Journal of Autonomous Intelligence (2024) Volume 7 Issue 4
doi: 10.32629/jai.v7i4.1218

ORIGINAL RESEARCH ARTICLE

A novel deep learning and Internet of Things (IoT) enabled precision agricultural framework for crop yield production

D. J. Anusha[1], R. Anandan[1,*], P. Venkata Krishna[2]

[1] *Department of Computer Science & Engineering, VELS Institute of Science, Technology & Advanced Studies (VISTAS), Chennai 603203, India*

[2] *Department of Computer Science, Sri Padmavati Mahila University, Tirupati 517502, India*

*** Corresponding author:** R. Anandan, anandan.se@velsuniv.ac.in

ABSTRACT

Precision agriculture is a growing concept that frequently refers to enhancing farms via the use of up-to-date knowledge and cutting-edge technology, which in turn aids farmers by automating and improving them to increase rural profitability. This paper suggests the novel framework Deep-Plant-IoT which amalgamates the Internet of Things (IoT) and Deep learning framework for an effective prediction of crop yields which act as intelligent recommendation systems that can significantly improve the production. The framework incorporates IoT sensors and devices to collect and store the soil parameters in the cloud. Then these data are downloaded offline and the Harris Hawk Optimized Long Short Term Memory network is deployed to effectively predict crop yields that can aid in better production. Nearly 15902 data were collected for two months and Extensive testing was undertaken to employ these data to evaluate and analyze the proposed framework. Moreover, the prediction algorithm proposed in the framework is evaluated in comparison to other cutting-edge learning models. The suggested algorithm has demonstrated greater performance such that 98% accuracy, 97.23% precision, 97.0% recall, and 97.2% F1-score respectively.

Keywords: Internet of Things (IoT); artificial intelligence; precision agriculture; Harris Hawk optimization; Long Short Term Memory

ARTICLE INFO

Received: 5 September 2023
Accepted: 13 October 2023
Available online: 4 February 2024

1. Introduction

The National Crime Record Bureau (NCRB) reported that roughly 10,281 farmers passed away in 2019–2020[1]. Despite the stereotype of farmers, suicide in the agricultural industry has increased. Inefficient farming practices are still used by many workers, which results in low productivity. According to research by the Centre for Research of Developing Societies (CSDS)[2], 76% of farmers desire to give up their jobs. It further states that 74% of farmers did not get any fundamental farming-related information, such as fertilizer doses, from agriculture department personnel. The precision agriculture systems therefore support and aid farmers in automating and modernizing their operations to increase rural profitability and support the development of intelligent farming systems[3]. IoT-based devices and other emerging technologies would undoubtedly benefit society. Large-scale installations are typically challenging to implement in practical situations. This study makes use of already-made innovations, such as difficult-yet-simple-to-implement sensor-based modules.

A balanced water level in the field is maintained by a smart irrigation system built on the Internet of Things. Whenever the field is dry, it puts the water pump "on", and when it is wet, it turns it "off". Additionally, this technology will assist in monitoring the level of water while only spraying fields when necessary, which is important given that water scarcity is one of the key issues facing the globe today[4]. The three macronutrients N (nitrogen), P (phosphorus), and k (potassium) were mostly needed by crops. Lack of nutrients can result in deficiencies and harm the health of crops. Fertilizers, which can be either natural or synthetic, are used to add nutrients to the soil. The excessive use of fertilizer can harm crops and groundwater in addition to the crops themselves. Therefore, a method for recommending fertilizer doses can aid in boosting agricultural output and reducing fertilizer usage. Data can be stored using Firebase with the use of an IoT-based fertilizer dose guidance system. Results from previously acquired data stored in a Firebase database may be easily tracked. Since Schwalbert et al.[5] introduced ImageNet, deep learning has gained popularity and is becoming more widespread. It is possible to use pre-trained CNN models like VGG16, VGG19, ResNet50, and InceptionV3 to aid with imager-related classification tasks. These models can unquestionably help farmers with the automatic early detection of infections and damage to crops or plants.

Besides combining both deep learning algorithms and IoT devices for precision agriculture remains to be too complex due to the dynamic behaviour of the system. Hence the idea behind this research article is to come up with a framework that gathers all the advantages of both IoT and Deep learning frameworks. This paper's greatest contribution is not the formulation of any strategy but the intelligent framework to elevate precision agriculture that can improve the farmer's life style. The points that follow are the research article's salient contributions:

1) As far as we're aware, this is only one of the first attempts to present the intelligent framework that combines both IoT and Artificial Intelligence related to the precision agriculture.
2) The paper successfully leverages the Hyper parameter a deep learning framework that has been designed for the accurate prediction of the crop yields based on the real time collected datasets.
3) Using the real-time data that was gathered, a hawk's eye experiment was conducted to assess the suggested framework, and its performance was compared to that of other cutting-edge learning models.

The following is how the paper is structured: 1) The associated research and advancements realized in the smart precision-based agricultural system are discussed in Section 2. 2) Section 3 details about the Data collection unit using Internet of Things (IoT), and proposed deep learning frameworks. 3) Section 4 is dedicated to the experimentation and evaluation of the proposed framework by comparing with the other state-of-art learning models. 4) The paper is ended with a discussion of future research in Section 5.

2. Related works

A number of remotely sensed time-series datasets and several algorithms were tried by Ghazaryan et al.[6] in 2020 to estimate yield at the county and field scales in the United States. Long-short term memory (LSTM) after a 3D convolutional neural network (CNN) was employed in this architecture. The CNN-LSTM model showed the best accuracy for county-level analysis, with a mean percentage error of 10.3% for maize and 9.6% for soybean. The proposed model has the benefit of being applied to a real-time dataset derived from reliable geographical sources. However, the created model lowered the effectiveness of crop production forecast while reducing the relative error.

To identify the most lucrative crop given the current weather conditions, Teja et al.[7] proposed a website that used Support Vector Machine (SVM) algorithms in conjunction with past meteorological data. This approach can also forecast crop yields using weather, soil, and previous yield data. The goal of this effort is to develop a system that combines data from many sources, data analytics, and forecast analysis in order to increase agricultural production productivity and, in the long term, increase farmer profitability. The created model's advantage was its ability in identifying various pests and illnesses by handling challenging local circumstances. The robust deep learning approach uses sophisticated pre-processing techniques, which takes more time and has a higher computing cost.

Sindhu Madhuri et al.[8] introduced a decision tree supervised machine learning method in 2022 to enhance the predictions of agricultural output based on soil moisture characteristics and to obtain improved error rate and accuracy for economic growth. This approach looked at additional geographical variables and revealed phonological traits, all of which are important for predicting crop yields. The approach utilized to fuse diverse remote sensing data using histogram-based tensor modification, which integrated multisource data with varying resolutions, remained difficult.

A deep learning approach was developed by Bharathi et al.[9] in 2022, and it is known as the Modified Deep Learning approach (MDLS). It is intended to help the agriculture sector forecast the crop production level in a precise manner. This MDLS is derived from the K-Nearest Neighbour and Decision Tree Algorithms, two common learning frameworks. The suggested method takes into account factors like pesticide use, rainfall ratios, and temperature levels as prediction limits when examining agricultural production characteristics. This framework was more effective. The smart irrigation system for farms, which would have increased yields, was not, however.

