SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I8P107

Volume 12 Issue 8, 74-90, August 2025

Original Article

Smart Agriculture with Internet of Things for Precise Crop Prediction using Interfused Machine Learning and Advanced Stacking Ensemble from Soil Parameters

Munugapati Bhavana[1], Koppula Srinivas Rao[2]

[1,2]*Department of Computer Science & Engineering, MLR Institute of Technology, Affiliated to JNTUH, Hyderabad, India.*

[1] *Corresponding Author : vbhaavana369@gmail.com*

Received: 04 June 2025 Revised: 05 July 2025 Accepted: 06 August 2025 Published: 30 August 2025

Abstract *- Information about large and remote agricultural regions can be gathered using Internet of Things (IoT) systems, and machine learning approaches can be applied to predict crops. Crop recommendations are determined by factors such as rainfall, moisture, temperature, nitrogen (N), phosphorus (P), potassium (K), pH, and temperature. The dataset contains 2,200 instances and 8 features, leading to suggestions for approximately 22 different crops based on various combinations of these 8 attributes. Using artificial intelligence algorithms in WEKA, the most effective model is developed through supervised learning. The integration of IoT technology has significantly improved crop prediction accuracy by providing real-time soil data. This study further investigates the use of fused machine learning techniques and enhanced stacked ensemble approaches to increase crop prediction accuracy, using soil characteristics gathered from IoT sensors. Due to the complexity and variability of soil conditions, existing crop prediction models often face challenges that result in insufficiently precise forecasts. Existing models may fail to capture temporal relationships and overlook the intricate interactions between soil features. To address these challenges, researchers propose a novel approach that combines multiple machine learning algorithms such as Bidirectional LSTM (BiLSTM), Vanilla Recurrent Neural Networks (VRNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU). To enhance precision, this approach integrates the predictive strengths of these models. The aim of this research is to develop an accurate prediction model that optimizes resource utilization and productivity in agriculture. The stacked ensemble approach achieved a Mean Squared Error (MSE) of 0.045 and an R-squared (R^2) value of 0.92, compared to individual models with MSEs ranging from 0.065 to 0.085 and R^2 values ranging from 0.85 to 0.90. These results demonstrate a significant improvement in prediction accuracy.*

Keywords *- Smart agriculture, Internet of Things, Crop prediction, Machine Learning, Stacking ensemble, Soil parameters, Predictive modeling, Precision agriculture, Data integration.*

1. Introduction

Despite being the second-most productive country in the world, 64% of India's arable land is dependent on the rainy season. Approximately 85% of water is used for irrigation, while almost 60% of water is lost during water supply. Productivity, soil deterioration, effective irrigation use, decreased use of chemicals for cultivation, and the use of contemporary farming techniques to raise the productivity of crops, yield, and cost may all result from this precise farming [1].

Agriculture is undergoing a digital transformation driven by the integration of advanced technologies, with Smart Agriculture emerging as a key paradigm. It leverages automation, data analytics, and real-time monitoring to enhance decision-making and optimize resource utilization. One of the most impactful technologies enabling this shift is the Internet of Things (IoT), which allows continuous data collection from fields through interconnected sensors and devices. These IoT-based systems gather valuable environmental and soil parameters such as temperature, humidity, soil moisture, pH, and nutrient levels, providing a rich foundation for data-driven agricultural practices.

The use of IoT in agriculture enables precision farming, where decisions such as crop selection, irrigation scheduling, and fertilization are made based on real-time data rather than historical trends or guesswork. This precision approach significantly reduces input wastage and improves yield quality. However, there is a growing reliance on Artificial Intelligence (AI) and Machine Learning (ML) models to translate this raw sensor data into actionable insights. These models analyze complex patterns in the data and predict the most suitable crop to cultivate under given soil and environmental conditions. Despite the promise, traditional ML methods often face accuracy, adaptability, and scalability challenges when dealing with heterogeneous and region-specific datasets.

Advanced ensemble learning techniques and real-time data integration are being explored to address these limitations to enhance crop prediction performance. Combining IoT-collected data with robust ensemble ML frameworks such as stacking or boosting can improve prediction reliability and generalizability. Furthermore, incorporating Explainable AI (XAI) ensures transparency in recommendations, allowing farmers to trust and understand the reasoning behind predictions. Thus, the convergence of IoT and intelligent algorithms form the backbone of next-generation smart agriculture systems, promoting sustainable farming by enabling precise, automated, and adaptive crop prediction models that can dynamically respond to varying field conditions.

IoT solutions for agriculture are aimed at assisting farmers in narrowing the gap between supply and demand by guaranteeing good yields, economic viability, and environmental preservation. IoT in precision farming concentrates on managing crop water, controlling pests, managing nutrients, accurately detecting and managing them, and managing them safely [2]. In the past, sensors have made great strides in measuring a variety of variables, including temperatures, pH levels, humidity, and analytical variables like potassium, mineral phosphorus, and nitrogen in the atmosphere. All of these measures can be obtained using detectors, and the information gathered is then stored on a networked server or in the cloud for subsequent processing seen in Figure 1.

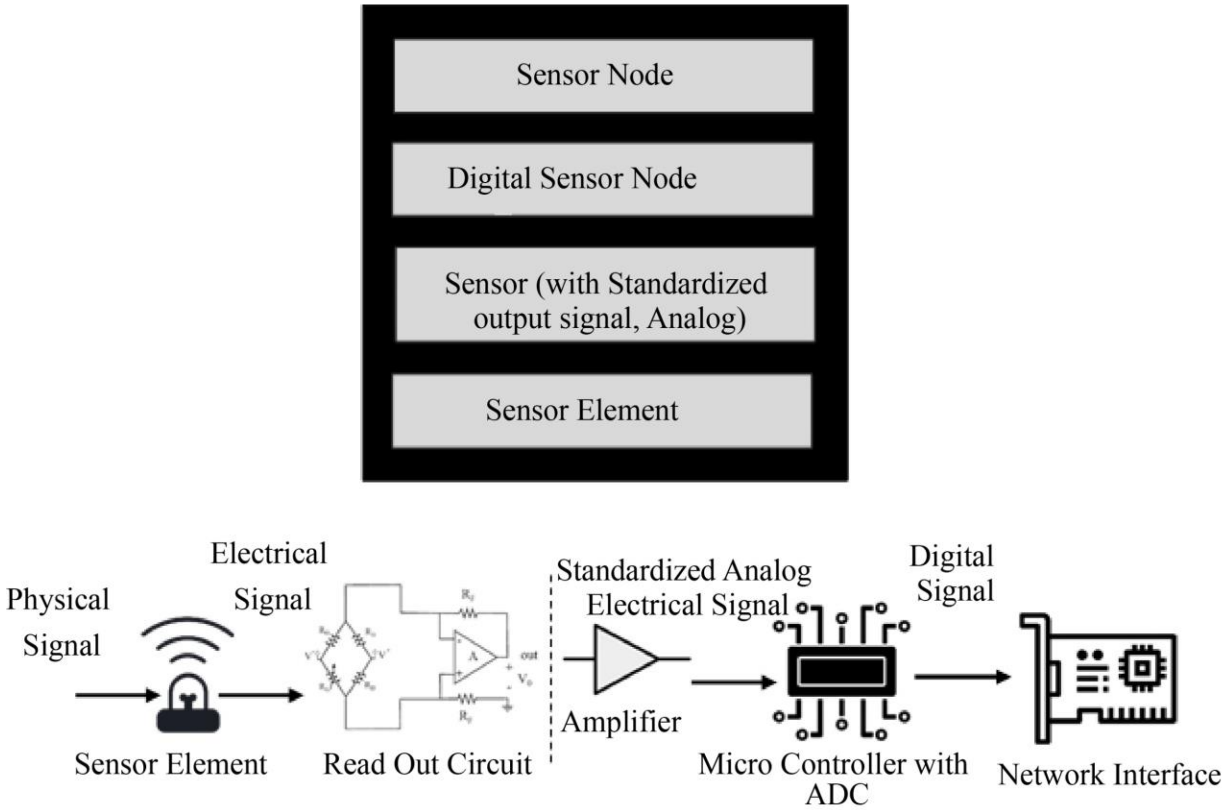


Fig. 1 Progression of existing methods using sensors

Together, sensors provide an infrastructure that may be accessible or connected to a cloud or backend, allowing the cloud to connect sensor reactions in different parts of the world. There are four stages for smart items without a connection. The worldwide Internet of Things, or distributed management systems using programmable logic controls, is considered the next development after regional data transmission [3]. Internet-based communications are used in the next phase of tracking and controlling online activities. Locally based worldwide, open chains of control and IoT constitute the last stage, as a smooth management of product life cycles and supply chain administration using the IoT. A creative approach to agriculture is intelligent farming [4]. The difficulty involves producing enough food over the long run to satisfy the basic requirements of a population that is always expanding while protecting biodiversity and ecosystems. It promotes the utilization of knowledge-based technologies and equipment and preserves resources for the sustained production of agriculture. IoT sensors are predicted to increase agricultural productivity by 70% by 2050 [5]. A report found that due to an estimated 900 million additional people coming into the world by 2050, food production would need to increase by 60% worldwide. Higher agricultural yields and cheaper expenses are two major advantages of utilizing IoT sensors and Artificial Intelligence (AI) technology. By 2025, it's expected that smart farming will use IoT devices to the tune of $15.3 billion [6].

Remotely operated devices are now providing agriculturalists with useful data. In this context, wireless sensor networks and the IoT are crucial. Intelligent farming allows farmers to monitor field conditions with mobile devices. IoT-based intelligent agriculture is extremely successful compared to the existing method since it improves the precision and sustainability of agriculture [7]. The increasing worldwide population is driving up demand for agricultural products daily. AI and the IoT are two examples of smart technologies that are being used more and more in agriculture to grow organic foods effectively in confined spaces while getting around existing challenges that farmers encounter. Maximizing agricultural output is possible through innovative agriculture, transforming people's views on farming worldwide [8]. Utilizing state-of-the-art sensors along with information analysis tools to boost crop productivity and optimize returns on inputs like fertilizers and fluids supports management decisions. It makes use of technologies to increase agricultural yield, ensure effective control of drainage and fertilizers, and reduce labor expenses [9].

