



RESEARCH ARTICLE

OPEN ACCESS

OPTIMIZING CROP YIELD THROUGH DATA-DRIVEN DECISION MAKING: A COMPREHENSIVE APPROACH TO CROP SELECTION USING MACHINE LEARNING

C. Pandi^{1*}, and M. Vedaraj²

¹Assistant Professor, Department of Computer Science and Engineering, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Avadi, Chennai, Tamilnadu, India.

²Professor, Department of Computer Science and Engineering, RMD Engineering College, Chennai, Tamilnadu, India.

¹<http://orcid.org/0000-0003-2918-5304>, ²<http://orcid.org/0000-0002-0727-9791>

E-mail: *pandic24@gmail.com, vedaraj84@gmail.com

ARTICLE INFO

Article History

Received: February 24, 2026

Reviewed: March 24, 2026

Accepted: May 15, 2026

Published: September 30, 2026

Keywords:

Crop Recommendation System (CRS),
ML,
EDA,
WaOA based MLP,
Bi-GRU.

ABSTRACT

The crop cultivation generally depends on the environmental circumstances and the production of crops plays a major part to meet the increasing requirement of food. Therefore, attaining improved efficiency and sustainability is considered highly essential in farming, due to this reason, a novel Crop Recommendation System (CRS) is innovated which enables to analyse environmental and soil conditions. The introduction CRS effectively optimises the crop selection process leading to maximized crop yield, thus increasing both productivity and agricultural reliability, which is accomplished by collecting agricultural data like temperature, humidity, rainfall and soil nutrients (nitrogen, phosphorous and potassium) including pH levels. Furthermore, pre-processing is carried out in two stages which are data cleaning to prevent data inaccuracies or missing values, and label encoding which transforms categorical data into a format applicable for Machine Learning (ML) based classifier models. In addition to this, Exploratory Data Analysis (EDA) is deployed to visualize data patterns and relationships within dataset. Moreover, for model training and evaluation, the data is then split into features and labels along with train-test split. Lastly, data classification is performed using Walrus Optimized Algorithm (WaOA) based Multi-Layer Perceptron (MLP) and Bi-Directional Gated Recurrent Unit (Bi-GRU) based neural network which enables to capture sequential dependencies within the data. Therefore, the proposed system as a whole ensures improved system stability with accurate prediction and selection of crop on the basis of the environmental and soil conditions by attaining enhanced efficiency (98%), Precision (98%), Recall (98%), F1 Score (98%) respectively. Furthermore, the developed model is implemented using Python software and the attained results depicts the enhanced model performance with highly efficient classifier function.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Agriculture is one among the leading and growing part in daily life as most of the food production depends on the crop yield [1]. Significantly, due to fast growing population, the need for food keeps on rising, which makes it essential to increase the crop production. Thus, farmers need to keep up with the increasing requirement for food by ensuring reduced loss during crop production [2]. Therefore, achieving accurate agriculture leads to improved productivity, quality, quantity and production of crops. In addition to this, forecasting and analysing the crop growth and condition is regarded as the vital part of the modern agriculture. Similarly, agricultural monitoring, specifically in observing crop production is important to attain enhanced food quality and quantity [3]. Significantly, crop production mainly depends on the environmental circumstances such as soil quality, soil condition, humidity, water quality and availability [4]. Farming or prediction agriculture is a modern farming approach which utilizes advanced techniques for increasing the crop production with minimised wastage. The major intention of smart farming is to improve the crop production by analysing the conditions of the environment and the soil which includes temperature, humidity and rainfall [5].

To accomplish this, various modelling methods are implemented for different crop production. However, these methods were incapable of providing spatial data dependencies of small farms for optimizing crop production [6]. Thus, considering these issues, ML based model have gained increased attention due to their enhanced solving and understanding of complex data patterns [7]. ML based crop analysis and recommendation systems are developed. Figure 1 illustrates the ML based crop recommendation and analysis process.