Ishak et al.[10]'s proposal from 2021 describes an intelligent system that can forecast the best harvests based just on a farmer's current location, general instructions for crop preparation and producing, and a systematic strategy for marketing commodities from grower to customer. For crop modelling, we employed Random Forest Regression, Support Vector Regression, and Voting Regression methods. Using current climate, weather, and soil data for the particular location, yield projection was made. On the other hand, the market monitoring system will aid in accurate crop pricing and offer transparency to all parties involved in agricultural marketing so that they may use our system to purchase and sell their goods. The maximum degree of crop production prediction was only achieved in sugarcane, cotton, and turmeric, which was a benefit of the constructed model. For other crops like wheat, rice, etc., the range was modest.

With the help of the Kalman Filter Algorithm, we performed data pre-processing, extracted some features using Linear Discriminant Analysis, and used an improved version of the Extreme Linear Machine to predict crop yield, as suggested by Vashisht et al.[11] in 2022. Based on location, season, and cultivation area, rice crop yield has been forecasted. The created model's ability to estimate crop yield even in the presence of fertilizer—which is also used to assist soil analysis and allow farmers to make informed decisions in cases of low crop yield prediction—was a benefit. However, in a large data environment that demonstrated system complexity, the proposed model crop yield prediction proved challenging with a vast soil dataset.

A research was carried out by Hussain et al.[12] in 2022 to analyze the data and create a yield forecast model using machine learning. The farmers may use this model to determine whether or not a specific climatic element will affect their harvest. Therefore, under these circumstances, specific decision-making techniques can be applied with the goal of effectively increasing crop output. For this, we have utilized a variety of ML-based techniques. The logistic regression has archived the most appropriate accuracy. The fundamental restriction, though, was that increasingly sophisticated models weren't producing reliable findings.

A crop yield forecast model based on an optimum bidirectional gated recurrent neural network (OBGRNN) was created by ThangaSelvi and Sathish[13] in 2023. Using previous agricultural data, the OBGRNN approach seeks to anticipate crop productivity. The prediction and parameter optimization operations are carried out via the OBGRNN approach. With this approach, prediction accuracy was improved. The proposed model used a physical model to analyze direct inversion, however the combination of residual learning and multi-resolution decomposition produced artefacts. As a result, the model was rejected because of the excessive noise level.

In order to forecast the yield or success rate using the provided data for various locations, Kuriakose and Singh[14], 2022 used LSTM. This framework, which takes into consideration soil type, soil fertility, climatic conditions, rainfall, and the specific seed requirements for each crop, aids farmers in determining the sort of crop that will generate a satisfactory crop for a given season. This paradigm offers more accurate prediction. The model, however, needed an optimization model since it presented challenges throughout the process and had poor assessment performance.

For the unsupervised domain adaptation (UDA) on county-level maize yield prediction, Ma and Zhang[15], 2022 introduced a Bayesian domain adversarial neural network (BDANN). By collecting domain-invariant and task-informative features from both the source and the target domains, BDANN was trained to minimize domain shift and reliably forecast maize yield using adversarial learning and Bayesian inference. The outcomes also showed that the BDANN model generalized successfully on tiny training sets. This approach was more effective, although it had a significant level of computational complexity. **Table 1** gives the quick summary of related works.

Table 1. Quick summary of literature survey.

Authors	Techniques incorporated	Merits	Demerits
Ghazaryan et al.[6]	CNN and LSTM	Better accuracy, less relative error	Decreased efficiency
Teja et al.[7]	SVM	Better accuracy	Required more time for training and High computational cost
Sindhu Madhuri et al.[8]	Decision Tree	Better performance in terms of error rate and prediction accuracy	This framework struggles at the phase of feature extraction
Bharathi et al.[9]	K-Nearest Neighbor and the Decision Tree Algorithms	Better efficiency	It requires more time for training
Ishak et al.[10]	Random Forest Regression, Support Vector Regression and Voting Regression techniques	Better prediction accuracy	Prediction range was low for crops like wheat, rice, etc.
Vashisht et al.[11]	ELM	Better prediction accuracy	High system complexity
Hussain et al.[12]	Logistic Regression	Better prediction accuracy	High time complexity
ThangaSelvi and Sathish[13]	bidirectional gated recurrent neural network	Better prediction accuracy	Not suitable for Noisy environment
Kuriakose and Singh[14]	LSTM	Better prediction accuracy	Optimization model was required for the better performance.
Ma and Zhang[15]	LSTM	Bayesian domain adversarial neural network	Computational complexity is high

It is evident from **Table 1** that the current frameworks take more training time, have greater computational complexity, are less accurate, and are not appropriate for real-time environments. Additionally, the current framework needs optimization strategies to improve performance in terms of plant production.

3. System model

The two main phases of the proposed framework are (i) data collecting and (ii) data analytics. **Figure 1** displays the block structure of the suggested framework. The suggested structure includes crop yield can be predicted by using the Optimized training model. To achieve an accurate prediction, system undergoes training and testing. Real-time dataset is collected from IoT devices which are implanted on the two different soils such as loamy and clay. The model is validated with the real time data obtained from the IoT devices installed on the different categories of soils. The detailed description of the complete framework is described in the section before.