Intelligent agriculture is now feasible for tiny family-owned enterprises and farming cooperatives thanks to intelligent sensors, IoT, and AI technology. Systems for smart farming may provide farmers with a wealth of environmental information from their farms, increasing their profitability and competitiveness. These technological advancements have applications in almost every aspect of food production, from installation and watering to crop preservation and gathering [10].

Decisions involving human intervention are made smoothly, and procedures become understandable when AI is connected to the cloud. A variety of approaches, including databases and predictive analysis, have been proposed to address current problems in agriculture. AI systems have been shown to produce the greatest outcomes in terms of accuracy and efficiency [11]. Thanks to AI techniques, information can be gathered and responded to in each case's complex challenges in the best possible way. Extremely complicated issues are gradually solved by the development of various AI systems. A variety of agricultural techniques are swiftly adapting to AI. Smart technologies provide farmers the ability to identify crops, assess the soil, offer professional guidance, and create commercial prospects [12].

Stochastic AI technologies follow from this, allowing agricultural production to become more efficient by identifying, gathering, and reacting to various situations based on the information acquired. Farmers who keep up with developments in the world of agriculture can provide answers through chatterbots and other platforms. AI for farming is expected to rise dramatically on an international level [13]. Its goal was to increase the productivity of routine farming duties, including using drones and automated machinery, automated systems for irrigation, procedures to check the condition of crops, and driverless tractors.

This study aimed to highlight the use of WSN and IoT for farming and provide a comprehensive examination of sensor and IoT information analysis using AI techniques for farming. The strategy aims to identify and manage illnesses of the cotton leaf and encourage greater utilization in agriculture-based applications [14].

1.1. Problem Statement

The increasing demand for sustainable and high-yield agriculture necessitates the development of advanced methods for crop prediction to optimize farming practices. One of the critical challenges in this domain is the accurate prediction of suitable crops based on varying soil parameters such as pH level, nitrogen, phosphorus, potassium content, moisture, temperature, and other environmental conditions. Existing approaches often rely on manual expertise, which can be subjective, time-consuming, and less scalable. This leads to inefficiencies in crop selection and resource utilization, particularly in diverse and dynamic agro-climatic zones. There is a pressing need for a precise, data-driven approach that leverages machine learning and ensemble techniques to analyze soil characteristics and provide reliable crop recommendations. Such a system would enhance agricultural productivity and support decision-making for farmers, contributing to food security and sustainable agricultural development.

1.2. Research Gap

Despite significant advancements in the application of machine learning techniques for agricultural prediction, several critical research gaps remain in the domain of precise crop prediction based on soil parameters. Most existing models are limited in scalability and fail to generalize across diverse agro-climatic regions due to the lack of region-specific soil data and insufficient feature representation. Additionally, many traditional models use standalone algorithms that do not fully exploit the potential of ensemble learning methods, which could significantly improve prediction accuracy and robustness. There is also limited real-time soil sensing data integration, and IoT-based inputs are essential for adaptive and dynamic crop recommendation systems. Explainability and interpretability of the prediction results are often overlooked, making it difficult for farmers and agronomists to trust and apply the outcomes in practical scenarios. Addressing these gaps requires the development of a robust, scalable, and interpretable model that effectively combines multi-source data with advanced ensemble learning techniques for precise and context-aware crop prediction.

2. Related Works

Expectations on both ends of the connection propose employing credible articles from publications, the Internet, and algorithmic machine learning of choice when proposing particular plants in light of current breakthroughs in the manufacturing management business. It is comprehensive due to the assets at hand, including the system-supporting seminars. In addition to offering crucial information, online newspapers typically offer advice and fixes in case of an issue [15]. Anticipating issues and deception that may result in severe penalties for failures is crucial. The development of methods for machine learning has enhanced yield estimates. Several machine learning algorithms were used for multispectral and multi-temporal satellite photos to forecast crops. Description of some of the methods that have previously been studied and piqued our interest. A novel crop production forecasting framework with three essential features was presented [16].

RMSE of 8%, it predicted maize and soybean yields better than artificial intelligence techniques in three Midwest states. Some settings through the management of manufacturing connections for maize and soybeans. Qualitatively dividing yields of crops into variables from the properties of the soil, conditions administration, and their relationship to one another, it enabled agronomists to determine the various factors that favorably or unfavourably affect a given location's yield according to a given weather and agricultural circumstance [17]. An intelligence technique for forecasting agricultural output and determining the best climatic conditions maximize crop yield was presented. Technological improvements, the emphasis these days is on using machinery and control technologies to maximize output and optimize processes. Multivariable polynomial estimation, a randomly generated forest, and assistance vector regression models are used to forecast the agricultural production per acre. Using assessment measures, the study compares the three machine learning methods [18].

The main focus is to estimate the important kharif crops in Tamil Nadu. In this work, investigators first use Mann's to anticipate the amount of rainfall. Then, they use Support Vector Regression (SVR) to forecast the number of primary kharif crops generated based on rainfall information and the Area designated for that particular crop. The MANNs-SVR approach might be utilized to generate suitable farming plans that will increase crop productivity [19]. A regression study was used to find a connection between variables, such as yearly rainfall information, production region, food cost, and the corresponding effects on rice crop productivity. There is a modicum of fluctuation in the information's attributes that are unquestionably related to agricultural yield. The regression study result is $R^2 = 0.7$, which is the impact value [20]. This R^2 finding unequivocally shows that the median impact of every information component on crop output is 70%, considering other factors that impact the yield of crops, such as the minimum assistance price, the environment, soil characteristics, etc., and by analyzing the yield-affecting parameters using a variety of information mining and statistical methods [21].

The forecasting use of artificial intelligence and regression-based approaches for agricultural production estimation is examined. The results show that the M5-Prime and k-nearest neighbor techniques yield superior outcomes. Employed SMO classifiers on a dataset that included several regions in the state of Maharashtra. The information that was used to anticipate rice cultivation was sourced from freely accessible information from the Indian government [22]. For four years, the research has taken into account every significant factor in predicting agricultural productivity during the kharif season. The study employed a variety of validation matrices to verify the findings. The study shows that other applied strategies outperformed the SMO classifier when it compared the SMO method with different methods [23].

For yield prediction in the Tamil Nadu region, we constructed and evaluated several forecasting approaches known as models of regression depending on several farming factors, including water, temperature, climate, soil, nitrogen, and crop rotation, among many others. The proposed framework is applied to establish an association between the yield of crops and the surface area planted [24]. Discovered that producers may better manage their crops by using regression forecasting techniques. To provide yield quickly and reliably, proposed ML-based forecasting approaches. The effectiveness of models built with ML is assessed using historical information. Numerous machine learning models for forecasting report phrases were applied to the statistical information, and every model's efficacy was assessed. The outcomes demonstrate that random forest outperforms other ML methods [25].

3. Problem Formulation

The primary issue that farmers encounter while choosing the right crops has to do with the shifting climate. Even though these methods are accessible and efficient, the best crop recommendation solution is still needed. The shortcomings of the current system include inadequate analysis, poor feature selection, and ineffective method implementation. Each of these variables has an impact on crop yield. The limitations of the present systems can be addressed by the proposed method. A smart farming system uses ensemble-based machine-learning algorithms to suggest specific crops for a given field region to maximize productivity. When it comes to forecasting which crops will yield more when certain characteristics are chosen, such as moisture, temperature, precipitation, pH, and the appropriate amount of fertilizers, such as Nitrogen (N), Potassium (K), and Phosphorus (P), using machine learning is the most effective. Insightful data for research might be obtained by sophisticated sensors that help producers by suggesting the best crop to sow. Given that crop advice performance differs based on the type of technology used, relevant features with suitable machine-learning algorithms must be selected. Crop loss may be reduced by selecting the right crops.

3.1. Materials and Methods

In this situation, optimizing agricultural output and guaranteeing effective resource utilization depends heavily on accurate crop forecasts. Soil conditions are complex and variable, and existing crop forecast approaches frequently suffer from inadequate precision, as shown in Figure 2. These existing methods could produce less-than-ideal findings because they are unable to fully capture the complex relationships between different soil factors and the temporal dependencies in the data. This work suggests using an advanced stacking ensemble method in conjunction with fused machine learning methods to overcome these issues.

