AVE Trends in Intelligent Computing Systems

Optimised XG-Boost Model and RSA Application-Based Intelligent Scheme for Internet of Agriculture Things

S. Sujitha[1,*], B. Venkatesh[2], S. Gopikha[3], M. Arunadevi Thirumalraj[4], Jouma Ali Al-Mohamad[5]

[1]Department of Electrical and Electronics Engineering, New Horizon College of Engineering, Bengaluru, Karnataka, India.
[2]Department of Electrical and Electronics Engineering, B.M.S College of Engineering, Bengaluru, Karnataka, India.
[3]Department of Information Technology, St. Joseph's College of Engineering, Chennai, Tamil Nadu, India.
[4]Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamil Nadu, India.
[4]Department of Computer Science and Business Management, Saranathan College of Engineering, Tiruchirappalli, Tamil Nadu, India.
[5]Department of Computer and Mobile Communications Engineering, Faculty of Information Engineering, Al-Shahbaa Private University, Aleppo, Aleppo Governorate, Syria.
sujithas@newhorizonindia.edu[1], venkateshb.eee@bmsce.ac.in[2], gopikha.re@gmail.com[3], aruna.devi96@gmail.com[4], jalmohamad@su.edu.sy[5]

*Corresponding author

Abstract: Internet of Things (IoT) advancements have brought about noticeable lifestyle shifts. In many nations, agriculture plays a crucial role; this industry needs to evolve towards the "Smart" economy. The key conclusion is a lack of soil expertise. Soil comes in a wide variety, each with its own set of properties. An in-depth understanding of soil conditions provides a wealth of data to improve harvests. In agriculture, machine learning is a cutting-edge tool for improving precision and addressing issues in crop productivity. New possibilities for relaxing, measuring, and understanding data are paired with massive data growth and improved computing power. An improved version of the sparrow search algorithm (ISSA) is presented in this study for picking out useful soil data. Based on the sparrow search algorithm (SSA), the ISSA adds a nonlinear weighting element to enhance its global search capabilities. This study presented a new extreme gradient boosting (XGBoost) model optimised using the Reptile Search optimisation strategy. Specifically, the impact of dataset size on model accuracy was investigated. Good agreement with experimental results is demonstrated by the findings that the prediction accuracy of the projected model declines with reduced dataset size, yet total (IoT) advancements have brought about noticeable lifestyle shifts.

Keywords: Soil Expertise; Good Agreement; Lifestyle Shifts; Reptile Search Optimisation; Crop Yield Problem; Sparrow Search Algorithm; Gradient Boosting; Optimised XGBoost Model.

Cite as: S. Sujitha, B. Venkatesh, S. Gopikha, M. A. Thirumalraj, and J. A. Al-Mohamad, "Optimised XG-Boost Model and RSA Application-Based Intelligent Scheme for Internet of Agriculture Things," *AVE Trends in Intelligent Computing Systems*, vol. 2, no. 3, pp. 155–173, 2025.

Journal Homepage: https://www.avepubs.com/user/journals/details/ATICS

Received on: 17/11/2024, **Revised on:** 01/03/2025, **Accepted on:** 25/04/2025, **Published on:** 09/09/2025

DOI: https://doi.org/10.64091/ATICS.2025.000199

1. Introduction

The phrase "smart farming" describes an increasingly popular approach to agricultural management. Monitoring crop growth and viability is now possible thanks to advances in agricultural and information technology. Field crop circumstances and other indicators are being monitored [1]. To sum up, the goal of farming is to reduce the expenditure on farm inputs without compromising on yield quality. Applying a large amount of pesticide or fertiliser to an entire field at once has the same effect on the entire field. Sand, salt, and clay are all present in the soil, among other sorts [2]. There are benefits and drawbacks to every soil type. Sand, which drains quickly, is an excellent example of this type of soil. Conversely, soil nutrients are quickly lost through drainage. Soil characteristics are critical for estimating a plant's water requirements [3]. The data mining techniques for the agricultural sector are particularly promising. Managing water use in agricultural regions is one such activity. In addition, IoT data collection and storage methods have enabled smart farming [4]. In contemporary irrigation systems, sophisticated sensor networks and smart production are just a few of the many real-world uses for machine learning. Soil and data, which are subsequently uploaded to a cloud server as part of the framework [5]; [6]. Different analytics are performed using machine learning algorithms on a cloud server. The precise amount of water needed by a crop may be calculated using this approach. Recently, deep learning has seen great success [7]. These applications are used to manage various agrarian operations by leveraging data from diverse sources.

There are many AI-based intelligent systems on the market, each with its own unique capacity to capture and understand data and provide timely guidance to farmers. IoT nodes (sensors) may be deployed to collect data, and learning approaches and decisions may be applied to operational domains via actuators [8]. The AI system is supplemented by other cutting-edge technologies for real-time nursing and agricultural organisation, such as remote sensing and global control [9]. In addition, AI-based smart agriculture can plan optimal use of resources such as fertiliser, insecticides, and water, reducing pollution and operational/production costs while increasing cumulative output [10]. Because AI aids in the early identification and control of plant diseases, fewer drugs are needed to stop their spread, which greatly lowers environmental pollution [11]. Water must be supplied regularly to plants to ensure their growth, development, and productivity [12]. Abiotic and biotic stress may result from a lack of any of these factors. Only AI can account for both current and anticipated needs when allocating resources. This study investigated the potential of deep learning in agriculture. Researchers also examined IoT-monitored agricultural factors and fed them into deep learning algorithms for analysis [13]. DL methods also provide information on agricultural yields under varying climate conditions, which is a major plus [14]. The learning process is sped up, and the results are enhanced. However, these methods have several drawbacks, including low sustainability, high computational cost, excessive complexity, and inaccurate predictions [20]. This paper proposes a more effective ML-based model for agricultural production prediction to address these issues [15].

1.1. Motivation

India relies heavily on its agricultural sector as a source of revenue. Conventional agricultural planting faces high costs and complex administration. Still, IoT is being used to enable real-time detection, intelligent crop growth management, and a shift away from this method. For several crops, the efficacy of both mathematical and empirical yield estimation methods has been assessed. These models require extensive information on soil and crops, which hinders their adaptability to different regions. The field of yield modelling has also seen the development of several satellite-based remote sensing systems. However, these methods do not yield sufficient geospatial data on small farms for crop optimisation. Researchers can now solve and interpret difficult predictions thanks to recent breakthroughs in machine learning algorithms. XGBoost is utilised for agricultural production prediction, resulting in improved accuracy compared to previous ML methods such as decision trees, ANNs, and SVMs.

1.2. Contributions

The most significant new elements in this planned work are: To reliably estimate agricultural yields, the authors propose an IoT-based farming system. Preprocessing, fuzzy set theory, and categorisation are all part of this task. FS algorithms perform preliminary data analysis and select the features to use:

- To improve agricultural production prediction, an IoT-based agriculture system that employs an ML model is presented.
- Given the characteristics of the soil dataset, ML-based classification is used to categorise the soil trials into several groups.
- Researchers propose the XGBoost model for crop yield prediction and use the RSA method to update the model's weights to improve the accuracy of the prediction system.
- The suggested crop production prediction system is systematically evaluated using experimental results.

The first section of the paper is the introduction, which provides a summary of smart agriculture. In Part 2, researchers talk about some of the most recent developments in this area from 2022 to 2023. The fundamentals of machine learning for predicting crop yields via IoT are discussed in Section 3. The suggested structure is detailed in Section 4, along with key metrics for its success. In Section 5, the researchers will discuss the analysis and findings from the experiments. The provided study is summarised in Section 6.

2. Related Works

A sensor-based intelligent control system for smart agriculture with autonomous irrigation, based on the Internet of Things, was proposed by Manikandan et al. [16]. The suggested solution helps farmers increase cultivation by leveraging device data and an autonomous watering system. During the growth season, IoT devices collect crucial field data, including UV radiation, humidity, temperature, light intensity, and soil moisture. Using a user-specified location, users can continuously monitor the field using the composed data. The data is sent for analysis, and then a smart watering system is built around the findings using a controller. To demonstrate the efficacy of the suggested method, Researchers test and verify the smart agricultural module's performance across a range of settings. Mathi et al. [17] suggest an Internet of Things-based strategy for intelligent farming. The suggested system's primary functions are automated plant watering and disease detection. Algorithms enable predicting how much water crops will need and automatically identifying pests based on their needs. In the pest-detection module, Researchers employ machine learning techniques to accurately predict plant diseases. The useful properties of plant leaves were extracted. After collecting features, they are used in a categorisation process. Whether a pest has infected a plant can be determined by extracting and classifying relevant traits. To automatically irrigate and identify plant diseases, the suggested system monitors, analyzes, evaluates, and regulates agricultural areas. The importance of accurate categorisation is explored, as is the numerical analysis of machine learning methods. The numerical findings show that a precision of 84% is attained. Abdurahman et al. [18] investigated the need for more advanced agriculture technologies to solve this problem.