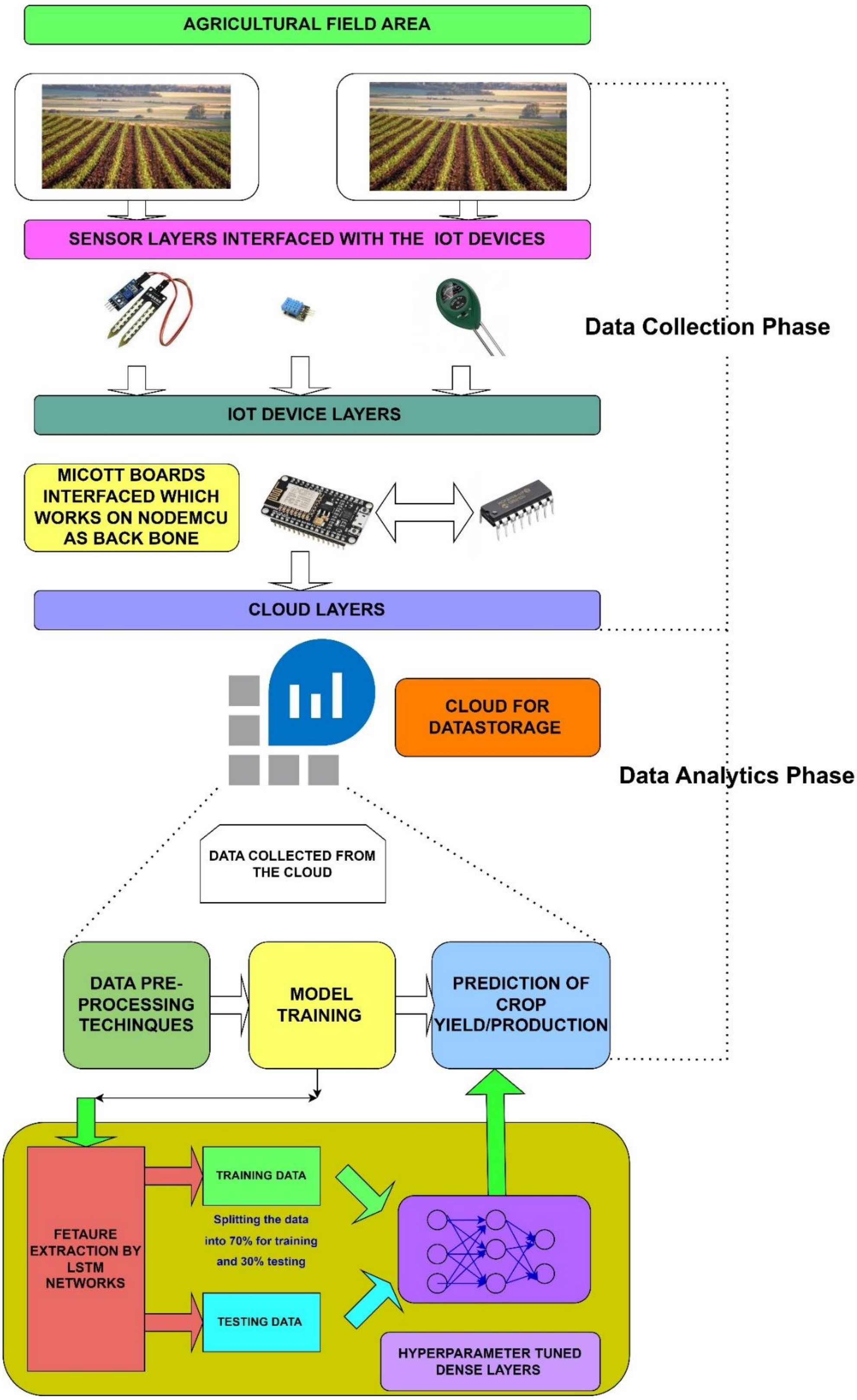

Figure 1. Overall representation of the proposed framework.

3.1. Data collection unit

In the suggested framework, this research utilizes the publicly accessible yield prediction datasets from the World Bank DataBank and the Food and Agriculture Organization (FAO) in the UN[16-17]. The variety of different crops is covered, including jute fiber maize, rice, sugarcane, & wheat for categorization purposes, including the seeds of flax lentils as well, grains, sugar cane, & wheat for yielding estimation. This research work uses the Internet of Things (IoT) for collecting the soil parameters from different types of soils as mentioned above. The primary IoT devices used to collect soil characteristics including temperature, pH, wetness, and humidity are MICOTT boards, which have an 8-bit NODEMCU as its main CPU and interface with 10-bit SPI (Serial Peripheral Interfaces) driven MCP3008 analog channels and ESP8266 WIFI transceivers. These boards are used to collect the soil parameters from the subjects and stores it in the AWS cloud for further testing. The IoT boards are powered with the 3.3V batteries and can be replaced with the other batteries when it is drained out. **Table 2** shows that total number of data

Table 2. Real time data used for the testing and evaluation.

Dataset Description	No of data	No of records	No of Attributes	Associated Tasks	Training Data /Testing
Real Time Datasets	15,900	212	05	Classification	70;30

3.2. Data pre-processing technique

In this research work, data preprocessing embraces three steps such as replacement of missing values, redundancy removal and separation. The missing value of a particular attribute is replaced by the checking with farmer’s consideration next step is reduction of data by eliminating the duplicated attributes. Finally, the data are separated into different classes based on the type of the soil and crops.

3.3. Proposed model training

This section discusses about the suggested deep learning framework to identify the different yields for the different crops.

3.3.1. Recurrent neural network

In RNN, each NN’s hidden layer is linked to the hidden layers of further nodes in a different new network. The nodes that make up of the identical layer that is concealed are connected to one another, as in recurrent NN. RNN’s ability to encode historical data in just a few milliseconds and its memory activity make it a popular choice for time series and big-data research. The RNN approach allows node with their sequences to directly form the graphs. Thus, it is possible to illustrate dynamic behavior for sequential synchronization. processes input sequences using internal memory (state). The RNN therefore uses historical data to forecast future values. Additionally, there still exists a disappearing gradient problems[18] with this method in practical applications when there is a significant interval of time between the past and the future data. As a result, the results are unsatisfactory in some real-world scenarios. With the addition of the LSTM network, RNN performance has increased in order to address this issue.

3.3.2. LSTM—An overview

Due to its flexibility in memory and suitability for large databases, a well-known learning model LSTM, is used in many applications. **Figure 2** shows the LSTM network in action.

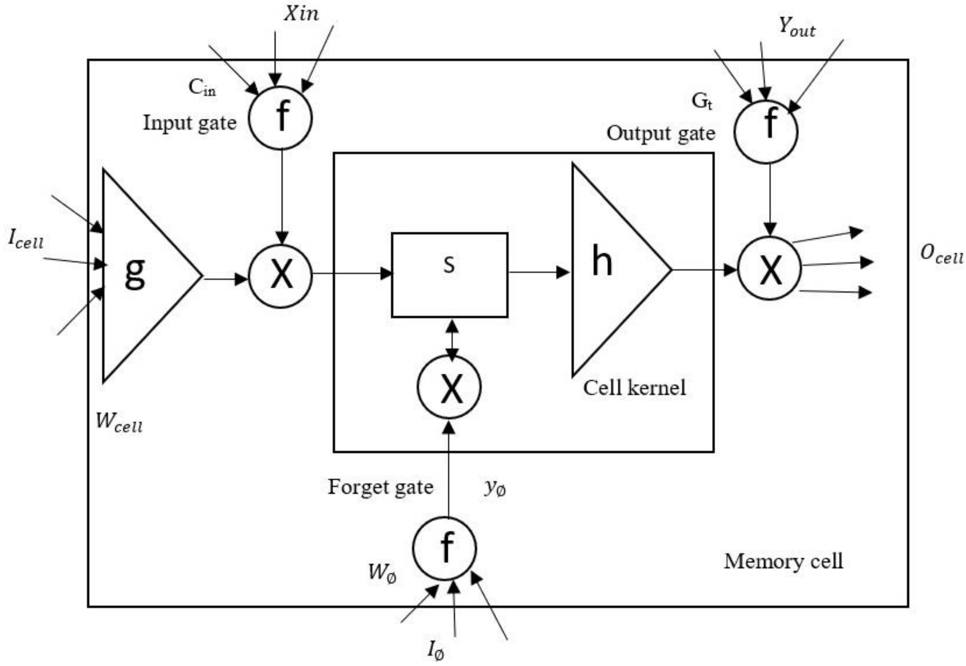

Figure 2. LSTM structure.