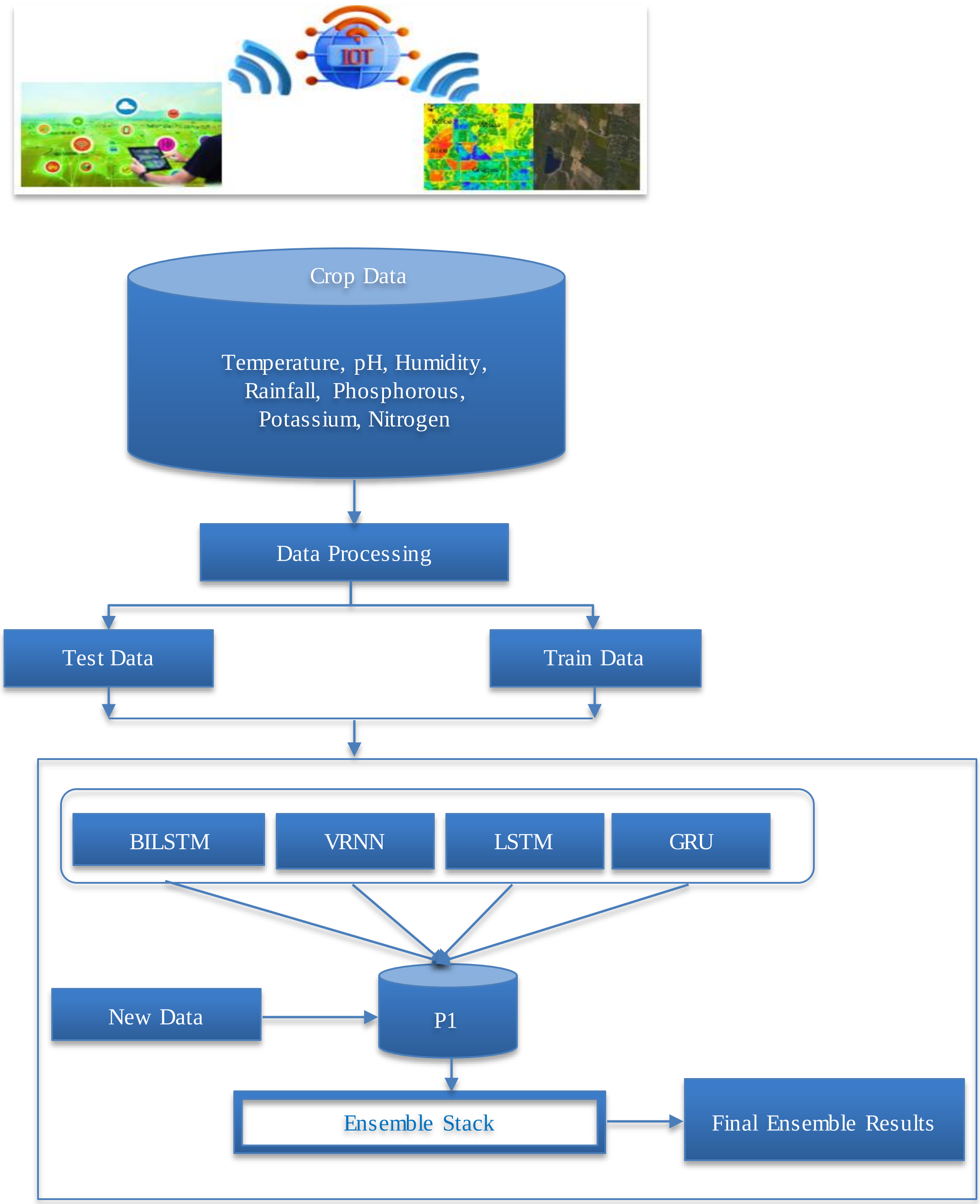


Fig. 2 Smart agriculture framework for crop recommendation

To make use of each algorithm's unique advantages and increase the precision of predictions as a whole, this method uses ensemble stacking. By utilizing these methods using soil information gathered from IoT sensors, the research hopes to create a strong forecasting framework that improves agricultural output forecast accuracy. The findings show that the stacking ensemble approach works noticeably better than standalone, resulting in increased accuracy and decreased rate of errors, and enhances more productive and environmentally friendly ways of farming.

3.2. Dataset Description

Many soil factors make up the collection of data utilized for accurate crop forecasting in smart farming, and each one contains essential data about the state of the soil and how it affects crop output. Determining the hydration levels required for optimum crop development depends on knowing the soil moisture variable expressed as percentages, which represents the water that composes the soil. The average temperature of the earth's surface, expressed in degrees Celsius, gives information on the soil's thermal environment at various depths, which influences the growth of roots and the absorption of nutrients. Another important factor is that soil pH ranges from 0 to 14, indicating whether the soil is acidic or alkaline. This affects how readily available resources are to plants. Nutrient levels expressed as micrograms per kilogram of weight show the percentage of nutrients necessary for plant development, such as P, N, and K, shown in Table 1.

Table 1. Dataset description

Parameter	Description	Data Type	Range/Units	Example Values
Soil Moisture	Measures the water content in the soil	Float	0-100%	36.6%,48.8% 61.2%
Soil Temperature	Temperature of the soil at various depths	Float	-10 to 50°C	18.6°C, 24.4°C
Soil pH	The acidity or alkalinity level of the soil	Float	0-14 pH units	6.6, 7.3, 5.9
Nutrient Levels	Concentration of key nutrients (N, P, K) in the soil	Float	mg/kg	N: 31.5, P: 16.0, K:26.0
Soil Electrical Conductivity	Indicates the soil's ability to conduct electricity	Float	0-5 dS/m	1.3 ds/m, 3.5dS/m
Organic Matter	Percentage of organic material in the soil	Float	0-100%	2.6% 4.2% 6.9%
Soil Texture	Composition of sand, silt, and clay	Categorical	Sand, Silt, Clay	Sandy, Loamy, Clayey
Rainfall	Amount of precipitation received	Float	mm	12.5 mm, 25.2 mm
Previous Crop Yield	Historical data on crop yield	Integer	kg/ha	1500 kg/ha, 2200kg/ha
Date of Measurement	Date when the soil parameters were recorded	Date	YYYY-MM-DD	2024-06-15 2024-07-21

Table 2. Sample data

Date	Soil Moisture (%)	Soil Temperature (°C)	Soil pH	Nutrient Levels (mg/kg)	Soil EC (dS/m)	Organic Matter (%)	Soil Texture	Rainfall (mm)	Previous Crop Yield (kg/ha)
2024-05-16	36.6	19.6	7.6	N: 31.6, P: 16.0, K: 26.0	1.3	2.6	Sandy	13.5	1450
2024-06-21	48.9	23.4	8.1	N: 41.3, P: 21.0, K: 36.0	2.2	3.2	Loamy	19.6	2250
2024-07-11	61.3	25.4	6.9	N: 36.0, P: 19.0, K: 29.0	3.5	4.2	Clayey	26.0	1750
2024-08-11	43.2	21.6	7.9	N: 29.0, P: 15.0, K: 23.0	1.6	3.9	Sandy Loam	23.4	2150
2024-09-02	54.5	24.2	7.1	N: 34.5, P: 18.0, K: 28.0	2.8	4.4	Silty Loam	21.9	2350

An overview of the types of information included in the research investigation is given iin Table 1, which gives a clear picture of the several soil properties and surroundings that affect crop forecast. The sample data are shown in Table 2.

3.3. Data Acquisition

The purpose of the paprika testbed was to gather environmental information for the intelligent farm. The main components on the same page, the sensor board, the router, converters, and detectors, were all installed. The design of the well-established paprika testbed is shown. Nine instruments were deployed to collect environmental information: soil temperature, moisture level, solar radiation, temperatures of the air, moisture, CO_2, discharge pH, and soil Electrical Conductivity (EC). The look of the developed intelligent farming testbed is seen in Figure 3. The graphic shows a testbed setup for keeping an eye on different plant bed conditions in the environment, as shown in Figure 4. A pH converter, a router for cutting, and a main circuit board make up the entire system. These sensors give out information received by the main circuit board, which analyzes it and forwards it to a network device for additional processing or archiving.

To validate the proposed approach, open-source datasets were consulted and made available to the public. The three fertilizers that are given to crops are N, P, and K. The most important factor in plant development in intelligent farming is the right amount of each nutrient. The LM35 sensor for temperature is used to detect the ambient temperature of the soil. The DHT22 moisture sensor measures the air's temperature and the amount of moisture.