Greenhouses often use the DHT22 sensor to obtain humidity readings from inside the greenhouse—the BH1750 sensor measures ambient temperature and light intensity, while the YL-69 measures soil moisture. Classification results generated by this method will include both optimum and suboptimal values for each criterion. The goal of this capstone project is to develop a categorisation scheme for the optimal growth of red spinach plants, with a focus on germination and early development. The test findings confirm the system's efficacy. A regular delay of 1.880 seconds was observed during QoS testing. The regular data-reading speed during QoS testing was 4,464 bps across IoT, sensor, MySQL, dataset, Raspberry Pi 3B+, and 3B+. When Zeng et al. [19] wanted to monitor farms more effectively, they used IoT and Blockchain technologies. To better manage water resources across communities, a smart water management system and an Internet-based seed quality monitoring system have been developed. Information security and community trust are two of the primary reasons for implementing a Blockchain network. Trust among commercially viable yet resource-limited systems that interact with the Blockchain network, which itself is a hardware platform, is another use case for the technology. The prototype's layout and its performance during implementation are also shown. An administration of fish farming is the subject of Al-Mutairi et al. Measuring several types of factors and using that data to regulate fish growth and boost output is crucial to the design of such a system—a wireless sensor network in which each fish pond functions as a node. The node has a wireless connection module, a collection of sensors, and actuators, all linked to an embedded microprocessor. The pond water quality and the surrounding environment are monitored and controlled by two fuzzy controllers, each with five sensors (5 in each pond and 3 in the surrounding area). Compared with the output of commercial instruments commonly used in agricultural settings, the practical findings indicate that the measuring method is accurate. These findings also indicated that the suggested method yields the highest efficiency in a real-world fish-pond system. Technological developments led Mamatha and Kavitha [21] to use the hydroponics smart greenhouse approach. Hydroponics is a method of plant cultivation that eliminates the need for soil, enabling faster growth and higher yields.

Normal hydroponic systems use only a small fraction of the greenhouse for harvesting, subjecting the plants to environmental factors such as humidity that can stunt their growth and deprive them of the nutrients they need to thrive. Since rock wool is not biodegradable and is made of volcanic elements, this technique proposes using organic coconut coir media for germination instead. In the proposed hydroponics study, the entire greenhouse is outfitted with automated systems for growing a variety of crops in controlled environments. The implementation takes place in a continuous greenhouse environment, where plants fill the entire available cultivation space, providing ideal conditions for achieving maximum production. The KNN algorithm is used to predict the absolute growth rate of leafy vegetable crops under different environmental conditions. It has been determined that NFT, in conjunction with a coconut coir medium, is the most effective hydroponic system for commercial crop development, with an accuracy of 93%. The accuracy of an average, which sums the instances across various class labels, is 93%, whereas the accuracy of a macro average, a presentation parameter for the F1 score, is 33%. The effective convolutional neural network (CNN) approach employed by Gupta and Shah [22] may identify the cause of leaf illnesses. Implementation procedures, such as data collection, training, testing, and classification, are outlined in our proposed publication, as is the use of a convolutional neural network (CNN) to determine whether leaves are ill. In this study, IoT and CNN alerts are used to diagnose potato leaf diseases and devise treatments. In this study, Researchers compare the performance of KNN and CNN methods for diagnosing potato leaf illness. The established method also helps mitigate sickness via email and IoT. CNN-based categorisation has achieved an accuracy of >90%.

3. Background Knowledge About Machine Learning Concepts for Agriculture

Smart agriculture is an emerging concept that improves agricultural efficiency through more precise algorithms. IoT-based smart agricultural systems may continually learn about soil properties using sensor data collected at discrete time intervals. Cellular M2M is a viable technology in agriculture, as wired networks offer certain advantages but also pose disadvantages in terms of cost, performance, usability, and other areas. However, the term "machine learning" was first used in the context of computing, and several vector quantisation techniques have been developed for use in signal processing and compression. Commonly used in agricultural IoT applications are wireless technologies such as Wi-Fi, Bluetooth, (WAN), and Zigbee, as well as mobile and cellular variants. An improved Internet of Things (IoT) will raise public awareness and reduce food waste through machine-to-machine (M2M) communication enabled by wireless technology. The tasks may be broken down into three groups: (a) animal welfare and production; (b) plant control and testing; and (c) disease detection and forecasting. Management of water in soils (c). Agronomic management systems provide the finest recommendations and guidance for future decisions and actions by using machine learning to analyse sensor data and build artificial intelligence systems that boost performance.

3.1. IoT Cloud Platforms for Machine Learning

Cloud computing is the use of remote servers and networks to store and process data. The term "cloud" has come to refer to both the hardware and software data centres. The cloud platform can be integrated with IoT systems using popular microcontroller boards such as Arduino and Raspberry Pi. Advantages include hardware flexibility, reduced risk to cloud stability, reduced downtime, support availability, internet access, and the ability to instantly install cloud-based devices without incurring any costs. The following are some ways in which IoT Cloud may be put to good use in IoT initiatives:

- **ThingSpeak:** This environment supports free, open-source IoT software. The Internet protocols HTTP and MQTT are used to both store and retrieve data. ThingSpeak enables the creation of sensor recording software, new areas of application, and the collection of informally gathered data.
- **Kaa:** A flexible, open-source middleware platform for creating smart devices and the Internet of Things from start to finish. This shortens time-to-market while reducing costs and the risk of failure. Kaa also provides a variety of IoT tools that are well-integrated with millions of devices and can be quickly linked to IoT applications.
- **Microsoft Azure:** A cloud infrastructure built for trying out and managing programs and services in a regulated environment provided by Microsoft. It supports several programming languages, methodologies, and frameworks, as well as third-party applications and cloud computing services, including PaaS, SaaS, and IaaS [23].
- **Firebase:** The Firebase platform is provided by Google (the company). Firebase leverages Google Cloud Messaging (GCM) to receive free alerts and messages sent via Android, iOS, and the web. As detailed in Stevenson [24].
- **AWS:** Amazon Web Services, which provides distributed computing services at the request of users, organisations, and governments [25]. Parameters for cloud integration with Internet of Things devices are listed in Table 1.

Table 1: Cloud stages for IoT

Stages	Data	Services Integration	Data Visualization	SDK API	Free account
ThingSpeak	store	Yes (Email, SMS)	Yes	Yes	Yes
Firebase	Yes	Yes (Email, SMS)	Yes	Yes	Yes
AWS	Yes	No	No	Yes	No
Kaa	Yes	Yes	No	Yes	Yes
Microsoft Azure	Yes	Yes	Yes	Yes	No

- **Service integration:** Connecting disparate software and hardware is what it's all about. Notifications can be sent via service integration.
- **Data Store:** The availability of cloud storage is detailed.
- **Data visualisation:** Graphs, charts, and graphics are used to depict data.
- **SDK and API:** A software toolkit for creating new programs is called a software development kit (SDK). API stands for "application programming interface," and it enables the exchange of data between applications.
- **Free Account:** This defines the cost of using a cloud service.

4. Proposed Methodology

4.1. System Model

The goal of the (IoT) smart agriculture project is to build a comprehensive crop forecasting system.

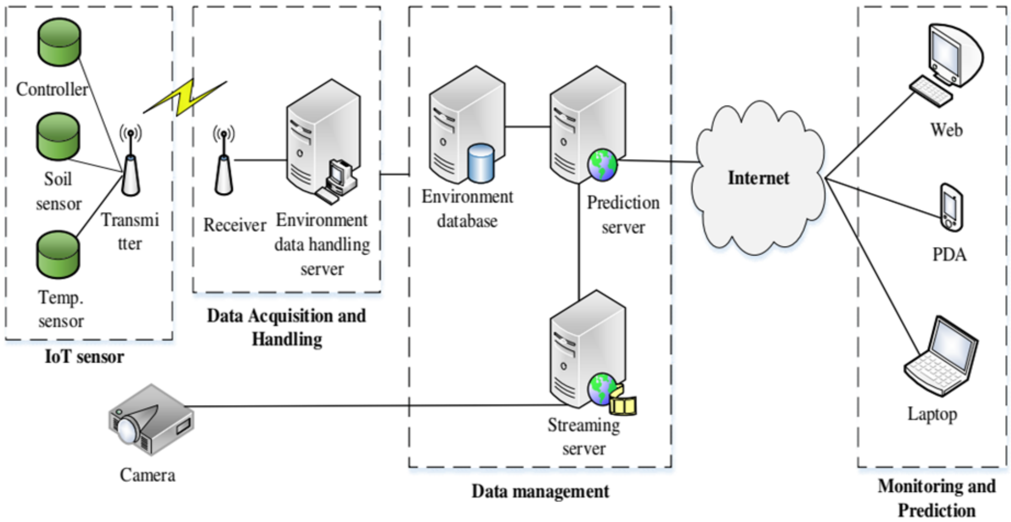

Figure 1: Service of IoT for agriculture

The primary intent of this model is to utilise data collected from (IoT) devices to make predictions about agricultural productivity. It helps the agricultural yield prediction system make more accurate crop yield predictions by analysing data from monitoring cameras. Camera data is combined with environmental data gathered by Internet of Things gadgets. Early access to environmental data is crucial for protecting crops and property during emergencies. Statistics from farms and modern information and communication technology (IoT) sensor data are used to create a reliable prediction model. Precipitation, temperature, pH, soil nutrients, and fertilisers are all measured and analysed using modern IoT sensor data and historical statistics. The architecture of IoT services for farming is exposed in Figure 1. The sensors, data-gathering devices, data-management devices, and prediction devices make up the bulk of an IoT-based agricultural model. The accuracy of projected agricultural yields was improved with the use of these tools. Predicting soil type, weather fluctuations, soil quality, and other factors can help an integrated IoT system for a smart farm increase crop yields.