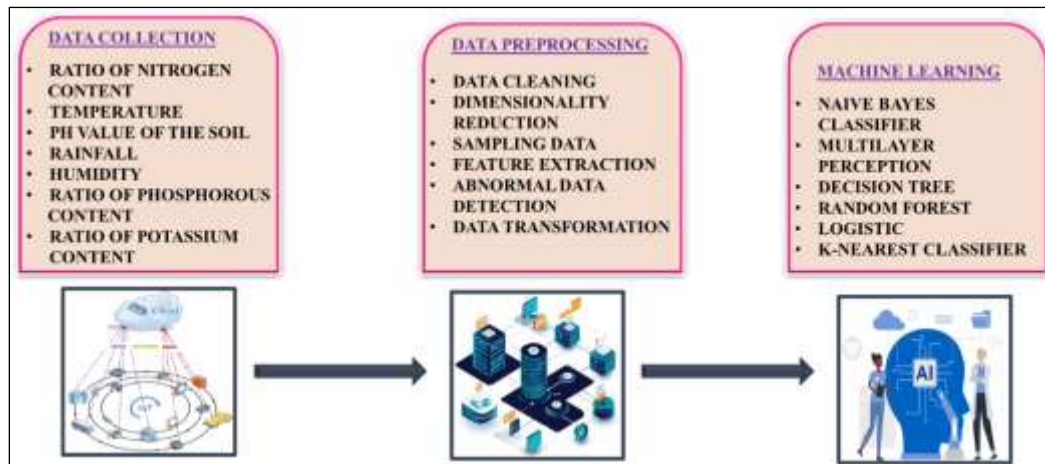


Figure 1: ML based Crop recommendation process.

Source: Authors, (2026).

Various ML based models are regarded and analysed for different crops, such in [8] authors utilized Artificial Neural Network (ANN) model which depicted improved performance in crop estimation by enabling precise prediction of crop yield even under varying environmental and agricultural factors. However, ANN efficiency depends on quality of input data. Significantly, in [9] authors developed a K-Nearest Neighbour (KNN) based crop recommendation model which achieves improved recommendation accuracy and precision for selecting the suitable crops based on environmental conditions such as soil type and water quality, thus, providing a more reliable and stable technique. Furthermore, the authors in [10] deployed Decision Tree (DT) based CRS, which allows enhanced crop prediction accuracy with reduced losses and increased system robustness. But DT faces certain limitations including overfitting and non-linearity issues. Consecutively, in [11] Random Forest (RF) based approach enables to identify intricate data features while providing increased robustness towards overfitting problems. Nevertheless, RF leads to increased complexity and are incapable of handling larger datasets. Furthermore, [12] developed Support Vector Machine (SVM) based approach is implemented to attain highly effective classification process in predicting crop suitability depending on the agricultural factors. Despite this, SVM is regarded as computationally expensive and are also highly sensitive towards parameters which makes it essential to attain enhanced tuning. Thus, various optimization approaches are considered for enhancing the performance of the classifier such as Genetic Algorithm (GA) [13], Particle Swarm Optimization (PSO) [14] and Grey Wolf Optimization (GWO) [15] which enhances the classification performance. To be specific, GA provides improved optimization by handling wide range of dataset with improved robustness against local optima. Despite this, GA struggles with increased complexity. Similarly, PSO offers enhanced convergence with enhanced ability to attain finest solution for choosing the optimum crop recommendation. In spite this, PSO produces increased complexity which affects the overall system performance. Significantly, implementing GWO based approach improves the classification process leading to enhanced accurate and precise prediction based on the environmental and agricultural factors. However, GWO is highly sensitive towards parameters which in turn decreases the convergence speed.

Recent Works

The summarized performance of various models for crop recommendation is illustrated in Table 1 with utilized dataset and drawbacks as discussed below.

Table 1: Recent development models for crop recommendation system.

Ref.no & Author	Dataset	models	Advantages & limitations
In [16] Sharma et al (2020)	Data collection from various sources	Bayesian and SVM	Offering flexibility and broader applicability. However, the heterogeneity in data sources introduce inconsistency and noise.
In [17] Ranaweera et al (2023)	Vegetable dataset of srilanka	MSP and SMO	Provided accurate results specific to regional crops, though the limited geographic scope constrains wider application.
In [18] singh et al (2023)	Collected from 45 rural areas in west Bengal	LR	Yielding interpretable and lightweight models ideal for deployment in low-resource settings, yet at the cost of lower prediction accuracy is complex.
In [19] Thilakarathne	From Kaggle, Crop recommendation dataset were collected	XGBoost	Benefiting from high model performance and robust handling of missing data, although it requires considerable computational resources.
In [20] charbuty et al (2021)	Data generation in private	DT	Enabling simple implementation and interpretation, also prone to overfitting and lack generalizability without pruning techniques.

In [21] priyadharshini et al (2021)	Yield dataset from kaggle and government websites	KNN	Achieving reasonable prediction accuracy suffering from high computational cost and sensitivity to irrelevant features.
In [22] Shehadeh et al (2021)	Economic analysis from bureau dataset	LightBGM	Capitalizing on its high speed and accuracy for large-scale data, but requires careful hyperparameters tuning to avoid overfitting.
Shams, M,Y et al (2021)	From historical Indian on crops	XAI-GNB	Enhanced interpretability and transparency through explainable AI, yet faced limitations in handling high-dimensional data
In [23] Shook et al (2022)	Unique genotypes dataset	SVM-RBF	Delivering high precision for non-linear classification tasks, but at the cost of longer training times and difficulty in tuning kernel parameters
In [24] Tabar et al (2023)	Time series from small farmlands	LSTM	Enabling effective learning of temporal dependencies, yet demanding substantial training data and computational power.