LSTM and the Whale optimizer are the components of the suggested hybrid learning model. Input gate (I.G), forget gate (F.G), cell input (C.I), and output gate (O.G) are the three different building components that make up an LSTM. LSTM is a type of neural network that relies on memory and remembers values after each iteration. Letting x_t, the unnoticed layer output be h_t, its former output be h_{t-1}, the cell input be C_t, the cell output be G_t, and the states of the three gates be $j_t, T_f \& T_0$. Similar to how both "Gt and ht" are transmitted to the following neural network in RNN, LSTM creation is similar. In order to update the memory, LSTM uses forget and output gates to combine the outcome of the prior unit with the present input state. The following equations are used to determine G_t and h_t.

$$I.G: j_t = \theta(G_l^i.O_t + G_h^i.e_{t-1} + s_i) \tag{1}$$
$$F.G: T_f = \theta(G_l^f.O_t + G_h^f.e_{t-1} + s_f) \tag{2}$$
$$O.G: T_o = \theta(G_l^0.O_t + G_h^o.e_{t-1} + s_o) \tag{3}$$
$$C.I: \widetilde{T_C} = \tanh\,(G_l^C.O_t + G_h^C.e_{t-1} + s_C) \tag{4}$$

where $G_l^0, G_l^f, G_l^i, G_l^C \Rightarrow$ input gates & output layers' weight matrices &$G_h^i, G_h^f, G_h^o, G_h^C \Rightarrow$the computed weight conditions among the input & hidden layers. The "s_i, s_f, s_o, s_C are the bias vectors and tanh is considered to be hyperbolic function". The computed cell outputs state is reported as follows:

$$T_C = k_t * \widetilde{T_C} + T_f * T_{t-1} \tag{5}$$
$$e_t = T_o * \tanh\,(T_C) \tag{6}$$

Using the equation above, the final outcome score is determined.

3.3.3. Reasons for the suggested model

When using big variety datasets, LSTM has a number of drawbacks[19]. This results in the need of a lot of memory cells, which often makes computations more complex and causes overfitting. It is necessary for having a computationally properly organized framework that can forecast the various crop yields in order to get around this problem. A simple framework for learning has been looked at below in order to meet the requirements mentioned above. By incorporating Harris Hawk techniques into LSTM networks, this hybrid model's main goal is to create a new hybrid algorithm.

3.4. Harris Hawk optimization

The HHO algorithm[20] was inspired by the various ways that hawks hunt and attack their prey. The three steps of HHO, a population-based optimization technology, are exploration, transforming exploration, and exploitation. In **Figure 3**, the various stages of HHO are depicted.

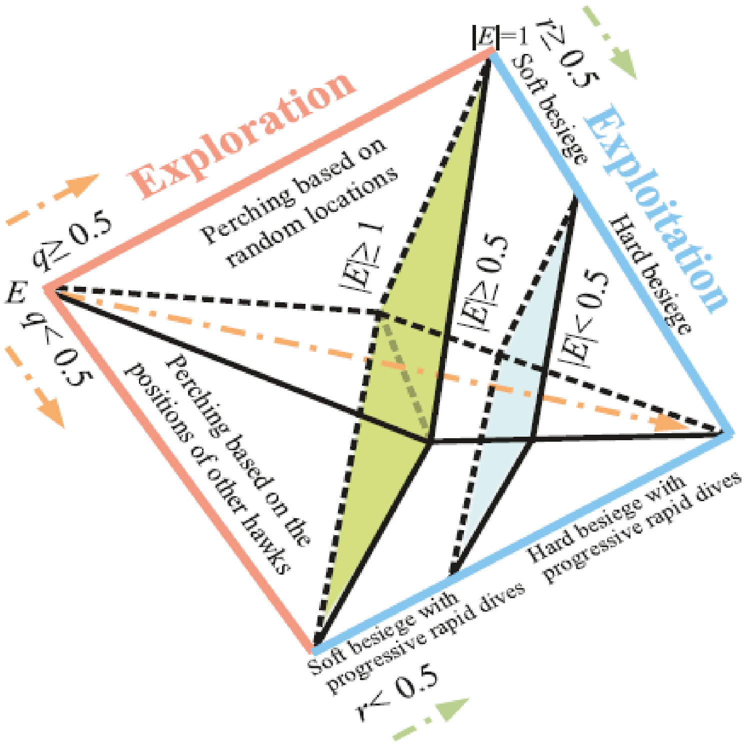

Figure 3. Illustration of various steps in Harris Hawk optimization process of catching the prey.

At this point, hawks perch in arbitrary locations dependent on the locations of other members or rabbits, which are represented as follows:

$$Z(i+1) = \begin{cases} Z_{rand}(i) - r_1|Z_{rand}(i) - 2r_2 Z(i)|, & q \geq 0.5, \\ (Z_{rabbit}(i) - Z_m(i)) - r_3(Lb + r_4(Ub - Lb)), & q < 0.5, \end{cases} \tag{7}$$

$$Z_m(t) = \frac{1}{N}\sum_{t=1}^{N} Z_i(i) \tag{8}$$

where $Z(i+1)$⇨ the updated placement of hawks with the subsequent version, $Z_{rabbit}(i)$⇨ a prey's location, & $Z(i)$⇨ the location of hawks. The modulus (absolute value) of the elements. r_1, r_2, r_3, r_4 & q⇨ random numbers among 0 and 1. **Ub** and **Lb**⇨ the variables' upper & lower bounds, $Z_{rand}(i)$⇨ the location for a randomhawkpopulation.$Z_m(i)$⇨ The typical position of the current hawk population.

3.4.1. Exploitation and exploitation transformation

The following equations are used to assess the transition stages, which takes into account the prey's escape energy:

$$E_1 = 2\left(1 - \frac{t}{T}\right) \tag{9}$$

$$E = E_0 E_1 \tag{10}$$

where t ⇨ iteration currently in use; E_0⇨ the prey's initial energy, randomly ranging among [1, 1]; and T⇨ the most possible iterations.

3.4.2. Exploitation stage

At this moment, the hawks attack the victim using its attempts to flee and four different pursuit techniques. A successful capture necessitates the presence of escaping energy (E) with the possibility of escape (r).

When $r \geq 0.5$ and $|E| \geq 0.5$, a soft besiege was conducted by hawks in the following equations, which means the prey has enough energy but gets a failed try for escaping:

$$Z(i+1) = \Delta Z(t) - E|JZ_{rabbit}(i) - Z(i)| \tag{11}$$

$$\Delta Z(i) = Z_{rabbit}(i) - Z(i) \tag{12}$$

where $\Delta Z(i)$⇨ the comparison of hawk positions at different iterations I and the present location of prey and $Z_{rabbit}(i)$⇨ the leap strength, which fluctuates at random with every iteration. r_5⇨ an arbitrary figure among 0 and 1.