4.1.1. Sensors

Using IoT sensors, farmers may monitor environmental factors such as soil and water pH, temperature, humidity, and fertiliser application [26]. In the sections that follow, Researchers discuss moisture, temperature, humidity, and pH sensors:

- **Soil Moisture Sensor:** It's a sensor that measures soil moisture levels. A probe has multiple sensors that work together to reduce water use and optimise irrigation system efficiency. The use of these sensors also improves both crop quality and production. It's fast, cheap, and accurate. What more could you want? This sensor measures soil water content and provides an average value across its entire range. This sensor was designed to track soil moisture levels for irrigation management, analyse water loss over time through transpiration and root absorption, and determine the ideal soil moisture level for plant growth. Once the threshold value is established in the software, the sensor values are sent to the transmitter. The threshold value controls the water motor's operation in the field; if the threshold value is higher than the water level, the motor will turn on to provide water. Once the sensor indicates that the desired water level has been reached, the irrigation process is terminated.
- **Humidity Temperature Sensor:** Predicting precipitation, tracking temperature swings, logging how much rain fell in a given time frame, and tracking how much water was utilised all need this. It consists of a resistive element and a moisture-sensitive NTC temperature sensor. This sensor has excellent reaction time, low power consumption, good quality, and interference resistance.
- **pH Sensor:** The pH scale measures the acidity and alkalinity of a given solution. Typically, a pH scale spanning from 0 to 14 is used. It is also a measure of the concentration of some solutions containing hydrogen ions. This sensor measures the voltage between electrodes, such as a hydrogen-sensitive glass electrode and a ground electrode. It is compatible with microcontrollers such as Arduino.

4.1.2. Controller

The farm's climate is monitored and managed using an Arduino controller. The controller's inputs are the distance away from the farm and the pH level. Also included is the nutritional breakdown (as percentages): calcium, phosphorus, magnesium, iron, sulphur, nitrogen, potassium, and boron. The controller then makes the necessary forecast decisions based on either the pretrained network or the current data.

4.1.3. Data acquisition

The first step in collecting data is adding each gadget in the farm to Mobius. The &cube device is used for registration and also serves as a bridge between the Mobius and the fielded devices. Then, the virtual representation of each gadget is generated based on its resource category. Sensors report their findings to &cube, which then relays the information to Mobius [27]. Ultimately, farmers or end users use the accessible virtual representations of the equipment to monitor and control their linked farms using IoT apps on devices such as tablets, laptops, and smartphones:

- **Mobius:** It's a free and public Internet of Things service platform that runs on a single, unified M2M protocol. This Mobius facilitates the creation of virtual replicas of actual IoT devices. It's built to meet the standards of a single M2M specification and aids the operation of common M2M/IoT services, such as authentication, device registration, subscription, notification, and data storage and administration. In addition, REST APIs are used to access Mobius-stored data resources and manage Internet of Things devices.
- **Cube:** A platform for device software that can be used in IoT gateways. The information from the gadgets is sent to Mobius via the common REST APIs. &cube supports several protocols, including HTTP, MQTT, and CoAP. In this paper, a Raspberry Pi is used to construct a networked farm. It's a Linux-powered single-board computer.

4.1.4. Data Management

The sensors capture relevant exterior data and send it to servers, where it is stored until needed. The camera's data is also saved so that the harvest may be forecasted. This data management system stores, updates, and manages information to keep related services running smoothly.

4.2. Proposed Framework

Soil nutrients like phosphorus, potassium, and nitrogen, as well as factors such as crop rotation and weather, are crucial to successful farming. When it comes to making educated guesses about what crops will provide the most profit, ML techniques are a crucial decision-making tool. Estimating harvest yield extensively uses ML algorithms. For crop production prediction, the suggested smart agricultural outline uses pre-processing, Feature Selection (FS), and a modified machine learning model. Compared to other algorithms, XGBoost generally offers faster learning, higher performance, and lower training error, making it a highly influential model. This study proposes a classification method to enhance the system's precision and produce superior results compared to existing models. Data noise is first removed during the preprocessing phase, and then features are selected using the improved sparrow search algorithm (ISSA). Finally, a reptile search strategy is used in conjunction with an XGBoost model to classify expected crop yields. This model enhances the system's precision, resulting in lower error rates.

4.2.1. Pre-processing

Data is composed and preprocessed from multiple sources. Because it cannot interpret noisy input, ML requires a preprocessing phase. When data is "noisy," it contains mistakes and anomalies. Classification requires preprocessing the data to fill missing values, remove redundant information, extract useful features, and keep the data within a suitable range. This work uses the isnull() method to detect missing values, then uses the label encoder() to convert categorical data from text to numerical format. Categorical data must be translated to a numeric representation before being used with Python. After the information is reduced to numbers, it is subjected to a feature selection process.

4.2.2. Feature Selection using Improved Sparrow Search Procedure

The (SSA) mimics the foraging behaviour of sparrows, which is commonly accomplished by splitting the flock into producers and followers [28]. In SSA, the most disadvantaged member of the population is given priority in food searches. Producers can explore a wider area in search of food than consumers can. Equation (1) displays the approach used to update the explorer's location in each iteration:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t}.\exp\left(\frac{-i}{a.iter_{max}}\right), \text{if } R_2 < ST \\ X_{i,j}^{t} + Q.L, \quad \text{if } R_2 \geq ST \end{cases} \tag{1}$$

when R_2< ST, it means they have seen a potential threat, and the rest of the flock must take precautions. Some of the flock keep an eye on the producers at all times, even while they're out foraging. The farmers will abandon their existing settlement at the first sign of better food elsewhere. The food is available immediately if they win the tournament; otherwise, they must continue to follow Equation (2):

$$X_{i,j}^{t+1} = \begin{cases} Q.\exp\left(\frac{X_{worst}^{t} - X_{i,j}^{t}}{i^2}\right), \quad \text{if } i > \frac{n}{2} \\ X_p^{t+i} + \left|X_i, j^t - X_p^{t+i}\right|.\ A^+.\ L, \ \text{otherwise} \end{cases} \tag{2}$$

where X_p^{t+i} is the optimal creator position, X_{worst}^{t} Is the current global worst site, and n is the population size. A is a 1 × d matrix where each element's random amplitude is either 1 or -1. In Equation (3), Researchers see the formula for calculating A+:

$$A^+ = A^T(AA^T)^{-1} \tag{3}$$

when i > $\frac{n}{2}$It means the i-th follower with the lowest fitness level is in bad shape and has to go elsewhere for sustenance. The initial placements of those people in the population who are aware of the threat are produced at random, as illustrated in Equation (4):

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^{t} + \beta.\ \left|X_{i,j}^{t} - X_{best}^{t}\right|, \quad \text{if } f_i > f_g \\ X_{i,j}^{t} + K.\left(\frac{X_{i,j}^{t} - X_{worst}^{t}}{f_i - f_W + \omega}\right) \quad \text{if } \ f_i = f_g \end{cases} \tag{4}$$

where X_{best}^{t} K is a random number in (1,1), f is the value, f_i is the present optimal fitness value, and f_g is the value. Is the position, b is the step-size control parameter (its value follows a normal distribution), and f is the fitness value. To prevent a null denominator, w is a constant. Therefore, $f_i > f_g$ Population. When $f_i = f_g$This behaviour suggests that the sparrow in the middle of the population is aware of the risk of being alone and, as a result, must remain in proximity to other sparrows. Here, K not only indicates the step control parameter but also the direction in which the sparrow will fly. In SSA, sparrows converge on the optimal solution by leaping to the region around it rather than gradually approaching it. This causes SSA to struggle during a global search and to become mired in local optima. Equation (1) demonstrates the limitations of SSA's capacity to perform a global search. As a first step toward improved performance, the SSA's Equation (1) was modified to eliminate convergence to the origin and to redefine the ideal position as a region near the origin. As seen in Equation (5), this study also introduces a nonlinear weight component:

$$\Delta = 2 - \left(\frac{t_{iter}}{T_{max}}\right)^2 \tag{5}$$

The fully improved equation is obtainable as Equation (6):