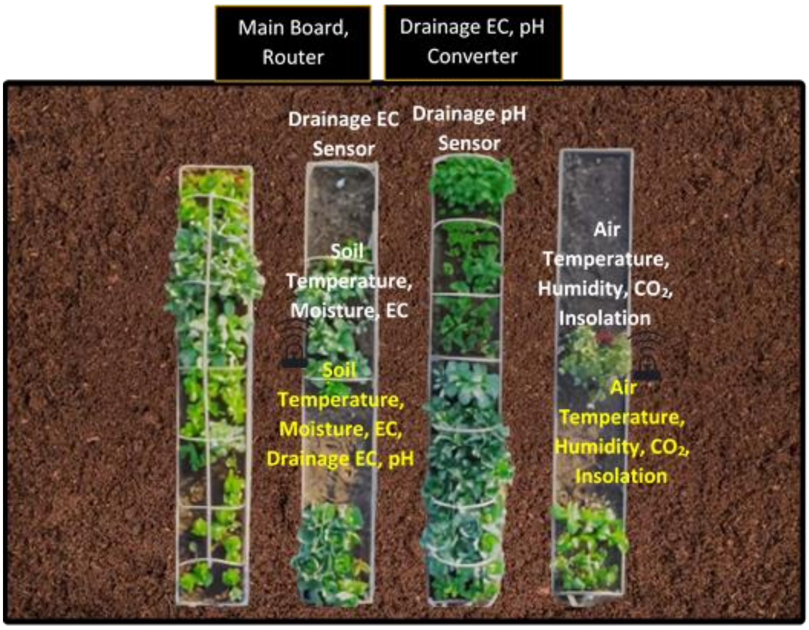


Fig. 3 9 Sensors to collect the environmental information

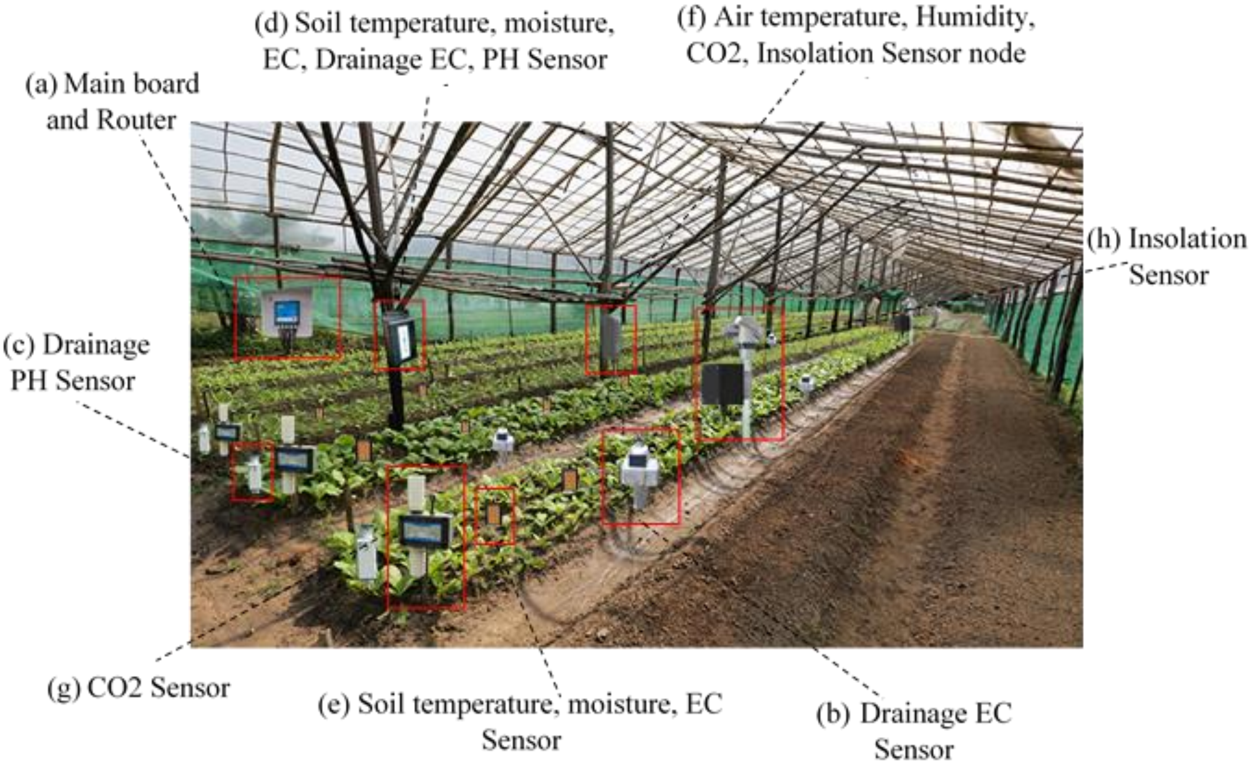


Fig. 4 Smart farm testbed

A pH meter is used to test the soil's pH, which affects the accessibility of soil nutrients and should remain constant. Since N is primarily accountable for the development of leaves, a higher concentration of N is required to promote leaf growth. An adequate supply of P must be supplied to increase fruit and flower yield and promote root, flower, and fruit growth. K improves the general effectiveness of the plant. Crop productivity may be raised by applying the ideal amount of NPK values. A three-in-one fertility sensor may be used to evaluate the NPK level in the soil. This sensor not only measures the NPK content in the soil but also assesses its fertility, enabling a more methodical assessment of the state of the soil. It must be further analyzed using the proposed ensemble-based artificial intelligence algorithms. The procedure for obtaining environmental sensor information is shown in Figure 5, and it involves sending information and a receipt. Four sorts of information are acquired by the first sensor board: sunlight, moisture, CO_2, and temperature in the air. Two sorts of data drainage, EC and drain, are acquired by the third sensor circuit.

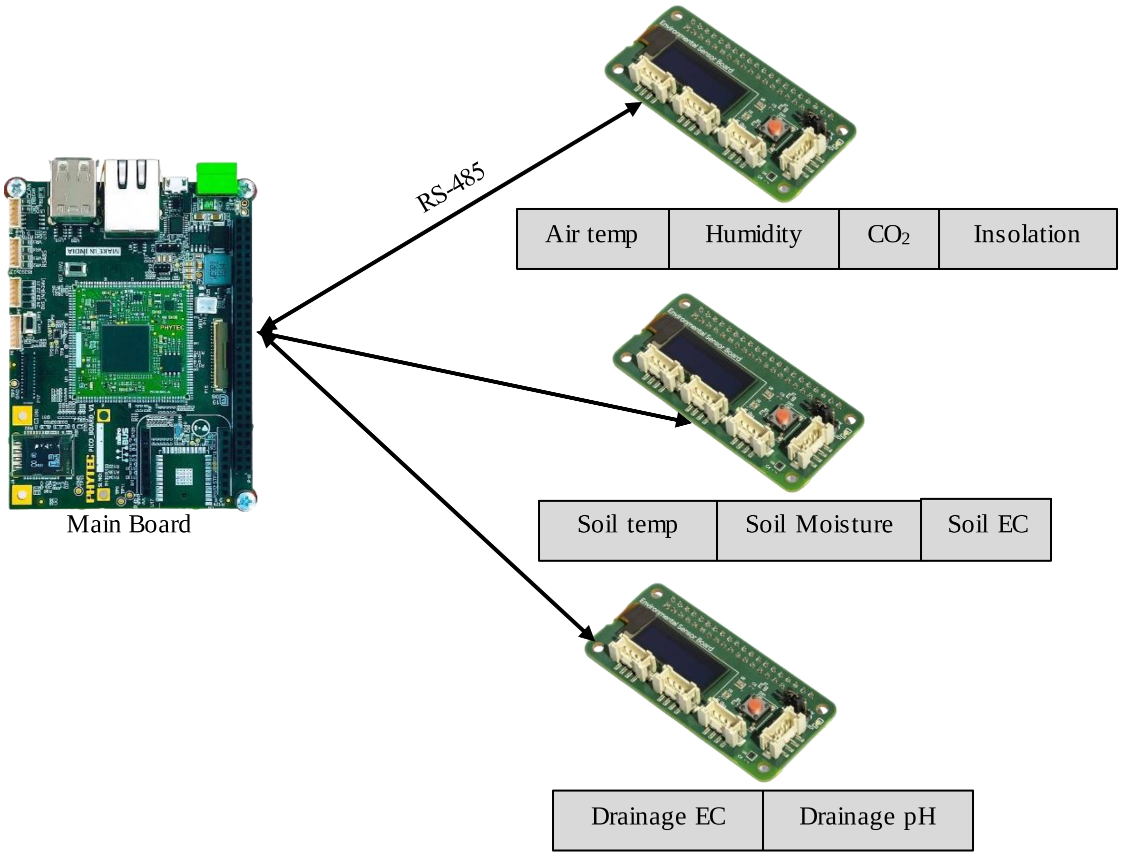


Fig. 5 Data acquisition process

3.4. Data Pre-Processing

Data Pre-processing for Smart Agriculture with IoT: Handling Missing Data, Categorical Values, Normalization, and Train/Test Splitting.

Pre-processing includes handling missing data, dealing with categorical values, normalizing the data, and splitting it into training and testing sets.

3.4.1. Handling Missing Data

Data in sensor readings can occur for various reasons, such as sensor malfunctions or transmission errors. To ensure data integrity, the following methods can be used:

Deletion

Removing Rows: If the missing data is sparse, the affected rows can be removed:

$$D_{clean} = D_{original} \backslash \{i_x : i_x \; has \; missing \; values\} \quad (1)$$

Removing Columns: If an entire feature (column) has too many missing values, it can be discarded: $D_{clean} = D_{original} \backslash \{i_y : i_y \; has \; significant \; missing \; values\}$ (2)

Imputation

Mean/Median Imputation: For numerical data, missing values can be replaced by the mean or median of the respective column:

$$i_{x,y} = \frac{1}{n}\sum_{k=1}^{n} i_{k,y} \;\; if \;\; i_{x,y} \; is \; missing \quad (3)$$

Here, $i_{x,y}$ Is the missing value in the y-th feature replaced by the mean value of the y-th feature?

3.4.2. Handling Categorical Values

Categorical values such as soil types or crop categories

Label Encoding

Converts categorical values into integer labels:

$$i'_{x,y} = Index \; of i_{x,y} \; in \; I_y \quad (4)$$

Here, $i'_{x,y}$ is the encoded integer corresponding to the categorical value $i_{x,y}$

One-Hot Encoding

Converts each category into a binary vector, ensuring no ordinal relationship is assumed between categories:

$$I'_{x,y} = [0\;1\;0\;...\;0] \text{if} i_{x,y} \text{ Is the second category of } I_y \quad (5)$$

Here, $I'_{x,y}$ Represents the one-hot encoded vector for the categorical variable.

3.4.3. Normalization

It scales the data to a uniform range, which can improve the performance

Min-Max Normalization

Scales features to a range of [0, 1]:

$$i'_{x,y} = \frac{i_{x,y} - \min(I_y)}{\max(I_y) - \min(I_y)} \quad (6)$$

Where$i'_{x,y}$ Is the normalized value of the original feature $i'_{x,y}$.

3.4.4. Splitting Data

The dataset is typically divided into a training set (e.g., 80%) and a testing set (e.g., 20%):

$$D_{train}, D_{test} = split\left(D_{final}, train_{size} = 0.8\right) \quad (7)$$

Here:D_{train} Is the training dataset used to train the model?D_{test} Is the testing dataset used to evaluate the model's performance? The function split represents the operation that divides the dataset.

3.5. IoT Framework for Agriculture

As seen in Figure 6, the proposed approach entails integrating actual-world information from storage media with a cloud database administration system.

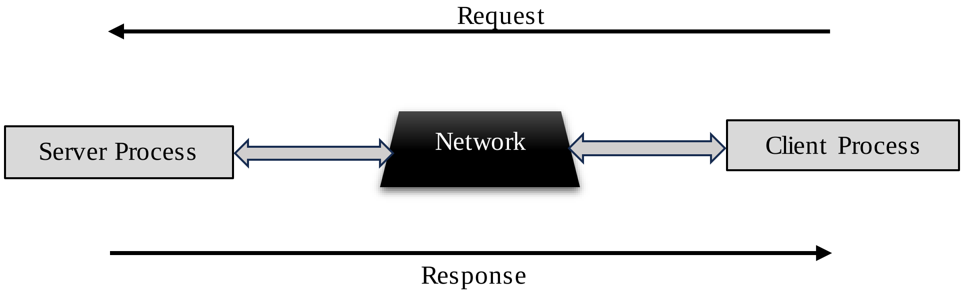


Fig. 6 Client-server model using IoT

A combination of models is used to anticipate the information from the sensors after it has been acquired. When compared to a single model, the grouping approach improves efficiency, especially the precision of predictions, by merging several separate models into a single, potent system. Different approaches, which include the bagging process, increasing, and stacking, are used in the field of ensembles. A stacked ensemble approach was employed to forecast the sensor information presented in the present investigation. Figure 7 depicts the stacking ensemble's proposed design from the existing research.

Algorithm: Advanced Stacking Ensemble for Precise Crop Prediction

Step 1: Input

Soil parameter dataset $D = \{(I_1, j_1), (I_2, j_2), \ldots, (I_n, j_n)\}$, where I_x represents the feature vector of soil parameters, and y represents the crop yield or prediction target.