Source: Authors, (2026).

Therefore, considering these limitations and shortcomings, this work develops a CRS using WaOA based MLP and Bi-GRU model for attaining accurate and precise classification for analysing crops and maintaining their production in order to achieve increased crop quality and yield. The major outline and contributions of the developed model is listed as follows,

- To develop a CRS to attain optimum crop selection for increasing the crop quality and production, thus improving framing reliability and sustainability.
- To integrate data pre-processing stages such as data cleaning and label encoding to enhance input data quality which leads to improved classification
- To deploy EDA to visualize data patterns and relationships within the dataset for simple and easy data interpretation.
- To introduce a novel WaOA optimized MLP and Bi-GRU for capturing sequential dependencies of the data, thus, providing improved crop sustainability on the basis of the environmental data collected.

II. PROPOSED METHODOLOGY

Figure 2 comprises of the developed CRS model using ML based approach, which is developed using WaOA optimized MLP and Bi-GRU for achieving enhanced crop prediction. The process with collecting agricultural and environmental data like temperature, humidity, soil and water quality. Originally, the data collected contains uncertainties and inaccuracies which needs to be removed, thus, these collected data undergo pre-preprocessing in two stages, where data cleaning is performed to ensure higher quality output by handling missing values. Secondly, label encoding is performed which converts data into form that is suitable and applicable for ML based model processing.

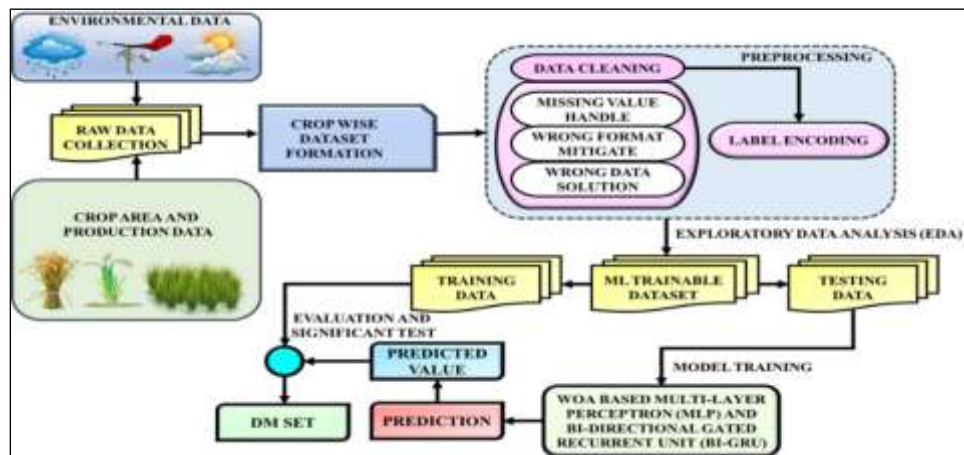


Figure 2: Block Diagram of the proposed CRS model.

Source: Authors, (2026).

After pre-processing stage, EDA is deployed for analysing and visualizing data patterns and relationships within dataset. Now, pre-processed dataset is divided into two sets of data namely, training set and testing set. Here, training set is utilized to train WaOA optimized MLP and Bi-GRU based ML model. This trained model later produces predicted values, which then goes through evaluation and analysis. The testing set is used for analysing model accuracy. Finally, the predicted value enables to attain enhanced crop prediction based on agricultural and environmental circumstances. Therefore, the system as a whole offer significantly, improved system performance by providing highly accurate and precise prediction of crops.

II.1 DATASET DESCRIPTION

Data used to forecast which crop or crops is the most suitable for cultivation in a given region based on a number of variables usually makes up a Crop Recommendation dataset. These elements include environmental variables like sunshine and wind speed, soil characteristics like pH, texture, and fertility, and climatic conditions like temperature, humidity, and rainfall.

Historical crop production information, farmer practices, and other agricultural characteristics including irrigation and pest control methods, as well as local agricultural knowledge, included in the dataset. These elements are used by the model in DL for crop selection in order to identify patterns and correlations that allow it to recommend the best crops for particular circumstances. The model is trained with labeled data (such as the yield of a specific crop in a given environment) and then generalizes to suggest crops for other, unexplored regions with comparable characteristics.