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t}.\ (1 + \Delta.Q), \quad \text{if } R_2 < ST \\ X_{i,j}^{t} + Q, \quad \text{if } \ R_2 \geq ST \end{cases} \tag{6}$$

SSA's optimisation performance can be greatly improved by adding a nonlinear weighting element. The convergence speed of the method and the optimisation accuracy can be improved by using a factor that increases with iteration duration, allowing dynamic adjustments to the global search aptitude of SSA. Next, Researchers keep Equation (2) and enhance Equation (4) by having each sparrow randomly approach the discoverer in all three spatial dimensions. Equation (7) displays the exact formula:

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} + \beta.\left(X_{i,j}^{t} - X_{best}^{t}\right), \ \text{if } \ f_i > f_g \\ X_{best}^{t} + \beta.\left(X_{worst}^{t} - X_{best}^{t}\right), \ \text{if } \ f_i = f_g \end{cases} \tag{7}$$

The improvement eliminates the unwieldy parts of the unique procedure. When $f_i > f_g$The sparrow will flee to an accidental site. When $f_i = f_g$The sparrow, on the best possible spot, will leave and land anywhere between the best and worst spots. These enhancements fix the original algorithm's biggest flaw—its slow update rate—and significantly enhance its ability to conduct a global search. As a result, the entire algorithm has been strengthened, improving not only its accuracy and convergence speed but also its global search capacity. Each location is updated within the allotted number of iterations using the aforementioned formula to discover the ideal solution that is both practical and comprehensive.

4.2.3. Classification using XGBoost Model

When solving regression and classification issues, the XGBoost method employs a forest of CARTs. This research presents a logistic regression challenge for estimating future harvests. Numerous CART regression trees use XGBoost as a powerful regressor fusion. XGBoost's structure consists of multiple root nodes and branches, as shown in Figure 2. To generate the initial choices, this structure takes as input the i-th limit, xi, and sends it to the root nodes of all CARTs. A single CART's predictions are shown as leaf nodes, whereas internal nodes determine future judgments, and branch points indicate the decision to be made directly. The XGBoost model's predictions are obtained by summing the ratings of all nodes that terminate in leaves [29].

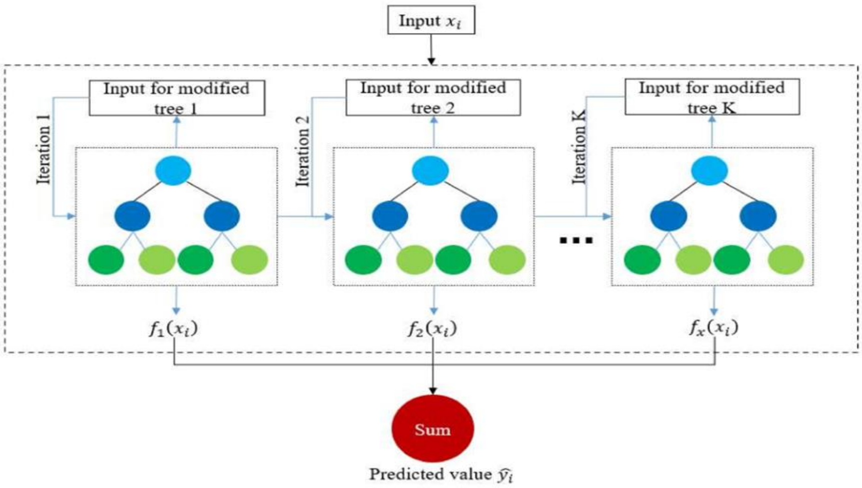

Figure 2: Graphic diagram of the XG boost tree perfect

As an instance, in the i-th set (x_i, y_i) (x_i Is the input data with manifold features? y_i The XGBoost regression tree is expressed exactly as Chen and Guestrin [30]:

$$\hat{y}_i = \alpha \sum_{k=1}^{K} f_k(x_i) \tag{8}$$

Where $\hat{y}_i$ Is f_i is the output of the ith CART given the input x_i, is the learning rate of that tree, and K is the total number of CARTs in the ith tree. The expected grade is as shown by Equation (8). $\hat{y}_i$ is the sum of all f_k Values. Once the prediction result was produced, the quality of the result was assessed using the objective function L, which is represented by:

$$L = \sum_{i}^{n} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_x) \tag{9}$$

There are two components to the objective function: (1) the loss function l, which quantifies the loss of the regulation item, and (2) the degree of complexity of the regression tree. In the context of a CART, it was written as:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{10}$$

In CARTs, the sum of leaf nodes is denoted by T, the forecasted value of the jth leaf node is denoted by w_j, and the control variables are used to prevent overfitting. The XGBoost classifier was trained, and the procedure was carried out step by step to optimise the objective function and prediction outcomes. By iteratively creating new CARTs from preexisting CARTs, the goal function was further minimised at each stage. A second-order Taylor expansion was applied to the equation after the present CARTs were substituted by the constant c. The t-th step's objective function, L((t-1)), was derived from the preceding step's L((t-1)), as:

$$L^{(t)} = \sum_{i}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)\right)\right] + \Omega(f_t) + c \tag{11}$$

Where:

$$g_i = \frac{\partial l\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial \hat{y}_i^{(t-1)}} \tag{12}$$

$$h_i = \frac{\partial^2 l\left(y_i, \hat{y}_i^{(t-1)}\right)}{\partial (\hat{y}_i^{(t-1)})^2} \tag{13}$$

In this study, the loss purpose selects the (RSE). Each input mutable x_i was charted to a leaf node of a CART, so $f_k(x_i)$ was articulated as:

$$f_k(x_i) = \omega_q(x_i), \omega \in R^T, q(x_i): R^d \rightarrow \{1,2, \dots, T\} \tag{14}$$

Where q(x_i) is the index of a particular leaf node, d is the input x_i, RT is a T-dimensional vector, and Rd is a vector. Using Equations (10) and (12)-(14), researchers obtained the first derivative using Equation (11), which researchers then solved for:

Letting $G_j = \sum_{i \in I_j} g_i$ and $H_j = \sum_{i \in I_j} h_i$ when $\omega_j = -\frac{G_j}{H_j + \lambda}$, L_{min} was written as,

$$L_{min} = \frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T + c \tag{15}$$

Therefore, the foretold value shown on the leaf nodes was the ideal value of the impartial function L, and a greedy method was used to improve the regression tree construction for each CART to discover this optimal structure [31].

4.2.3.1. Hyperparameter Optimisation with RSA Method

Phase 1: Initialisation of RSA's Limits

Before running the RSA, it is important to initialise the control settings and the algorithmic parameters. The supreme sum of iterations is set by the control parameter (T), while the number of crocodiles is represented by the parameter (N). Moreover, RSA uses two algorithmic parameters, such as and. To find the sweet spot in the search process, the algorithm uses these two parameters to balance exploitation and exploration [32].

Phase 2: RSA Population Initialisation

At this point, researchers arbitrarily produce a set of early keys using the subsequent equation:

$$X_{i,j} = X_j^{min} + rn\ d \times \left(X_j^{max} - X_j^{min}\right), \forall i = 1,2, \dots, N$$

$$\forall j = 1,2, \dots, d, \tag{16}$$

where $X_{i,j}$ The choice represented at the j th position is X_j^{max} and X_j^{min}. Rand is a randomly generated value between 0 and 1, while d denotes the total quantity of the decision variable at each solution. The keys, as many as N, are produced and stored in X as shadows:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,d-1} & X_{1,d} \\ X_{2,1} & X_{2,2} & \ddots & X_{2,d-1} & X_{2,d} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{N,1} & X_{N,2} & \cdots & X_{N,d-1} & X_{N,d} \end{bmatrix} \tag{17}$$

where each row $X_i = (X_{i,1}, X_{i,2}, \dots, X_{i,d-1}, X_{i,d})$ indicates the key of i th site.

Phase 3: Fitness Evaluation

The fitness value (i.e., excellence) of each key in the populace should be intended as $f(X_i)\ \forall i = 1, 2, \dots, N$.