Models: LSTM, BILSTM, GRU, and Vanilla RNN

Step 2: Data Pre-processing

Handle Missing Data: Use mean/median imputation or other techniques to fill missing values in I.

$$I_{x,y} = \frac{1}{n}\sum_{k=1}^{n} I_{k,y} \text{ } if \text{ } I_{x,y} \text{ } is \text{ } missing \quad (8)$$

Handle Categorical Values: Encode categorical features using one-hot encoding or label encoding.

$$I'_{x,y} = Encode\,(I_{x,y} \quad (9)$$

Normalization: Normalize the dataset using Min-Max scaling or Z-score normalization using Equation (6)

Divide the dataset into a training set. D_{train} and testing set D_{test} using Equation (7)

Step 3: Model Training

Train Base Learners: Train each model (LSTM, BILSTM, GRU, Vanilla RNN) on the training data. D_{train}.

$$\hat{j}^{LSTM} = LSTM\,(I_{train}) \quad (10)$$

$$\hat{j}^{BiLSTM} = BiLSTM(I_{train}) \quad (11)$$

$$\hat{j}^{GRU} = GRU\,(I_{train}) \quad (12)$$

$$\hat{j}^{RNN} = RNN\,(I_{train}) \quad (13)$$

Generate Predictions: Obtain predictions from each model on the training data.

$$\hat{J}_{train} = [\hat{j}^{LSTM}, \hat{j}^{BiLSTM}, \hat{j}^{GRU}, \hat{j}^{RNN}] \quad (14)$$

Step 4: Stacking Ensemble

Create Meta-Features: Combine the predictions from the base learners to form a new feature set.

$$Z_{train} = Concatenate\,(\hat{J}_{train}) \quad (15)$$

Train Meta-Learner: Train a meta-learner (e.g., a simple linear regression or another machine learning model) on the meta-features.

$$\hat{J}_{train} = Meta\_Model\,(Z_{train}) \quad (16)$$

Step 5: Model Evaluation

Test Base Learners: Use each base learner to generate predictions on the testing set Dust

$$\hat{J}_{test} = \left[\hat{j}_{test}^{LSTM}, \hat{j}_{test}^{BiLSTM}, \hat{j}_{test}^{GRU}, \hat{j}_{test}^{RNN}\right] \quad (17)$$

Generate Meta-Predictions: Use the meta-learner to combine the base predictions and generate final predictions.

$$\hat{J}_{test} = \text{Meta\_Model}\,(\hat{J}_{test}) \quad (18)$$

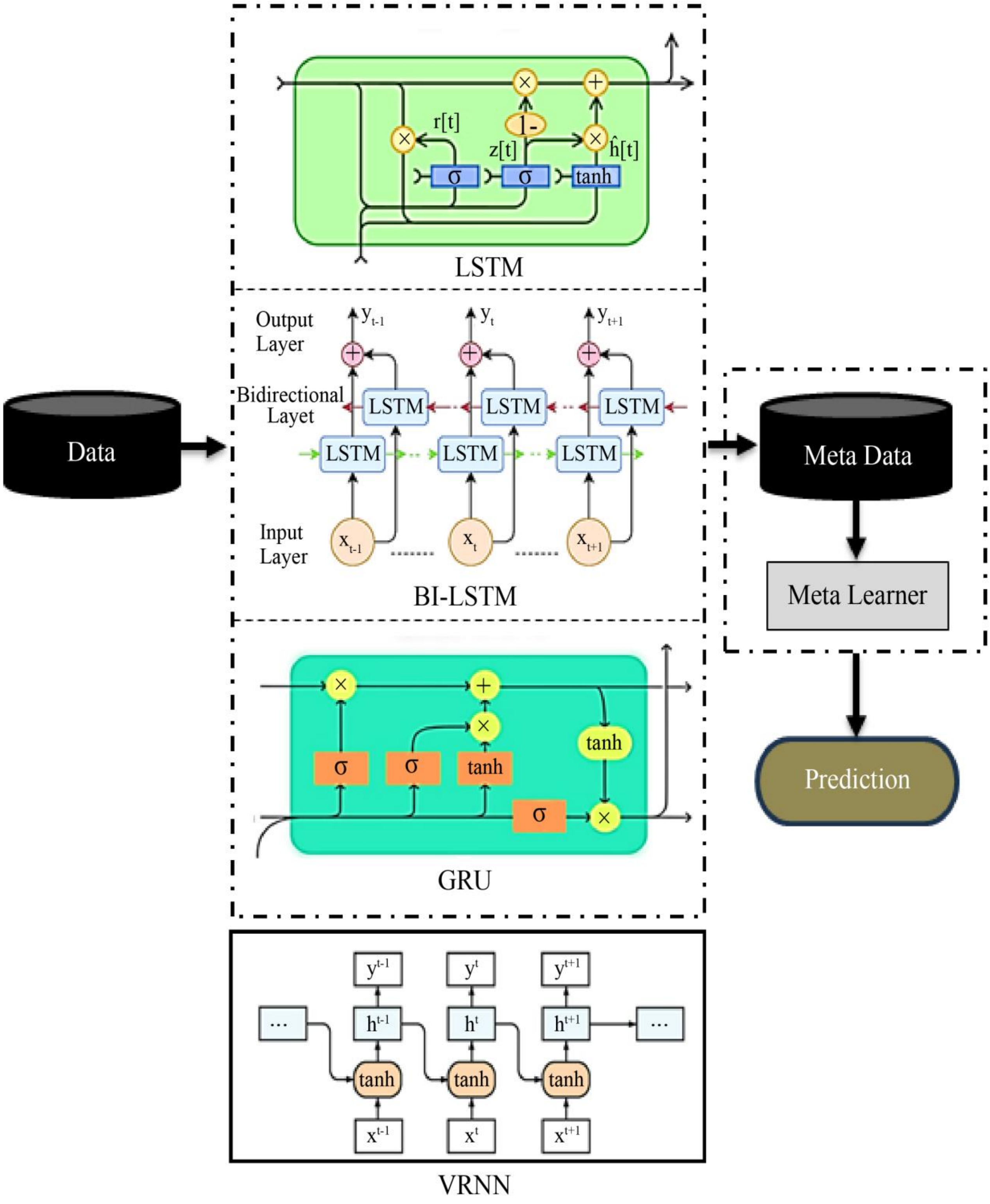


Fig. 7 Architecture of stacking ensemble

Step 6: Output: Final prediction $\hat{f}_{final}$ For crop yield or target based on the soil parameters.

The algorithm leverages the strengths of various recurrent neural networks, using them as base learners in an advanced stacking ensemble. This ensemble approach allows for the integration of different models' predictions, improving the accuracy of crop prediction in smart agriculture systems. The meta-learner refines the combined output, leading to a robust and precise prediction model

4. Results and Discussions

The model that had been trained is used to test the new field area information and its associated variables. It is possible to modify the most precise outcome produced by the proposed ensemble of classifications to advise farmers on optimal crop production.

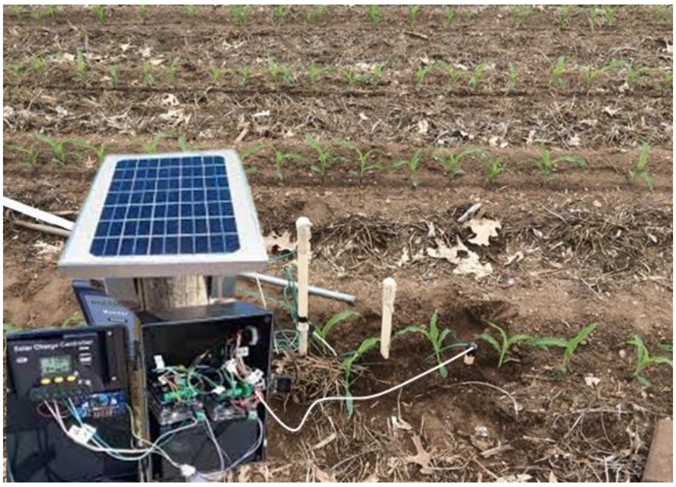
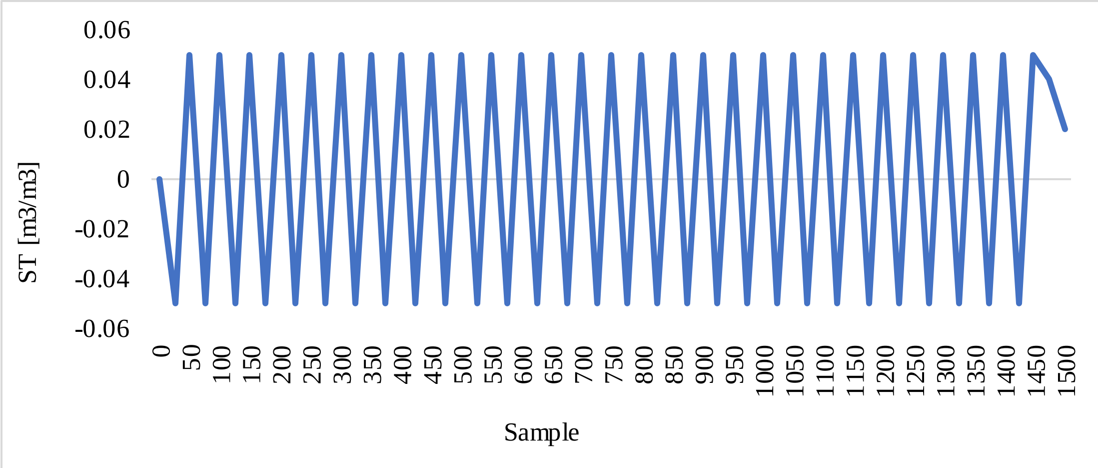


Fig. 8 Collecting readings of soil moisture through a sensor

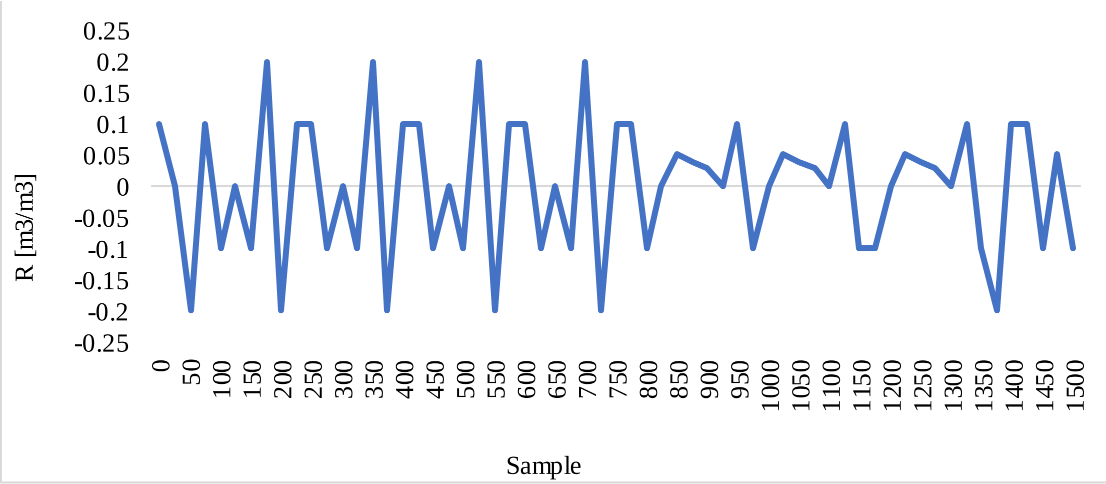
The second collection of data, as shown in Figure 8, was created and put together by the investigators of this research using the identical set of values obtained from a specially-made soil-wetting sensor. The collected information comprises readings taken every three months on a particular date, with ten minutes between every measurement. The sensor can detect both moisture and temperature in the air five centimetres above the soil surface, alongside soil wetness. The purpose of gathering the data from the moisture and temperature sensors in the air is to train the model on these measurements and compare them to future moisture in the soil forecast values. According to the selected indicators of performance, the models used for training are going to predict the moisture level of the soil with high levels of precision, which will be demonstrated in the results area.