Hawks applies a hard besiege to prey with low escaping energy and fails to escape, which is indicated by $r \geq 0.5$ and $|E|< 0.5$, modeled as follows:

$$Z(i+1) = Z_{rabbit}(i) - E|\Delta Z(i)| \tag{13}$$

Hawks hunt through a more intelligent soft encirclement known as gentle besiege with progressive quick dives when r and |E| are below 0.5. This behavior is depicted as follows:

$$P = Z_{rabbit}(i)0_v - E|Z_{rabbit}(i) - Z(i)| \tag{14}$$

$$Q = P + SXLF(D) \tag{15}$$

where D ⇨ the size of the issue, S⇨ a size-random vector $1 \times D$, & the equations defining the Levy flight (LF) function.

$$LF(d) = 0.01 \times \frac{u \times \sigma}{|v|^{1/\beta}} \tag{16}$$

$$\sigma = \left(\frac{\Gamma(1+\beta) \times sin\pi\beta/2}{\Gamma(1+\beta/_2) \times \beta \times 2^{\beta-1/2}} \right)^{1/\beta} \tag{17}$$

where u, v ⇨ a random vector with a normal distribution of size $1 \times d$, β ⇨ a constant with a value restricted to 1.5, and Γ ⇨ a common Gamma function. The hawk's positions can be updated by modelling

$$Z(i+1) = \begin{cases} PifF(z) < F(Z(i)) \\ QifF(z) < F(Z(i)) \end{cases} \tag{18}$$

A hard besiege is produced when the prey's energy is exhausted (r 0.5 and |E| 0.5). Equations (19) and (20), which resemble the computation of P and Q, are used. The procedure for updating is as follows:

$$P = Z_{rabbit}(i) - E|JZ_{rabbit}(i) - Z_m(i)| \tag{19}$$

$$Q = P + S \times LF(D) \tag{20}$$

$$Z(i+1) = \begin{cases} PifF(z) < F(Z(i)) \\ QifF(z) < F(Z(i)) \end{cases} \tag{21}$$

3.5. Proposed model

Harris Hawk methods are employed for optimum the weighting of dense layers in LSTM networks, as was covered in Section 3.4. Harris Hawk's criteria for the exploration and exploitation phases of the prey finding process are used in this instance as the primary term for optimum the weighting of LSTM networks. The LSTM cells are initially fed a random quantity of weights & biases. The fitness function of the suggested model is defined as its accuracy. The mathematical Equations (7), (9) and (13) are effectively utilized to calculate input bias & weights for every iteration. The LSTM network subsequently processes the weights to determine the fitness function. The loop will either stop or continue if the level of fitness functions matches to the threshold. The fitness function of the proposed framework is presented as follows

$$Fitness\ Function(FF) = Max\ performance(accuracy, Recall, Specificity, Precision, F1 - Score) \tag{22}$$

As opposed to additional metaheuristic algorithms, which require lower time for optimization and also enhance detection times, Harris Hawk optimization in this method offers a slower rate of convergence. The proposed model's pseudo code is presented in Algorithm1.

Algorithm 1 Pseudocode for the model that is suggested

```
01: Input: Epochs, Concealed Layers, Biased Weights, rate of learning
02: Output—Prediction of Crop Yields
03: Biases weights, concealed layers, epochs, and learning rate should be assigned at random.
04: Set the three parameters such as
05: Start the While loop
06: Apply Equation (6) to the LSTM cells' output.
07:         Apply Equation (22) to FF to determine it.
08:       Start the For loop from 1 to iteration(maxi)
09: Bias weights & input layers should be assigned using Equations (20) and (21).
10:                 Using Equation (22), determine FF
11: Check for (FF equal to threshold)
12:                                 jump to step 17
13:                           otherwise
14:                                 jump to step 08
15: halt
16: halt
17:    Check for (output ≤ 1)
18:             // Calculate Crop l
19:   Otherwise check for (output ≤ 2 & output > 1)
20:            // calculate Crop 2
21: Otherwise check for (output ≤ 3 & output e > 2)
22:             // calculate Crop 3
23: otherwise
24: jump to Step 09
25: halt
26: halt
27: halt
```

4. Investigation analysis

The recommended framework was run on a PC workstation equipped with an i9 CPU, 240 GB SSD, NVIDIA Titan V4 graphics card, and 3.2 GHz Python, with Keras Libraries and Tensorflow v2.1 as the backend. Accuracy, precision, recall, specificity, and F1-score were used as criteria for evaluating the model. The mathematical equation used to determine the performance measurements is shown in **Table 3**. In order to show the superiority of the proposed framework, we also computed the AUC using the confusion matrix. Early pausing is employed to address the overfitting and generalization concerns. When the validation performance of the proposed model shows no improvement over time, this procedure is used to end the iteration. To train and test the data in order to address the problem of class inequality, the uniform distribution is utilized. **Table 3** provides the mathematical formula for computing the performance measures needed to assess the suggested model.

Table 3. Performance metrics used for the evaluation.

Sl.no	Performance Standards	Statements
01	Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
02	Recall	$\frac{TP}{TP+FN}$ *x100*
03	Specificity	$\frac{TN}{TN+FP}$

Table 3. (*Continued*).

Sl.no	Performance Standards	Statements
04	Precision	$\frac{TN}{TP+FP}$
05	F1-Score	$2\times\frac{Precison\times Recall}{Precision+Recall}$

TP & TN ⇨ "True positive & True negative", FP & FN ⇨ "False positive & False negative".

Figures 4–7 depicts the ROC values for the proposed model in predicting the different crop yields based on real time datasets. With the ROC curves, Area under curves (AUC) are calculated for each and every dataset. It is found that average AUC is found to be 0.98 for predicting the all the crop yields. **Figure 8** presents the confusion matrix of the proposed algorithm.

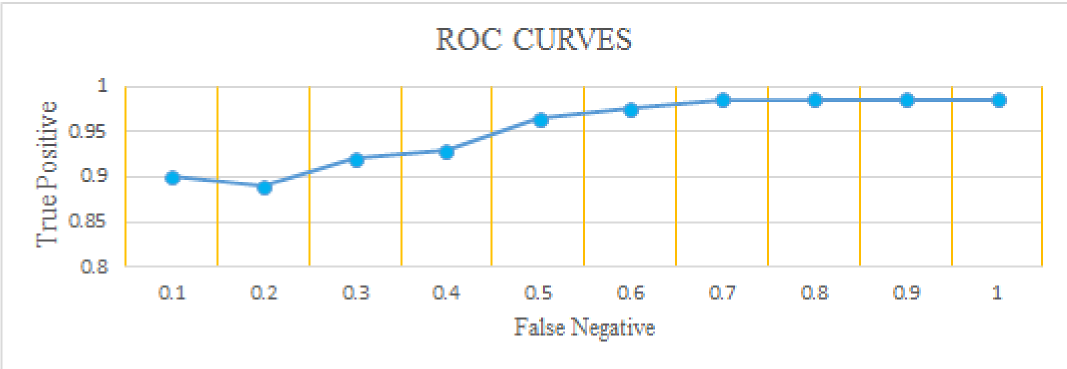

Figure 4. ROC of the suggested model in predicting the RICE crop from the real time datasets.