Phase 4: Encircling Phase

This is typical crocodile exploring behaviour in the RSA. During this stage, researchers present two strategies—high seek, which explores different parts of the problem's search space to find a better solution. The tactic of walking tall is managed by by t≤T/4, while the strategy is skillful by T/4<t≤2T/4:

$$X_{i,j}(t+1) = \begin{cases} X_j^{Best}(t) - \eta_{i,j}(t) \times \beta - R_{i,j}(t) \times rn\,d, & t \leq \frac{T}{4}, \\ X_j^{Best}(t) \times X_{r1,j}(t) \times ES(t) \times rn\,d, & \frac{T}{4} < t \leq \frac{2T}{4}, \end{cases} \tag{18}$$

Where $X_{i,j}$ Signifies the choice of the i th key at the j th position. $X_j^{Best}(t)$ Is the jth position in the best solution obtained at t repetitions? t + 1 is the novel repetition, and while the previous repetition is t. The hunting operator of the j th is signified as $\eta_{i,j}(t)$, which is calculated using (19). The limit β controls the walking policy during examinations. The value of β is set to 0.1. rn d is a randomly generated value ranging from zero to one. $X_{r1,j}(t)$ Is the decision mutable at the j th solution, where r1 ∈ [1, N]? $\eta_{i,j}$ (t), $P_{i,j}$ And Avg(Xi) are intended, correspondingly, as shadows:

$$\eta_{i,j} = X_j^{Best} \times P_{i,j} \tag{19}$$

$$P_{i,j} = a + \frac{X_{i,j} - Avg(X_i)}{X_j^{Best} \times \left(X_j^{max} - X_j^{min}\right) + \varepsilon} \tag{20}$$

$$Avg(X_i) = \frac{1}{d} \sum_{j=1}^{d} X_{i,j} \tag{21}$$

Where P_(i,j) is the percentage solution, X_i is set to 0.1, which is also used to regulate the RSA's exploratory prowess during the hunting cooperation. Is a number between 0 and 2 chosen at random. For this approach, the average of all choice variables, X_i, is denoted by Avg(X_i). Using the following formulae, researchers can get X_i and the search area reduction factor R_(i,j) (t) for the jth location in the ith solution, and the sense likelihood ES(t), which assigns a randomly decreasing value between 2 and -2:

$$R_{i,j} = \frac{X_j^{Best} - X_{r2,j}}{X_j^{Best} + \epsilon} \tag{22}$$

$$ES(t) = 2 \times r3 \times \left(1 - \frac{1}{T}\right) \tag{23}$$

Where r2 is an arbitrary sum between one and N, denoting the index of a randomly selected solution from the population. r3 is an arbitrary positive or negative integer between 0 and -1.

Phase 5: Hunting Phase

This is the crocodiles' exploitative style in the RSA. As illustrated in (24), this stage of the RSA uses the present study regions to identify optimal solutions using either the hunting coordination or hunting cooperation techniques. The hunting cooperative policy is governed by t > 3T/4, whereas the hunting policy is governed by t 3T/4:

$$X_{i,j}(t+1) = \begin{cases} X_j^{Best}(t) \times P_{i,j}(t) \times rn\,d, & \frac{2T}{4} < t \leq \frac{3T}{4}, \\ X_j^{Best}(t) - \eta_{i,j}(t) \times \epsilon - R_{i,j}(t) \times rn\,d, & \frac{3T}{4} < t \leq T \end{cases} \tag{24}$$

Phase 6: Stop Criterion

Iterate through Steps 3–5 until the maximum number of iterations, T, is reached. Hyperparameters such as the maximum tree depth dmax (e.g., dmax = 3 in Figure 1), an in Equation (8), K in Equation (8), g in Equation (10), and l in Equation (10), are crucial to the XGBoost concept and hence to the procedure's strengths and weaknesses. The random starting values for these hyperparameters in this study are as follows: dmax = 3, a = 0.3, K = 300, and g = l = 0. XGBoost produces a training dataset-concordant model once hyperparameter values are specified. In this study, researchers used a ten-fold cross-validation strategy, which randomly splits the training dataset into 10 subsets to enhance training accuracy. Figure 3 depicts the results of training and evaluating the established XGBoost model 10 times, with each iteration using a different set of nine subsets for training and one for assessment. Researchers ended up with a total of 10 evaluation ratings (E) and the average of those values.

Figure 3: Diagram of ten-fold cross-validation

Table 2 displays the hyperparameters and typical differences used to fine-tune the model.

Table 2: Hyperparameter varieties for model fine-tuning

Element	Tolerance	Variety of Values
Dmax	1	1–10
K	50	100–600
γ	0.01	0–0.05
λ	0.1	0–1
A	0.02	0.01–0.3

5. Experimentation, Results, and Discussion

In this part, researchers explain the scheme they created and analyse its performance (Figure 4).

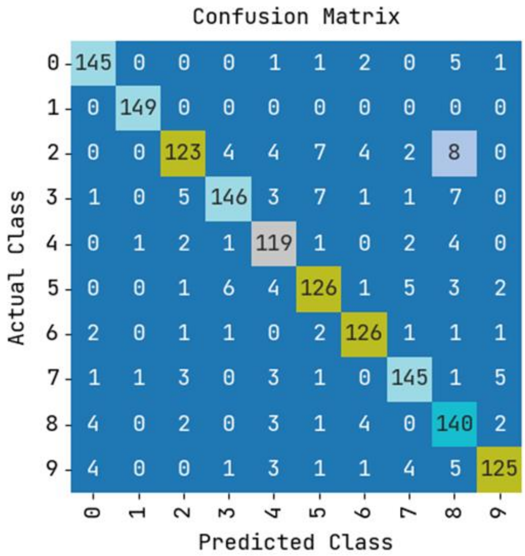

Figure 4: Testing confusion matrix

A computer with 8 GB RAM and an Intel Core i5 processor has handled the full implementation. Python 3.8 is used to implement the plan. In this study, the researchers use a soil dataset for their experiments. Figures 4 and 5 display the proposed model's training and testing confusion matrices, respectively (Figure 5).

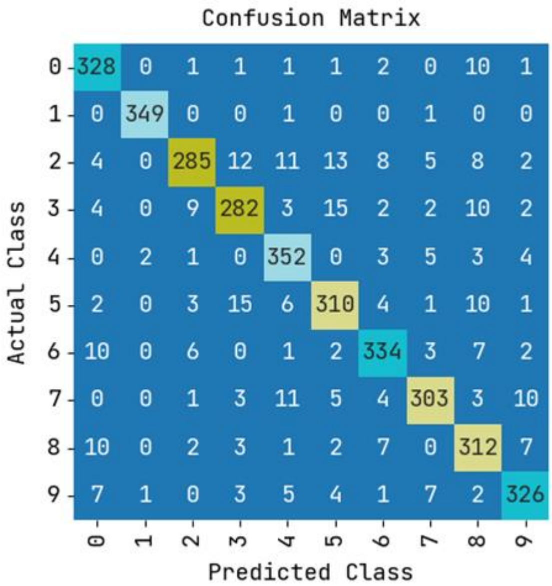

Figure 5: Training confusion matrix

The graphical representations of the precision-recall curves for the training and test data are shown in Figures 6 and 7, respectively.

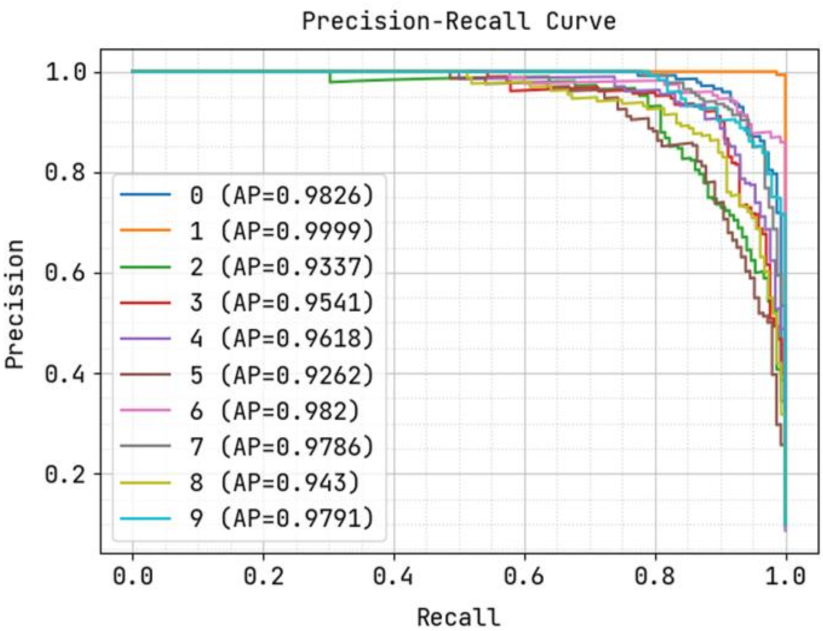

Figure 6: Testing the precision-recall curve

Figure 7 shows the training precision-recall curves for all classes. It shows that high precision is kept throughout a wide range of recall values. This shows that the model can properly identify positive examples while minimising false positives. The consistently high average accuracy (AP) across classes indicates that the learnt representations are strong and reliable. The curves are also smooth and stable, which implies that the training process is working well and that the model performs well across diverse class distributions. This means the model performs well across different decision thresholds and delivers consistent classification performance, even when recall is increased to obtain more relevant data.

