Leaf Disease Identification Using Improved Deep Convolutional Neural Networks. In: 2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC). IEEE. 2020;p. 1267–1272. Available from: https://doi.org/10.1109/COMPSAC48688.2020.00-82.
15) Ahmad J, Jan B, Farman H, Ahmad W, Ullah A. Disease Detection in Plum Using Convolutional Neural Network under True Field Conditions. *Sensors*. 2020;20(19):1–18. Available from: https://doi.org/10.3390/s20195569.
16) Yebasse M, Shimelis B, Warku H, Ko J, Cheoi KJ. Coffee Disease Visualization and Classification. *Plants*. 2021;10(6):1–13. Available from: https://doi.org/10.3390/plants10061257.
17) Wakhare PB, Kandalkar JA, Kawtikwar RS, Kalme SA, Patil RV. Development of Automated Leaf Disease Detection in Pomegranate Using Alexnet Algorithm. *Current Agriculture Research Journal*. 2023;11(1):177–185. Available from: http://dx.doi.org/10.12944/CARJ.11.1.15.
18) Xiao Q, Li W, Kai Y, Chen P, Zhang J, Wang B. Occurrence prediction of pests and diseases in cotton on the basis of weather factors by long short term memory network. In: BMC Bioinformatics: Proceedings of the 2018 International Conference on Intelligent Computing (ICIC 2018) and Intelligent Computing and Biomedical Informatics (ICBI) 2018 conference: bioinformatics;vol. 20, Supplement 25. Springer Science and Business Media LLC. 2019;p. 1–15. Available from: https://doi.org/10.1186/s12859-019-3262-y.
19) Jawade P, Chaugule D, Patil DM, Shinde HV. Disease Prediction of Mango Crop Using Machine Learning and IoT. In: Advances in Decision Sciences, Image Processing, Security and Computer Vision;vol. 3 of Learning and Analytics in Intelligent Systems. Springer, Cham. 2019;p. 254–260. Available from: https://doi.org/10.1007/978-3-030-24322-7_33.
20) Wahyono T, Heryadi Y, Soeparno H, Abbas BS. Enhanced LSTM Multivariate Time Series Forecasting for Crop Pest prediction. *ICIC Express Letters*. 2020;14(10):943–949. Available from: http://www.icicel.org/ell/contents/2020/10/el-14-10-02.pdf.
21) Suleman MAR, Shridevi S. Short-Term Weather Forecasting Using Spatial Feature Attention Based LSTM Model. *IEEE Access*. 2022;10:82456–82468. Available from: https://doi.org/10.1109/ACCESS.2022.3196381.
22) Nirgude V, Rathi S. Exploratory Data Analysis and Optimal Feature Selection on Pomegranate Dataset for Enhancing the Performance of Pomegranate Disease Prediction. *Annals of Forest Research*. 2022;65(1):746–764. Available from: https://www.e-afr.org/login/pdf/Vaishali%20Nirgude.pdf.
23) Doddaraju P, Kumar P, Gunnaiah R, Gowda AA, Lokesh V, Pujer P, et al. Reliable and early diagnosis of bacterial blight in pomegranate caused by Xanthomonas axonopodis pv. punicae using sensitive PCR techniques. *Scientific Reports*. 2019;9(1):1–9. Available from: https://doi.org/10.1038/s41598-019-46588-9.