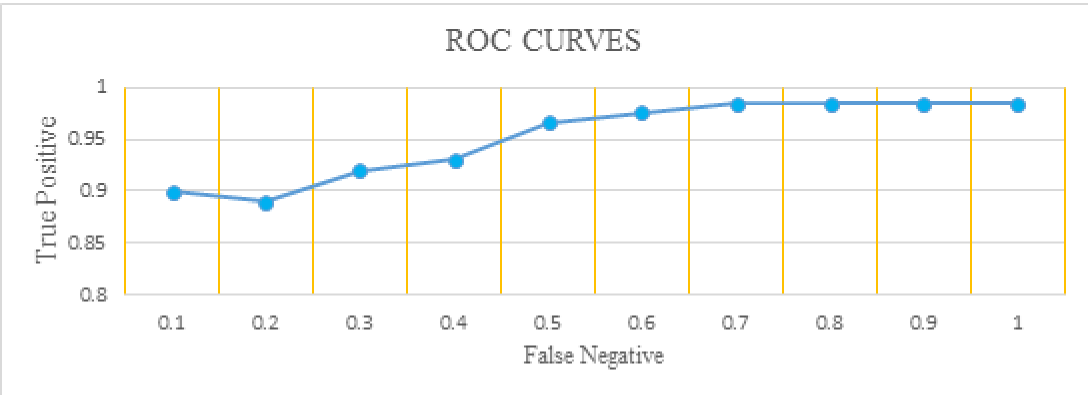

Figure 5. ROC of the suggested model in predicting the Jute crop from the real time datasets.

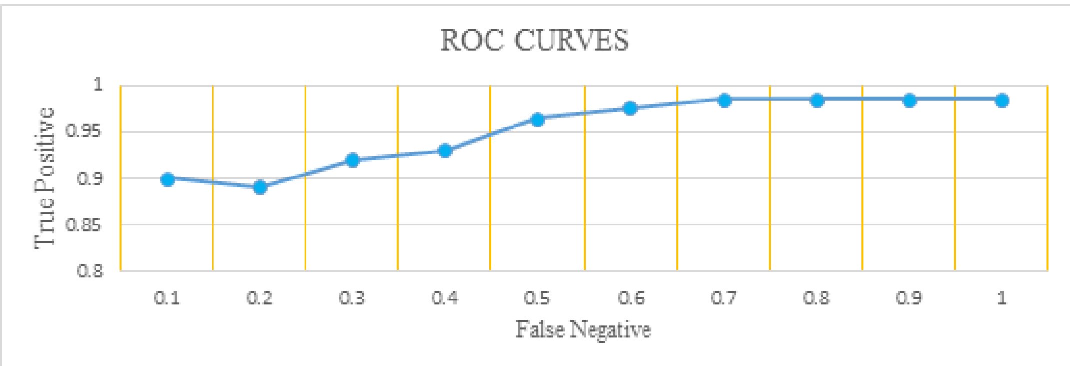

Figure 6. ROC of the suggested model in predicting the Sugar can crop from the real time datasets.

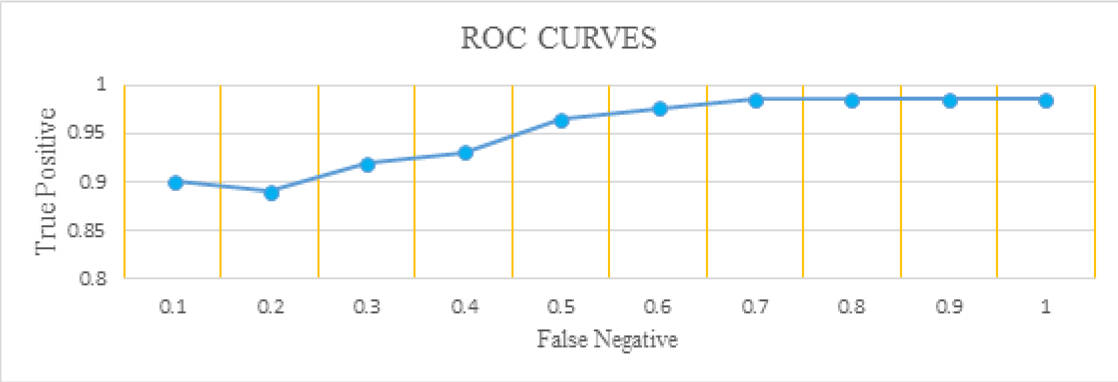

Figure 7. ROC of the suggested model in predicting the Maize crop from the real time datasets.

Actual Value

Label	Sugarcane	Jute	Rice	Maize
Sugarcane	98.0	1.2	0	0
Jute	1.0	98.2	0	0
Rice	0	0	98.2	1.0
Maize	0	0	1.0	98.2

Predicted Value

Figure 8. Matrix of confusion for the suggested model in predicting the various crop yields.

The suggested model's performance indicators are assessed for the different number of epochs. With the drop-out ratios. **Figures 9–11** shows the effectiveness of the suggested model at different drop-outs adopted during the process of training. The metrics for the suggested approach when processing the Framingham & public health datasets are shown in **Figures 9–11**. The data clearly show that the suggested model has generated the average accuracy of 98.5%,98.45% precision, 97.6% recall and F1-score 98.2% for 200 epochs for the drop-outs ranges from 0.2, 0.4, 0.6, 0.8 respectively. It proves that the integration of HHO algorithm for tuning the hyper parameters of LSTM has produced the uniform performance even though the drop-outs are increasing gradually. Moreover, the predicted values are validated with ground truth scenario in which the proposed model has produced the RMSE of 0.0034 as compared with ground truth data source.

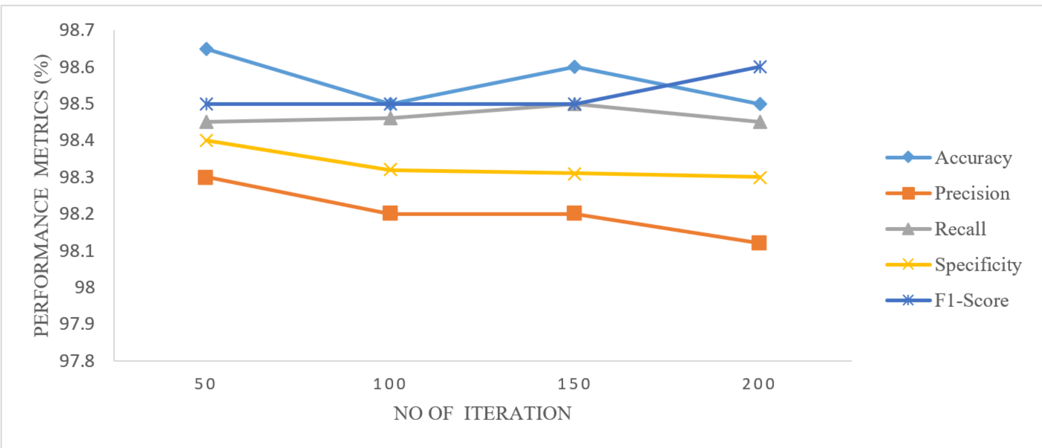

Figure 9. Effectiveness of the suggested framework with the drop-outs = 0.2 & 0.4.