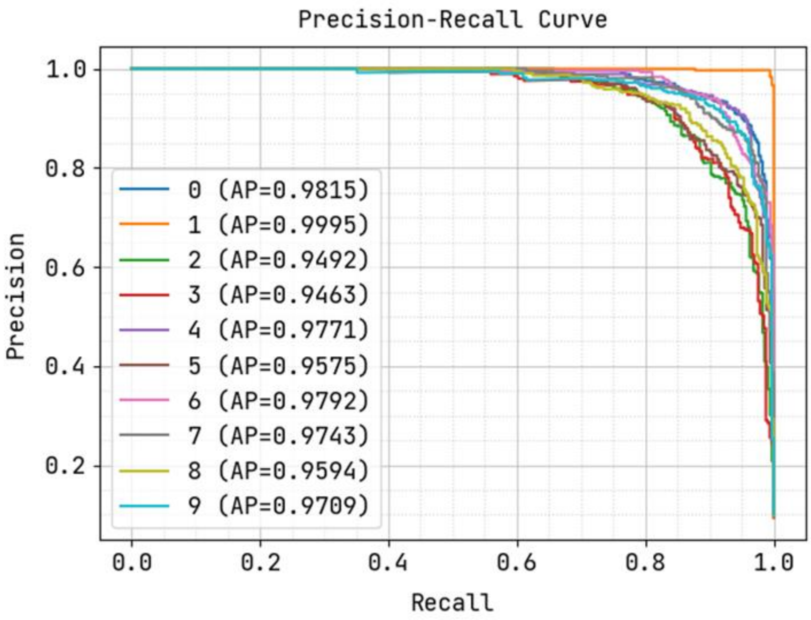

Figure 7: Training precision-recall curve

ROC curves for equal training and testing for the proposed model are provided in Figures 8 and 9.

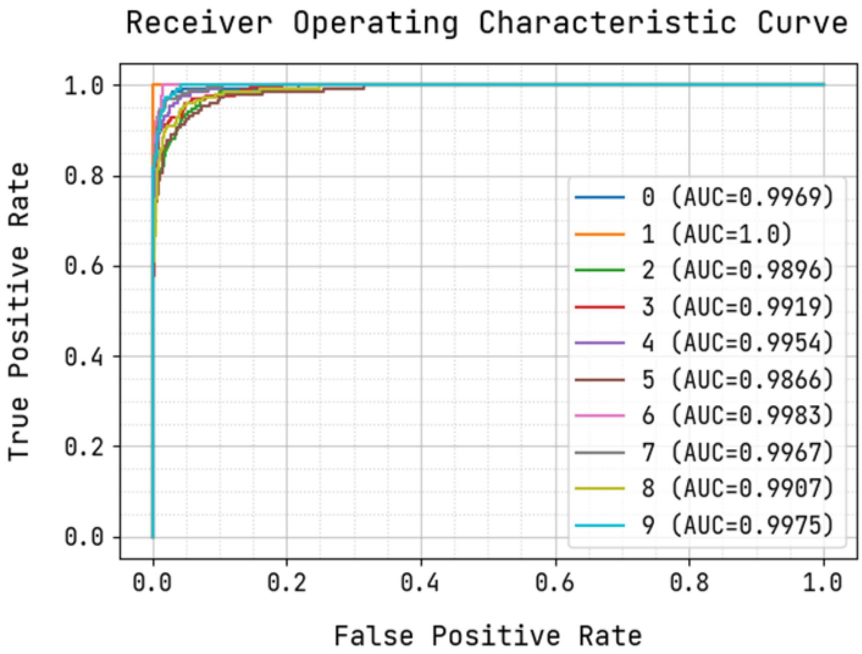

Figure 8: Testing the receiver operating characteristic curve

Figure 9 shows the training ROC curves for all classes. It illustrates the true positive rate and the false positive rate at different decision thresholds. The curves are near the top-left corner and have very high area under the curve (AUC) values. This means the model is very good at distinguishing between positive and negative samples. This performance indicates that the model is quite sensitive while keeping the false-positive rate low, suggesting it separated classes well during training. The fact that the ROC curves are the same across all classes also shows that learning is balanced and that there is less bias toward any one class. The fact that the curves are smooth also indicates that optimisation is stable and convergence is reliable. Overall, these findings show that the model is strong enough to correctly classify data at different threshold levels.

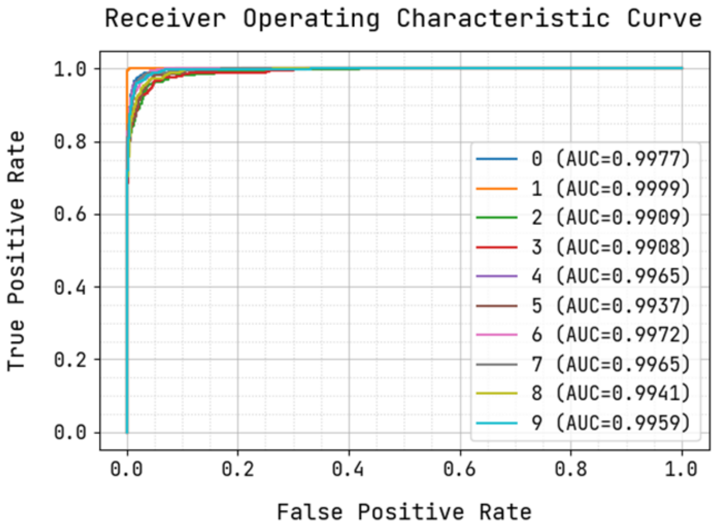

Figure 9: Training receiver operating characteristic curve

5.1. Performance Metrics

Some of the metrics used to assess the efficacy of the suggested model:

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+TN+FN}} \tag{25}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{26}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{27}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}} \tag{28}$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}} \tag{29}$$

TP and TN designate the total number of true results, whereas FP and FN indicate the total number of false results. Researchers also intended to use the ROC curve and calculated the AUC.

Table 3: Comparative analysis of the proposed model in testing values

Models	Accuracy	Precision	Recall	Specificity	F1 Score	AUC Score
SVM	0.7876	0.8654	0.7258	0.8627	0.7895	0.9089
LR	0.8407	0.8793	0.8226	0.8627	0.8500	0.9203
MLP	0.8407	0.8929	0.8065	0.8824	0.8475	0.9064
ELM	0.8407	0.8929	0.8065	0.8824	0.8475	0.9089
XGBoost	0.8053	0.8030	0.8548	0.7451	0.8281	0.8786
XGBoost-RSA	0.9792	0.897	0.897	0.9884	0.8962	0.9944

Table 3 represents the comparative analysis of the proposed model using test values. The analysis of the SVM model yielded an accuracy of 0.7876, precision of 0.8654, recall of 0.7258, specificity of 0.8627, F1-score of 0.7895, and AUC of 0.9089. In other LR models, the accuracy was 0.8407, the precision was 0.8793, the recall was 0.8226, the specificity was 0.8627, the F1-score was 0.8500, and the AUC was 0.9203. In another MLP model, the accuracy was 0.8407, the precision was 0.8929, the recall was 0.8065, the specificity was 0.8824, the F1-score was 0.8475, and the AUC was 0.9064. In another ELM model, the accuracy was 0.8407, the precision was 0.8929, the recall was 0.8065, the specificity was 0.8824, the F1-score was 0.8475, and the AUC was 0.9089. In another XGBoost model, the accuracy was 0.8053, the precision was 0.8030, the recall was 0.8548, the specificity was 0.7451, the F1-score was 0.8281, and the AUC was 0.8786. In the final XGBoost-RSA model, the accuracy was 0.9792, the precision was 0.897, the recall was 0.897, the specificity was 0.9884, the F1-score was 0.8962, and the AUC was 0.9944.

Table 4: Investigation of the projected model in training values

Models	Accuracy	Precision	Recall	Specificity	F1 Score	AUC Score
SVM	0.8407	0.8667	0.8387	0.8431	0.8525	0.9190
LR	0.8673	0.9273	0.8226	0.9216	0.8718	0.9219
MLP	0.8584	0.8966	0.8387	0.8824	0.8667	0.9330
ELM	0.8850	0.9016	0.8871	0.8824	0.8943	0.9374
XGBoost	0.8584	0.8594	0.8871	0.8235	0.8730	0.9282
XGBoost-RSA	0.9818	0.9093	0.9082	0.9899	0.9081	0.9953

In the above Table 4 represents the comparative analysis of the proposed model on training values. The analysis of the SVM model yielded an accuracy of 0.8407, precision of 0.8667, recall of 0.8387, specificity of 0.8431, F1-score of 0.8525, and AUC of 0.9190 (Figure 10).