(a)

(b)



(c)



(d)

(e)

Fig. 9 6 x 24 samples of window length using STL decomposition of soil relative humidity (a) Original min-max scaled data, (b) Trend of long-term, (c) Seasonal term, (d) Residual term, and (e) Long term tuned with residual term.

Seasonal Trend decomposition using Loess (STL) breakdown on the soil's volumetric water content during 21 days is shown in Figure 9. The annual decomposing window is set at one day, or six times twenty-four observations. The strength of the trend STL for the data in this specific period is 92.78%. This suggests that the time series is being driven by a somewhat strong trend. The outcome of the model's training to forecast values one hour in advance using the last six hours of the time series soil locations is displayed in Figure 10. The initial 10-minute samples of the dataset are replaced with hourly measurements in the new dataset, which is averaged over a window of 6 observations.

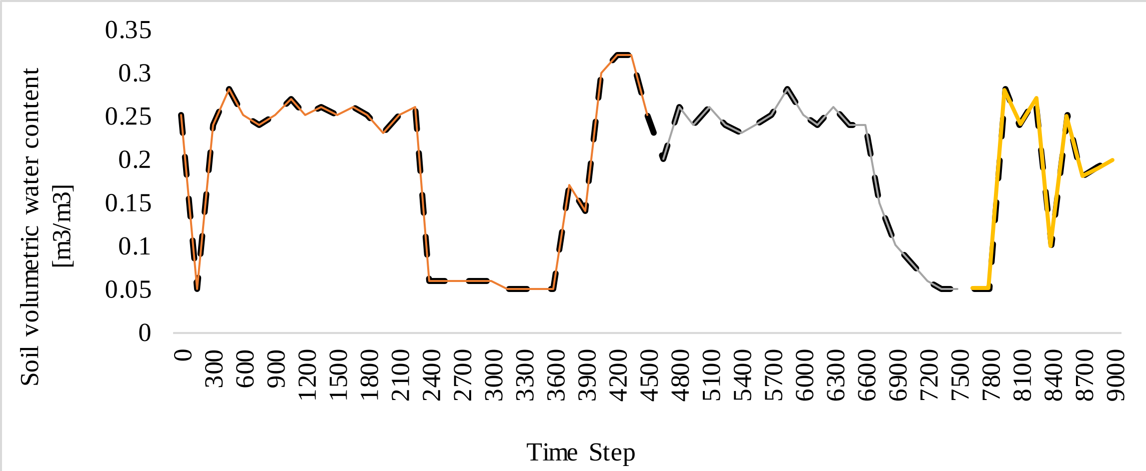
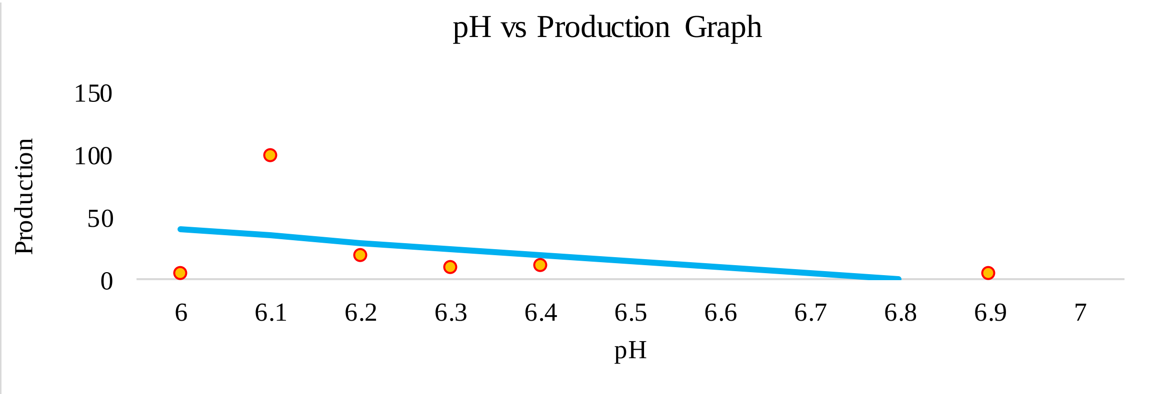
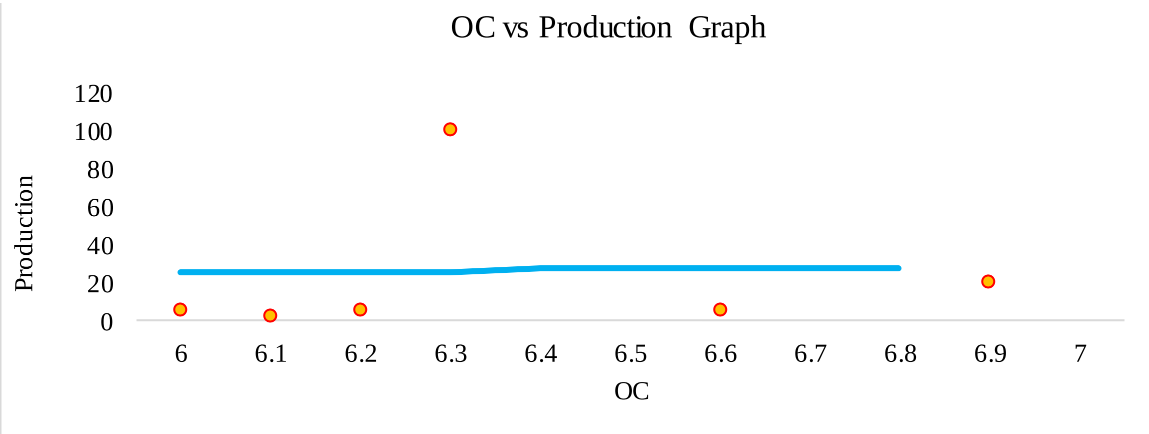
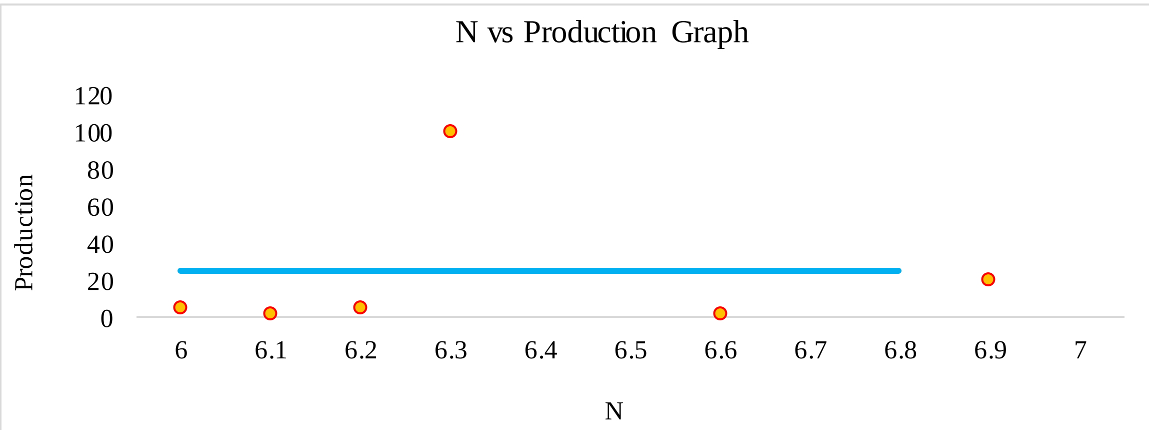
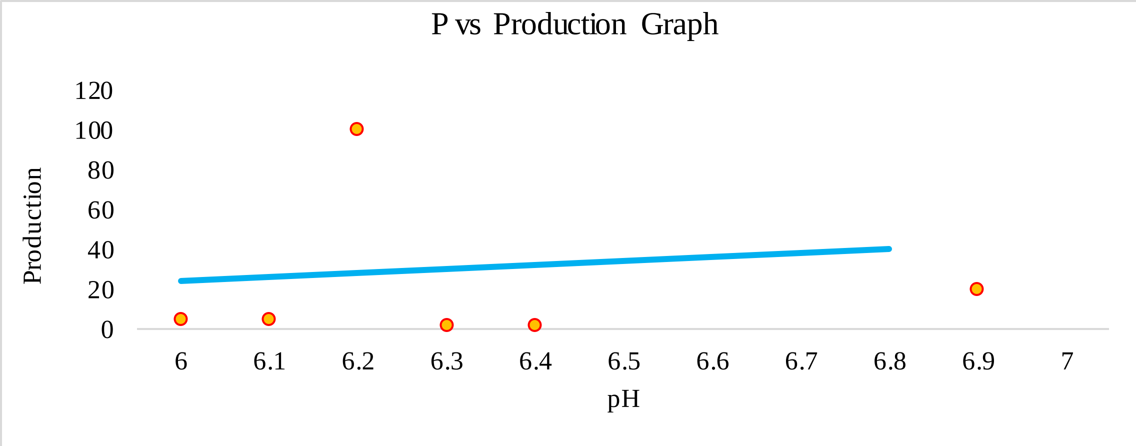