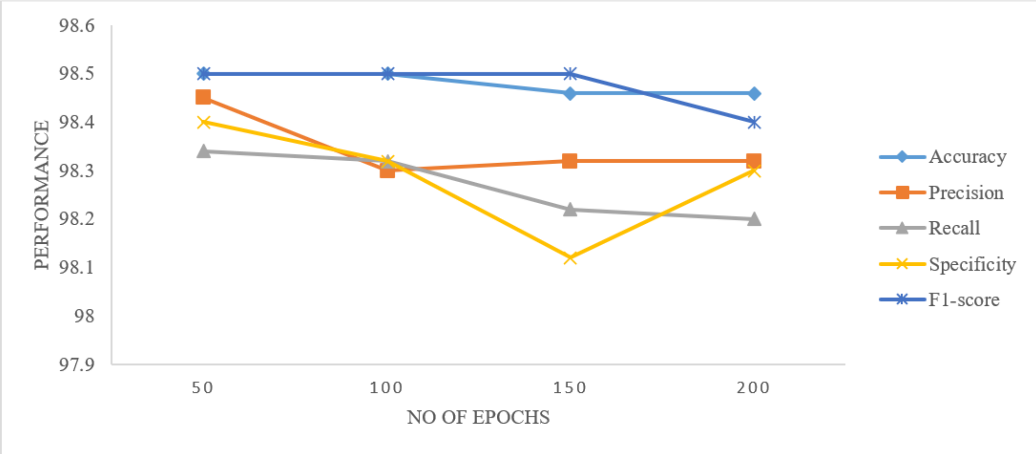

Figure 10. Effectiveness of the suggested framework with the drop-outs = 0.6 and 0.8.

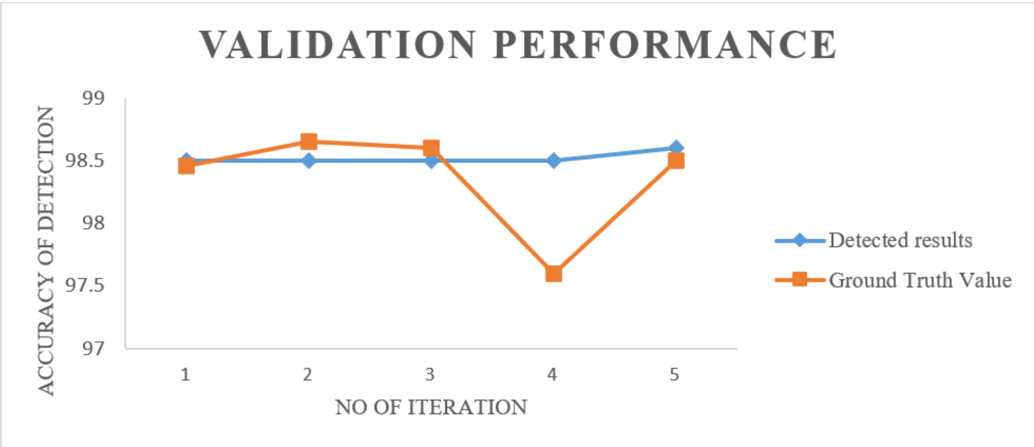

Figure 11. The suggested model's typical detection performance for different crop yields and compared with the ground truth model.

The suggested predictor has been compared with the most advanced options that are similar to this framework to demonstrate the superiority of the model that has been suggested. Tables shows the comparative analysis between the effectiveness of various algorithms while employing real time datasets. The similar state of art works which are included for the comparative studies are LSTM[20], 1D-CNN[21], DNN[22], PSO-ANN[23], BAT-LSTM[24],WHALE-SVM[25],and FIREFLY-MLP[26].

Tables 4–6 show how successfully different algorithms predicted different crop yields. It is evident from the tables and recommended model that it has provided the much better results in predicting various crop yields. On the other hand, several deep learning models have also produced impressive intermediate prediction performances (average performance of 90%). In addition, when compared to other optimized LSTM, the inclusion of HHO in LSTM has demonstrated to be crucial in forecasting crop yields. It is evident from **Table 4** through **Table 6** that the performance of the other optimization methods is clearly impacted by their propensity to become trapped in local minima.

Table 4. Comparative evaluation of the various algorithms in handling the real time datasets in predicting the rice crop yield.

Slno	Algorithms	Performance Standards				
		Accuracy	Precision	Recall	Specificity	F1-Score
01	LSTM	92.1%	91.3%	90.3%	88.4%	91.2%
02	1D-CNN	90.2%	90.23%	89.45%	88.45%	90.0%
03	DNN	87.3%	86.3%	85.3%	84.4%	85.2%
04	PSO-ANN	89.4%	87.5%	86.5%	85.6%	86.0%

Table 4. (*Continued*).

Slno	Algorithms	Performance Standards				
		Accuracy	Precision	Recall	Specificity	F1-Score
05	BAT-LSTM	92.5%	91.4%	90.4%	89.3%	91.3%
06	WHALE-SVM	93.4%	92.1%	91.5%	90.4%	90.2%
07	FIREFLY-MLP	92.1%	91.3%	90.3%	88.4%	91.2%
08	PROPOSED MODEL	98.5%	97.6%	97.0%	97.3%	97.24%

Table 5. Comparative evaluation of the various algorithms in handling the real time datasets in predicting the maize crop yield.

Slno	Algorithms	Performance Standards				
		Accuracy	Precision	Recall	Specificity	F1-Score
01	LSTM	91.2%	90.4%	89.2%	87.3%	90.2%
02	1D-CNN	90.2%	90.23%	89.45%	88.45%	90.0%
03	DNN	87.3%	86.3%	85.3%	84.4%	85.2%
04	PSO-ANN	88.4%	86.5%	86.0%	86.1%	86.2%
05	BAT-LSTM	92.5%	90.54%	89.45%	88.3%	90.2%
06	WHALE-SVM	93.2%	92.8%	92.5%	89.3%	92.9%
07	FIREFLY-MLP	92.1%	91.3%	90.3%	88.4%	91.2%
08	PROPOSED MODEL	98.5%	97.6%	97.0%	97.3%	97.24%

Table 6. Comparative evaluation of the various algorithms in handling the real time datasets in predicting the sugar cane and jute crops yield.

Slno	Algorithms	Performance Standards				
		Accuracy	Precision	Recall	Specificity	F1-Score
01	LSTM	91.2%	90.4%	89.2%	87.3%	90.2%
02	1D-CNN	90.2%	90.23%	89.45%	88.45%	90.0%
03	DNN	87.3%	86.3%	85.3%	84.4%	85.2%
04	PSO-ANN	88.4%	86.5%	86.0%	86.1%	86.2%
05	BAT-LSTM	92.5%	90.54%	89.45%	88.3%	90.2%
06	WHALE-SVM	93.2%	92.8%	92.5%	89.3%	92.9%
07	FIREFLY-MLP	92.1%	91.3%	90.3%	88.4%	91.2%
08	PROPOSED MODEL	98.42%	97.5%	97.2%	97.1%	97.30%

5. Conclusion and its future scope

This article suggests using the Internet of Things (IoT) effectively using Artificial Intelligence in precision agriculture in predicting the different yield that can aid for the better production. This kind of smart prediction systems can help the farmers to increase the yields in accordance to the soil and environmental conditions. To enable this intelligent precision agriculture, this research incorporates the IoT for the data collection process and deep learning framework for the analytics and prediction process. The IoT devices were designed and modelled using MICOTT boards interfaced with agricultural sensors such as temperature, humidity, moisture and Ph values of soils. The hyper parameter values of the dense layers are adjusted through Harris Hawk optimization technique in the LSTM that was trained on over 15,902 data. Both loamy and clay have been used in the comprehensive experiments, as well as performance standards like accuracy, precision, recall, specificity, and F1-score have been determined and contrasted to other deep learning techniques. Experimental findings showed that the suggested model worked better than the other models in use, demonstrating its strong place for the precision agriculture. In the future, these algorithms need more improvisation in terms of computational complexity which can be embedded into edge devices that can make the precision agriculture even more scalable and flexible.