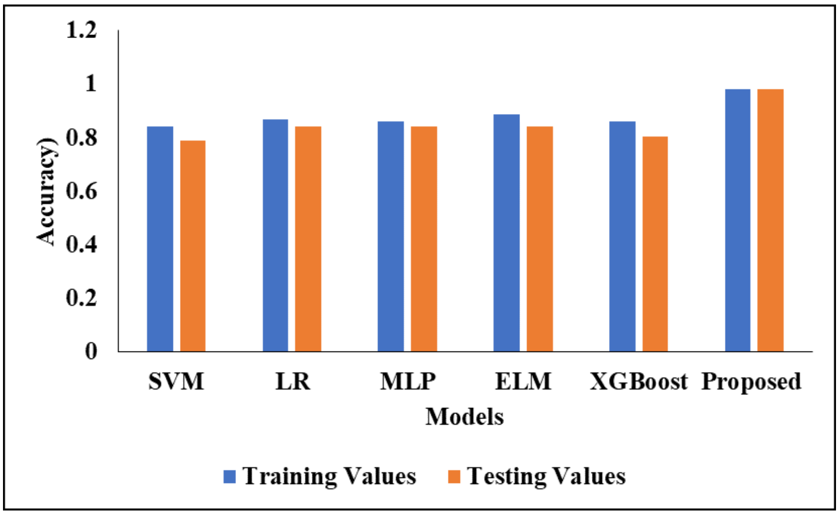

Figure 10: Comparison of the proposed model for accuracy

For another LR model, the accuracy was 0.8673, the precision was 0.9273, the recall was 0.8226, the specificity was 0.9216, the F1-score was 0.8718, and the AUC was 0.9219 (Figure 11).

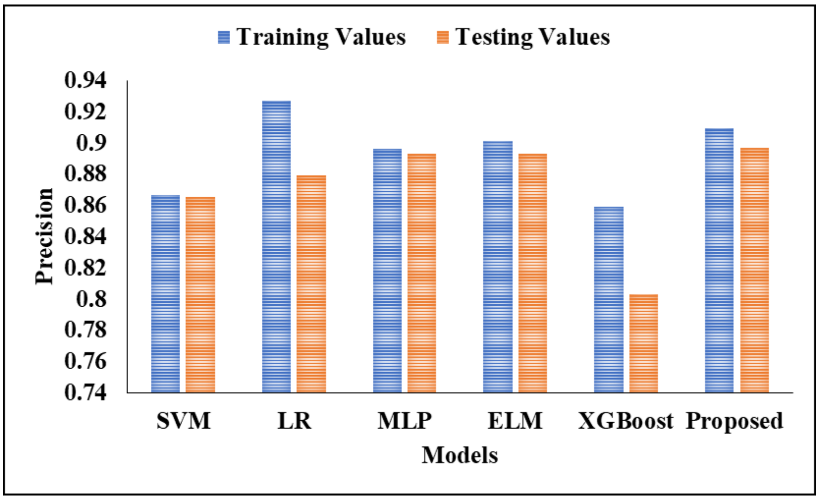

Figure 11: Validation analysis on precision

Additionally, the MLP model achieved an accuracy of 0.8584, precision of 0.8966, recall of 0.8387, specificity of 0.8824, F1-score of 0.8667, and AUC of 0.9330 (Figure 12).

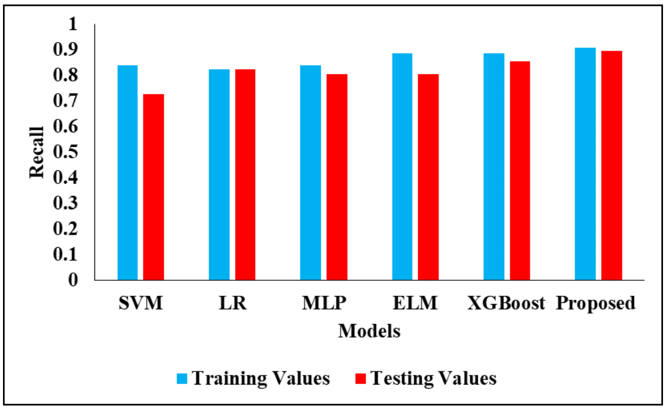

Figure 12: Recall analysis

According to the next ELM model, the accuracy is 0.8850, the precision is 0.9016, the recall is 0.8871, the specificity is 0.8824, the F1-score is 0.8943, and the AUC is 0.9374, respectively (Figure 13).

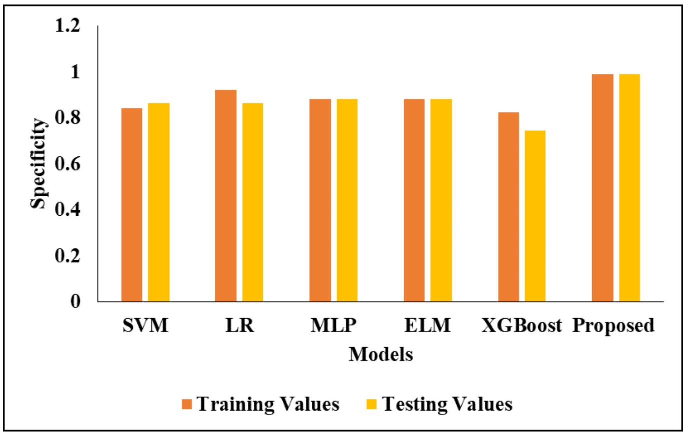

Figure 13: Specificity validation

Analysis of another kind: the XGBoost model reached an accuracy of 0.8584, a precision of 0.8594, a recall of 0.8871, a specificity of 0.8235, an F1-score of 0.8730, and an AUC of 0.928, respectively (Figure 14).

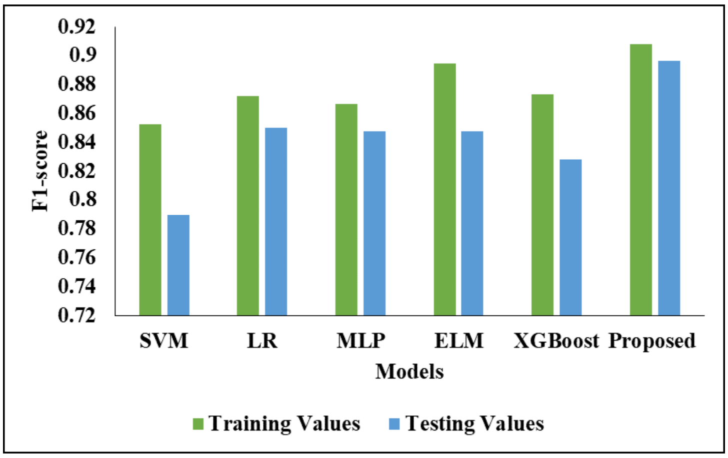

Figure 14: Graphical representation of the F1-score

And the final XGBoost-RSA model reached an accuracy of 0.9818, a precision of 0.9093, a recall of 0.9082, a specificity of 0.9899, an F1-score of 0.9081, and an AUC of 0.9953, respectively. In this comparison, the proposed model outperformed the others (Figure 15).

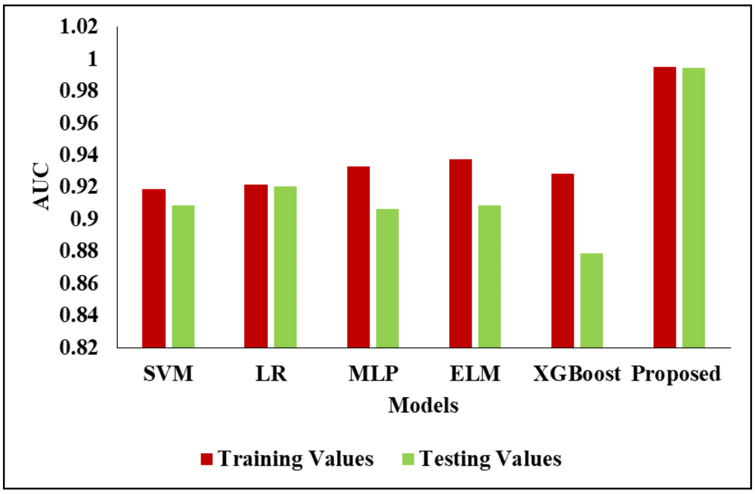

Figure 15: AUC comparison

6. Conclusion

This study proposes a cloud-based, machine learning–enabled, (IoT) intelligent system for farming. This study analyses the current state of IoT in farming by discussing key literature, advances in IoT, popular hardware, cloud platforms, agricultural applications, and IoT challenges. The quality and yield of crops can be increased by the use of IoT in agriculture. The goal of this work is to apply the presented ML method to forecast agricultural yield. In the research, the soil quality score is used as one input to an XGBoost model, alongside other characteristics known to affect crop productivity, such as temperature and rainfall. To improve XGBoost's predictive performance, RSA hyper-tunes the model's parameters. The suggested system is implemented in Python. The suggested prediction model is evaluated using the Soil dataset. Compared with previous classification models, the proposed scheme outperforms them in recall and F1-score. Extending and refining the investigation's scope by using characteristics such as soil nutrients, soil quality, and irrigated area points is possible. Together, the Internet of Things and smart agriculture built on deep learning can boost crop yields and quality. Irrigation is taken into account for productivity in the proposed work alone, and fertility factors may be considered in the future scope of work.