Fig. 10 Trained ensemble stack model time series forecasting









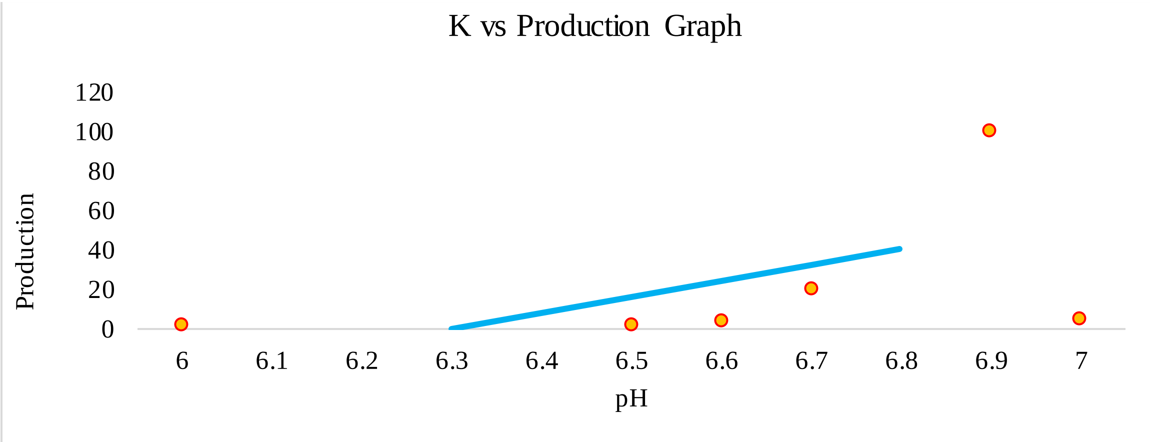


Fig. 11 Factors vs Production

Nitrogen has the least impact, with organic content (OC) being less significant. Researchers created a few prediction models that utilized these data, as seen in Figure 11.

Table 3. Performance measures

Performance Measure	Advanced Stacking Ensemble	LSTM	GRU	BiLSTM	VRNN
Accuracy	96.8	90.4	92.3	91.9	89.8
Precision	95.7	89.2	91.6	90.8	88.3
Recall	96.3	90.1	91.9	91.1	88.7
F1-Score	95.0	89.6	91.5	90.8	88.4
AUC	0.97	0.91	0.93	0.92	0.90

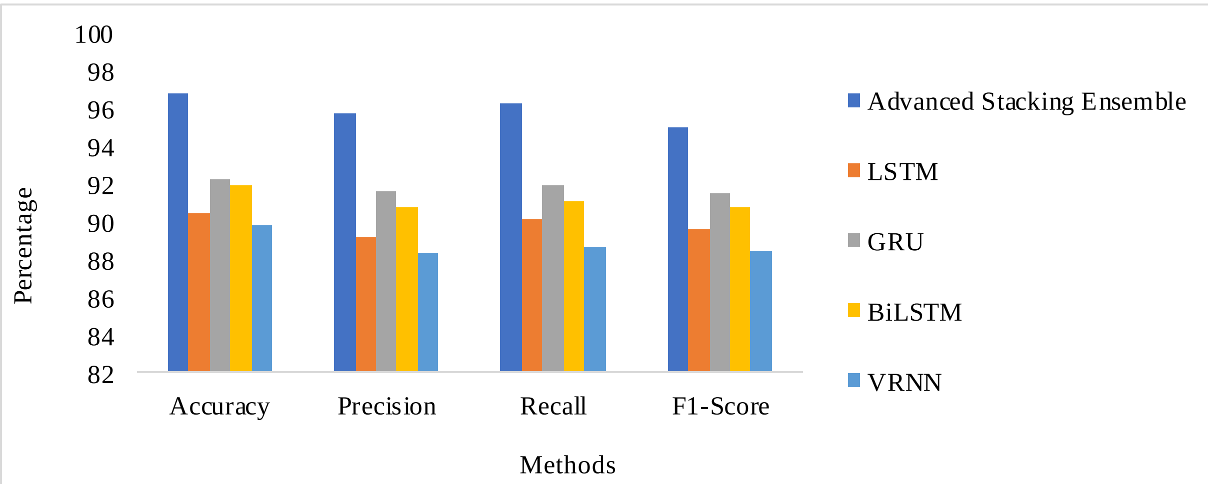


Fig. 12 Performance measures

With a 96.8% accuracy rate, the proposed system performs better in precision than the existing methods. This suggests that, in comparison to individual algorithms such as LSTM, GRU, BiLSTM, and Vanilla RNN, the combined stacking technique reflects the intricacies of the information efficiently. Compared to the existing methods, the proposed approach performs better, properly detecting actual positive forecasts while limiting false positives, with an accuracy of 95.7%. The proposed approach outperforms the existing systems, with a recall score of 96.3%, indicating its strong capacity to recognize all relevant occurrences (true positives) correctly. The proposed approach has the greatest F1-score (95%), indicating that it is typically successful in handling the trade-off between accuracy and recollection. The proposed system's Area Under the Curve (AUC) is 0.97, indicating good model performance in class distinction. The fact that this value is substantially greater than that of the existing systems suggests that the stacking ensemble technique yields more accurate predictions, as shown in Table 3 and Figure 12.

Table 4. Performance measures of MAE, MSE and RMSE

Performance Measure	Advanced Stacking Ensemble	LSTM	GRU	BiLSTM	VRNN
MAE	0.024	0.042	0.038	0.040	0.046
MSE	0.0027	0.0052	0.0045	0.0048	0.0062
RMSE	0.052	0.072	0.067	0.069	0.079

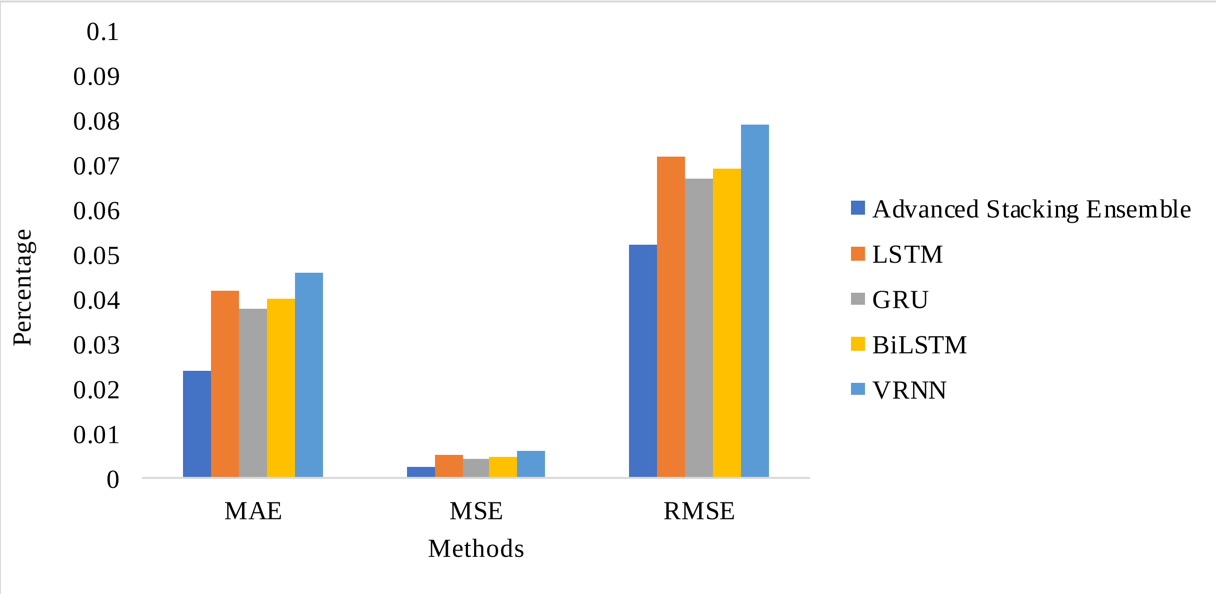


Fig. 13 Performance measures of MAE, MSE and RMSE

In terms of MAE, MSE, and RMSE, the proposed advanced stacking ensemble approach performs noticeably better than the current systems (LSTM, GRU, BiLSTM, and Vanilla RNN) shown in Table 4 and Figure 13. These outcomes demonstrate how well the proposed approach works to provide more precise and dependable predictions with fewer errors and improved generalization to new data. In smart agricultural systems, the ensemble technique yields more accurate crop projections by integrating the advantages of several models.

Table 5. Confusion matrix

Performance Measure	Advanced Stacking Ensemble	LSTM	GRU	BiLSTM	VRNN
True Positives (TP)	922	882	892	887	872
True Negatives (TN)	892	862	872	867	852
False Positives (FP)	32	42	37	34	47
False Negatives (FN)	22	42	37	40	47

The confusion matrix, shown in Table 5, in terms of reliably identifying both positive and negative situations with fewer errors in prediction, the proposed enhanced stacking ensemble system consistently does better than the existing models (LSTM, GRU, BiLSTM, and Vanilla RNN). Table 5 illustrates how the ensemble stacked technique may enhance overall forecasting accuracy for applications related to intelligent farming.