Author contributions

Conceptualization, DJ; methodology, DJ and RA; software, DJ; validation, DJ, RA and PV; formal analysis, DJ; investigation, DJ; resources, DJ and RA; data curation, PV; writing—original draft preparation, DJ; writing—review and editing, PV; visualization, DJ; supervision, RA and PV; project administration, RA; funding acquisition, RA. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

The authors declare no conflict of interest.

References

1. Boukhris L, Ben Abderrazak J, Besbes H. Tailored Deep Learning based Architecture for Smart Agriculture. 2020 International Wireless Communications and Mobile Computing (IWCMC). 2020. doi:10.1109/iwcmc48107.2020.9148182
2. Oswalt Manoj S, Ananth JP. MapReduce and Optimized Deep Network for Rainfall Prediction in Agriculture. The Computer Journal. 2020;63(6):900-912. doi:10.1093/comjnl/bxz164
3. Nosratabadi S, Imre F, Szell K, et al. Hybrid Machine Learning Models for Crop Yield Prediction. arXiv2020; arXiv:2005.04155. doi:10.48550/arXiv.2005.04155
4. Maimaitijiang M, Sagan V, Sidike P, et al. Soybean yield prediction from UAV using multimodal data fusion and deep learning. Remote Sensing of Environment. 2020; 237: 111599. doi: 10.1016/j.rse.2019.111599
5. Schwalbert RA, Amado T, Corassa G, et al. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. Agricultural and Forest Meteorology. 2020; 284: 107886. doi: 10.1016/j.agrformet.2019.107886
6. Ghazaryan G, Skakun S, Konig S, et al. Crop Yield Estimation Using Multi-Source Satellite Image Series and Deep Learning. IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. 2020. doi:10.1109/igarss39084.2020.9324027
7. Teja MS, Preetham TS, Sujihelen L, et al. Crop Recommendation and Yield Production using SVM Algorithm. In: Proceedings of 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS); 2022. doi:10.1109/iciccs53718.2022.9788274
8. Sindhu Madhuri G, Paudel S, Nakarmi R, et al. Prediction of Crop Yield Based-on Soil Moisture using Machine Learning Algorithms. In: Proceedings of 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS);10-12 October 2022; Tashkent, Uzbekistan.doi:10.1109/ictacs56270.2022.9988186
9. Bharathi PS, Amudha V, Ramkumar G, et al. An Experimental Analysis of Crop Yield Prediction using Modified Deep Learning Strategy. In: Proceedings of 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI); 28-29 January 2022; Chennai, India. doi:10.1109/accai53970.2022.9752492
10. Ishak M, Rahaman MS, Mahmud T. FarmEasy: An Intelligent Platform to Empower Crops Prediction and Crops Marketing. In: Proceedings of 2021 13th International Conference on Information & Communication Technology and System (ICTS); 20-21 October 2021; Surabaya, Indonesia. doi:10.1109/icts52701.2021.9608436
11. Vashisht S, Kumar P, Trivedi MC. Improvised Extreme Learning Machine for Crop Yield Prediction. In: Proceedings of 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM); 27-29 April 2022; London, United Kingdom. doi:10.1109/iciem54221.2022.9853054
12. Hussain N, Sarfraz MS, Ishaq M. Environmental Constraints of Optimization Crop-Yield Prediction using Machine learning. In: Proceedings of 2022 International Conference on IT and Industrial Technologies (ICIT); 3-4 October 2022; Chiniot, Pakistan. doi:10.1109/icit56493.2022.9989120
13. ThangaSelvi R, Sathish M. An Optimal Bidirectional Gated Recurrent Neural Network Model for Crop Yield Prediction. In: Proceedings of 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT); 23-25 January 2023; Tirunelveli, India. doi:10.1109/icssit55814.2023.10060890
14. Kuriakose SM, Singh T. Indian Crop Yield Prediction using LSTM Deep Learning Networks. In: Proceedings of 2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT); 2022. doi:10.1109/icccnt54827.2022.9984407

15. Ma Y, Zhang Z. A Bayesian Domain Adversarial Neural Network for Corn Yield Prediction. IEEE Geosci Remote Sensing Lett. 2022; 19: 1-5. doi: 10.1109/lgrs.2022.3211444
16. DataBank. Available online: https://databank.worldbank.org/home.aspx (accessed on 5 October 2022)
17. Food and Agriculture Organization of the United Nations. Available online: http://www.FAO.org (accessed on 5 October 2022)
18. Rashid M, Bari BS, Yusup Y, et al. A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction. IEEE Access. 2021; 9: 63406-63439. doi:10.1109/access.2021.3075159
19. Yu Z, Du J, Li G. Compact Harris Hawks Optimization Algorithm. In: Proceedings of 2021 40th Chinese Control Conference (CCC); 2021. doi:10.23919/ccc52363.2021.9550421
20. Saini P, Nagpal B. Deep-LSTM Model for Wheat Crop Yield Prediction in India. In: Proceedings of 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT); 8-9 July 2022; Sonepat, India. doi:10.1109/ccict56684.2022.00025
21. Yogapriya G, Kumari RS, Suganthi S, et al. Crop Yield Identification Using CNN. In: Proceedings of 2023 International Conference on Artificial Intelligence and Knowledge Discovery in Concurrent Engineering (ICECONF);5-7 January 2023; Chennai, India. doi:10.1109/iceconf57129.2023.10084304
22. Saini P, Nagpal B. Efficient Crop Yield Prediction of Kharif Crop using Deep Neural Network. In: Proceedings of 2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES);20-21 May 2022; Greater Noida, India. doi:10.1109/cises54857.2022.9844369
23. Aditya Shastry K, Sanjay HA. Hybrid prediction strategy to predict agricultural information. Applied Soft Computing. 2021; 98: 106811. doi: 10.1016/j.asoc.2020.106811
24. Bhimavarapu U, Battineni G, Chintalapudi N. Improved Optimization Algorithm in LSTM to Predict Crop Yield. Computers. 2023;12(1):10. doi:10.3390/computers12010010
25. He B, Jia B, Zhao Y, et al. Estimate soil moisture of maize by combining support vector machine and chaotic whale optimization algorithm. Agricultural Water Management. 2022; 267: 107618. doi: 10.1016/j.agwat.2022.107618
26. Bazrafshan O, Ehteram M, Dashti Latif S, et al. Predicting crop yields using a new robust Bayesian averaging model based on multiple hybrid ANFIS and MLP models. Ain Shams Engineering Journal. 2022; 13(5): 101724. doi: 10.1016/j.asej.2022.101724