Acknowledgement: The authors sincerely acknowledge the support and academic resources provided by New Horizon College of Engineering, B.M.S College of Engineering, St. Joseph's College of Engineering, Karunya Institute of Technology and Sciences, Saranathan College of Engineering, and Al-Shahbaa Private University.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding authors upon reasonable request, ensuring transparency and reproducibility of the research.

Funding Statement: The authors declare that no external funding was received for the conduct of this research or the preparation of this manuscript.

Conflicts of Interest Statement: The authors declare that there are no conflicts of interest that could have influenced the outcomes or interpretations of this study.

Ethics and Consent Statement: All authors confirm their consent for the publication of this work and its unrestricted availability to the academic community and general readership for scholarly and educational purposes.

References

1. N. G. Rezk, E. E. D. Hemdan, A. F. Attia, A. El-Sayed, and M. A. El-Rashidy, "An efficient IoT-based smart farming system using machine learning algorithms," *Multimedia Tools and Applications,* vol. 80, no. 1, pp. 773–797, 2021.

2. N. A. Farooqui, A. K. Mishra, and R. Mehra, "IoT based automated greenhouse using machine learning approach," *Int. J. Intell. Syst. Appl. Eng.*, vol. 10, no. 2, pp. 226–231, 2022.
3. K. N. E. A. Siddiquee, M. S. Islam, N. Singh, V. K. Gunjan, W. H. Yong, M. N. Huda, and D. S. B. Naik, "Development of algorithms for an IoT-based smart agriculture monitoring system," *Wireless Communications and Mobile Computing*, vol. 2022, no. 1, p. 7372053, 2022.
4. A. Bhoi, R. P. Nayak, S. K. Bhoi, S. Sethi, S. K. Panda, K. S. Sahoo, and A. Nayyar, "IoT-IIRS: Internet of Things-based intelligent-irrigation recommendation system using machine learning approach for efficient water usage," *PeerJ Computer Science*, vol. 7, no. 6, pp. 1-15, 2021.
5. N. Fatima, S. A. Siddiqui, and A. Ahmad, "IoT-based smart greenhouse with disease prediction using deep learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 7, pp. 113-121, 2021.
6. S. Biswas, L. K. Sharma, R. Ranjan, S. Saha, A. Chakraborty, and J. S. Banerjee, "Smart farming and water saving-based intelligent irrigation system implementation using the internet of things," in Recent Trends in Computational Intelligence Enabled Research, *Academic Press*, Cambridge, Massachusetts, United States of America, 2021.
7. S. A. Salim, M. R. Amin, M. S. Rahman, M. Y. Arafat, and R. Khan, "An IoT-based smart agriculture system with locust prevention and data prediction," in *Proc. 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, Semarang, Indonesia, 2021.
8. E. S. Mohamed, A. A. Belal, S. K. Abd-Elmabod, M. A. El-Shirbeny, A. Gad, and M. B. Zahran, "Smart farming for improving agricultural management," *Egyptian J. Remote Sensing Space Sci.*, vol. 24, no. 3, pp. 971–981, 2021.
9. A. K. Podder, A. A. Bukhari, S. Islam, S. Mia, M. A. Mohammed, N. M. Kumar, K. Cengiz, and K. H. Abdulkareem, "IoT-based smart agrotech system for verification of urban farming parameters," *Microprocessors and Microsystems*, vol. 82, no.4, p. 104025, 2021.
10. Á. L. P. Gómez, P. E. López-De-Teruel, A. Ruiz, G. García-Mateos, G. B. García, and F. J. G. Clemente, "FARMIT: Continuous assessment of crop quality using machine learning and deep learning techniques for IoT-based smart farming," *Cluster Computing*, vol. 25, no. 3, pp. 2163–2178, 2022.
11. Y. Zhou, Q. Xia, Z. Zhang, M. Quan, and H. Li, "Artificial intelligence and machine learning for the green development of agriculture in the emerging manufacturing industry in the IoT platform," *Acta Agriculturae Scandinavica, Section B—Soil and Plant Science*, vol. 72, no. 1, pp. 284–299, 2022.
12. L. Ting, M. Khan, A. Sharma, and M. D. Ansari, "A secure framework for IoT-based smart climate agriculture system: Toward blockchain and edge computing," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 221–236, 2022.
13. R. Shailendra, J. Anitha, S. Velayutham, A. Baladhandapani, A. Srivastava, S. K. Gupta, and M. Kumar, "An IoT and machine learning based intelligent system for the classification of therapeutic plants," *Neural Processing Letters*, vol. 54, no. 5, pp. 4465–4493, 2022.
14. P. Wang, B. A. Hafshejani, and D. Wang, "An improved multilayer perceptron approach for detecting sugarcane yield production in IoT-based smart agriculture," *Microprocessors and Microsystems*, vol. 82, no. 4, p. 103822, 2021.
15. A. Sen, R. Roy, and S. R. Dash, "Smart farming using machine learning and IoT," in Agricultural Informatics: Automation Using the IoT and Machine Learning, *Wiley*, Hoboken, New Jersey, United States of America, 2021.
16. R. Manikandan, G. Ranganathan, and V. Bindhu, "Deep learning based IoT module for smart farming in different environmental conditions," *Wireless Personal Communications*, vol. 128, no. 3, pp. 1715–1732, 2023.
17. S. Mathi, R. Akshaya, and K. Sreejith, "An Internet of Things-based efficient solution for smart farming," *Procedia Computer Science*, vol. 218, no. 1, pp. 2806–2819, 2023.
18. S. R. Abdurahman, N. B. A. Karna, and A. I. Irawan, "IoT-based smart farming using machine learning for red spinach," *eProceedings of Engineering*, vol. 8, no. 6, pp. 3146-3153, 2022.
19. H. Zeng, G. Dhiman, A. Sharma, A. Sharma, and A. Tselykh, "An IoT and blockchain-based approach for the smart water management system in agriculture," *Expert Systems*, vol. 40, no. 4, p. e12892, 2023.
20. W. Abdallah, A. Mutairi, M. Kasim, and K. Al-Aubidy, "IoT-based smart monitoring and management system for fish farming," *Bull. Electr. Eng. Informatics*, vol. 12, no. 3, pp. 1435–1446, 2023.
21. V. Mamatha and J. C. Kavitha, "Machine learning based crop growth management in greenhouse environment using hydroponics farming techniques," *Measurement: Sensors*, vol. 25, no. 2, p. 100665, 2023.
22. H. K. Gupta and H. R. Shah, "Deep learning-based approach to identify the potato leaf disease and help in mitigation using IoT," *SN Computer Science*, vol. 4, no. 4, p. 333, 2023.
23. D. Rana, "Top 11 Cloud Platforms for Internet of Things (IoT)," *DZone*, 2022. Available: https://dzone.com/articles /10-cloud-platforms-for-internet-of-things-iot [Accessed by 18/09/2024].
24. D. Stevenson, "What is Firebase? The complete story, abridged," *Medium*, 2018. Available: https://medium.com/ firebase-developers/what-is-firebase-the-complete-story-abridged-bcc730c5f2c0 [Accessed by 14/09/2024].
25. P. Kirvan, N. Barney, and A. S. Gillis, "What is AWS? Ultimate guide to Amazon Web Services," *Techtarget*, 2024. Available: https://www.techtarget.com/searchaws/definition/Amazon-Web-Services [Accessed by 12/09/2024].
26. P. B. Wakhare, S. Neduncheliyan, and G. S. Sonawane, "Automatic irrigation system based on Internet of Things for crop yield prediction," in *Proc. Int. Conf. Emerging Smart Comput. Informatics (ESCI)*, Pune, India, 2020.

27. S. M. Patil and R. Sakkaravarthi, “Internet of things based smart agriculture system using predictive analytics,” *Asian J. Pharm. Clin. Res.*, vol. 10, no. 13, pp. 148–152, 2017.
28. J. Xue and B. Shen, “A novel swarm intelligence optimization approach: Sparrow search algorithm,” *Syst. Sci. Control Eng.*, vol. 8, no. 1, pp. 22–34, 2020
29. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, “Machine learning for molecular and materials science,” *Nature*, vol. 559, no. 7, pp. 547–555, 2018.
30. T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, San Francisco, California, United States of America, 2016.
31. J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
32. L. Abualigah, M. A. Elaziz, P. Sumari, Z. W. Geem, and A. H. Gandomi, “Reptile search algorithm (RSA): A nature-inspired metaheuristic optimizer,” *Expert Systems with Applications*, vol. 191, no. 4, p. 116158, 2022.