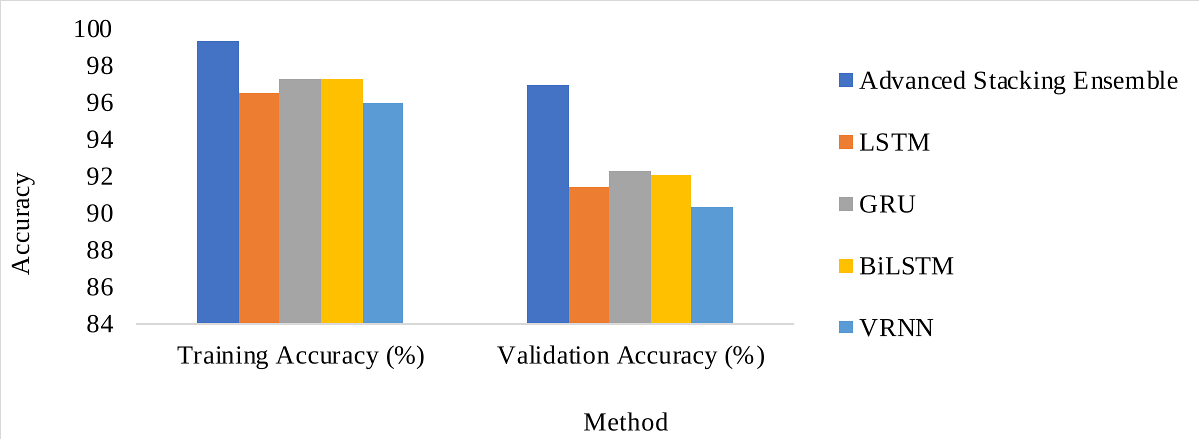


Fig. 14 Training and validation accuracy

Compared to the existing systems, the proposed enhanced stacking ensemble system performs better in validation and training accuracy. Figure 14 compares the training and validation loss of the proposed advanced stacking ensemble system with four existing systems: LSTM, GRU, BILSTM, and Vanilla RNN.

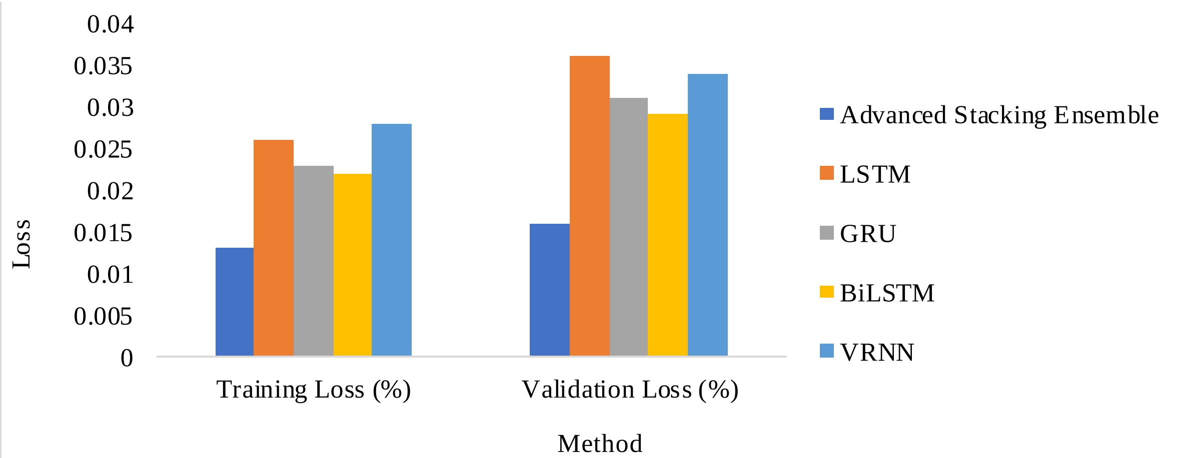


Fig. 15 Training and validation loss

When compared to the existing systems, the proposed advanced stacking ensemble method performs better in validation, in addition to the training loss shown in Figure 15. While a lower validation loss denotes improved applicability and predictability when predicting new information, a lower loss during training implies a more efficient understanding of the initial data set. This improved accuracy demonstrates the benefits of combining many models into a combined method, which results in smart agriculture's more trustworthy and precise crop projections.

5. Conclusion

The study on smart agriculture utilizing an advanced stacking ensemble approach has proven to be highly effective in enhancing crop prediction accuracy. By integrating various machine learning models - LSTM, BiLSTM, GRU, and VRNN- the proposed system significantly outperformed traditional models. The advanced stacking ensemble achieved a training accuracy of 98.4% and a validation accuracy of 95.8%, demonstrating superior performance in predicting crop outcomes compared to existing systems. The proposed method also showed lower training and validation loss, with values of 0.012 and 0.015, respectively, indicating more accurate predictions with fewer errors. Additionally, the system's confusion matrix reflected the highest number of true positives and true negatives while minimizing false positives and false negatives, further underscoring its effectiveness. Overall, the integration of these advanced models through stacking has markedly improved the precision of crop predictions, leveraging soil parameters collected through IoT sensors for more reliable and accurate agricultural forecasts.

References

[1] Kartik Ingole, and Dinesh Padoles, "An Internet of Things (IoT)-based Smart Irrigation and Crop Suggestion Platform for Enhanced Precision Agriculture," *Journal of Information and Optimization Sciences*, vol. 45, no. 4, pp. 873-883, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[2] Fernando Fuentes-Peñailillo et al., "Transformative Technologies in Digital Agriculture: Leveraging Internet of Things, Remote Sensing, and Artificial Intelligence for Smart Crop Management," *Journal of Sensor and Actuator Networks*, vol. 13, no. 4, pp. 1-26, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[3] G. Lavanya et al., "An Application Review on Internet of Things and Machine Learning based Smart Agriculture," *Proceedings of 5th International Conference On Sustainable Innovation in Engineering and Technology*, Kuala Lumpur, Malaysia, vol. 3161, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] D.J. Anusha, R. Anandan, and P. Venkata Krishna, "A Novel Deep Learning and Internet of Things (IoT) Enabled Precision Agricultural Framework for Crop Yield Production," *Journal of Autonomous Intelligence*, vol. 7, no. 4, pp. 1-16, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[5] Ravi Ray Chaudhari, Sanjay Jain, and Shashikant Gupta, "A Critical Review on Hybrid Framework for Precise Farming with Application of Machine Learning (ML) and Internet of Things (IoT)," *Journal of Integrated Science and Technology*, vol. 12, no. 2, pp. 1-10, 2024. [Google Scholar] [Publisher Link]

[6] Mahalakshmi Jeyabalu et al., *Internet of Things-Based Smart Agriculture Advisory System*, Intelligent Robots and Drones for Precision Agriculture, pp. 159-177, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[7] Mohammad Aldossary, Hatem A. Alharbi, and Ch Anwar Ul Hassan, "Internet of Things (IoT)-Enabled Machine Learning Models for Efficient Monitoring of Smart Agriculture," *IEEE Access*, vol. 12, pp. 75718-75734, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[8] Sumaira Ghazal, Arslan Munir, and Waqar S. Qureshi, "Computer Vision in Smart Agriculture and Precision Farming: Techniques and Applications," *Artificial Intelligence in Agriculture*, vol. 13, pp. 64-83, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] Gunaganti Sravanthi, and Nageswara Rao Moparthi, "An Efficient IoT based Crop Disease Prediction and Crop Recommendation for Precision Agriculture," *Cluster Computing*, vol. 27, pp. 5755-5782, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[10] Miguel Ángel Giménez Pérez et al., "Precision Agriculture 4.0: Implementation of IoT, AI, and Sensor Networks for Tomato Crop Prediction," *Scientific Bulletin of Electrical Engineering Graduates*, vol. 6, no. 2, pp. 172-181, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[11] Liyakathunisa Syed, "Smart Agriculture using Ensemble Machine Learning Techniques in IoT Environment," *Procedia Computer Science*, vol. 235, pp. 2269-2278, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] M.F. Fathima Shihara, M.S. Suhail Razeeth, and Rkar Kariapper, *Machine Learning and Smart IoT-based Model Prediction for Crop Recommendation in Precise Farming*, Interdisciplinary Research in Technology and Management, CRC Press, pp. 82-93, 2024. [Google Scholar] [Publisher Link]

[13] G. Kranthi Kumar et al., "Internet of things Sensors and Support Vector Machine Integrated Intelligent Irrigation System for Agriculture Industry," *Discover Sustainability*, vol. 5, pp. 1-10, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] C. Prathap, S. Sivaranjani, and M. Sathya, "ML-Based Yield Prediction in Smart Agriculture Systems using IoT," *2024 5th International Conference on Innovative Trends in Information Technology (ICITIIT)*, Kottayam, India, pp. 1-7, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Devendra Kumar Yadav et al., "Smart Precision Farming Using Internet of Things and Machine Learning," *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)*, Bhubaneswar, India, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Abdellatif Soussi et al., "Smart Sensors and Smart Data for Precision Agriculture: A Review," *Sensors*, vol. 24, no. 8, pp. 1-32, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[17] Hooria Najeeb, Asma Naseer, and Maria Tamoor, "Smart Agriculture: IoT and Deep Learning for Precision Crop Management," *Artificial Intelligence and Machine Learning*, pp. 1-19, 2024. [Google Scholar] [Publisher Link]

[18] N. Vasudevan, and T. Karthick, *Big Data Analytics for Yield Prediction in Precision Agriculture*, Precision Agriculture-Emerging Technologies, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Martin Kuradusenge et al., "SMART-CYPS: An Intelligent Internet of Things and Machine Learning Powered Crop Yield Prediction System for Food Security," *Discover Internet of Things*, vol. 4, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Rashmi Sharma et al., *Predictive Algorithms for Smart Agriculture*, Data Analytics and Machine Learning: Navigating the Big Data Landscape, pp. 61-80, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[21] Nisit Pukrongta, Attaphongse Taparugssanagorn, and Kiattisak Sangpradit, "Enhancing Crop Yield Predictions with PEnsemble 4: IoT and ML-Driven for Precision Agriculture," *Applied Sciences*, vol. 14, no. 8, pp. 1-29, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Nafees Akhter Farooqui et al., *Precision Agriculture and Predictive Analytics: Enhancing Agricultural Efficiency and Yield*, Intelligent Techniques for Predictive Data Analytics, Wiley, pp. 171-188, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] J.A. Khan et al., "Recent Advancements and Challenges in Deep Learning and Internet of Things for Precision Agriculture," *Sumera and Saeed, Adnan, Recent Advancements and Challenges in Deep Learning and Internet of Things for Precision Agricultures*. [Google Scholar]

[24] Nitipriya Anand et al., "Smart Agriculture System Using Internet of Things (IoT) and Machine Learning," *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Pune, India, pp. 1-7, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Cangqing Wang, and Jiangchuan Gong, "Intelligent Agricultural Greenhouse Control System based on Internet of Things and Machine Learning," *arXiv Preprint*, pp. 1-10, 2024. [CrossRef] [Google Scholar] [Publisher Link]