II.2 MODELLING OF DATA PRE-PROCESSING

Data pre-processing is considered as an important part, which is also the initial step to prepare the original data for further analysis. Here, data pre-processing comprises of two stages data cleaning and label encoding which improves the data quality enabling more accurate, reliable and stable prediction.

Data Cleaning:

Data cleaning is the process of rectifying and mitigating the data inaccuracy and missing data values to improve the data quality and reliability.

Label Encoding:

Label Encoding is where dependent variable or parameter 'LABEL' of the dataset is encoded. Where, this variable comprises of both categorized and non-numerical values. Here, the label variable consists of the crop names. As these data are non-numerical it is essential to encode them into numerical data to make it applicable and suitable for classification process. Hence, these labels are encoded before they are fed to the classifier.

Data splitting:

After, evaluating and representing the data using EDA, the next essential process is to divide the dataset into training and testing set for model construction. Generally, the training set is often referred as the data that is fed into the model for training. Utilizing this data, ML based model are trained to attain accurate predictions to obtain the desired output. Significantly, testing set is used to define and provide precise evaluation and analysis of finalized model fitness on training set. The major intention of the developed system is to determine the crops using ML based classifiers model, thus, here, composition of MLP and Bi-GRU is employed which is further optimized using WaOA for further enhanced and improved performance.

II.3 MODELLING OF WAOA OPTIMIZED MLP-BI-GRU

MLP

The integration of MLP with Bi-GRU provides enhanced classification and feature extraction by achieving the benefits of both MLP and Bi-GRU. Multi-Layer Perceptron is generally a supervised feed forward neural network, comprising of three layers namely input, hidden and output layer. Here, hidden layer and output layer utilizes a non-linear activation function, total input attained by each neuron in the layer is designed using,

$$x_j^{d+1} = \varepsilon_i y_i^d w_{ij}^d - \alpha_j^{d+1} \quad (1)$$

Where, y_i refers to the state of i_{th} neuron in d_{th} layer, w_{ij} implies weight of the i_{th} neuron in d_{th} layer and the j_{th} neuron in layer $d + 1$. Whereas, each neuron layer output is determined using,

$$y_j^{d+1} = \frac{1}{1 + e^{-x_j^d}} \quad (2)$$

Figure 3 demonstrates the MLP structure contains input, hidden and output layer.

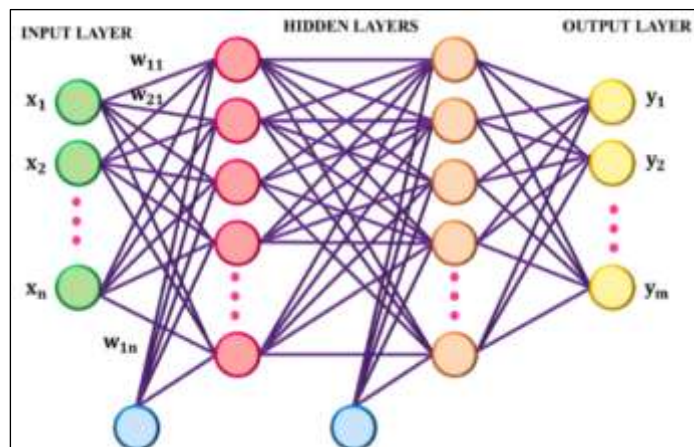


Figure 3: MLP Structure.
Source: Authors, (2026).

BiGRU

The Bi-GRU based network is represented in Figure 4, which comprises of a future layer where the data sequence is positioned in opposite direction. Here, two hidden layers are utilized for capturing intricate information from the current and previous layers which are connected to the same output layer. Utilizing this approach, enables bi-GRU to attain maximum extraction of information, thus improving the learning process.

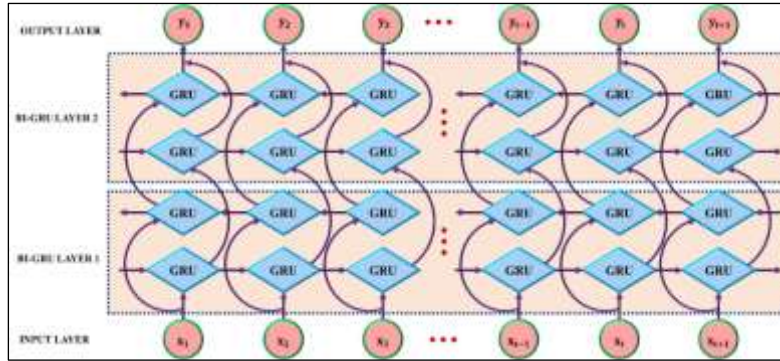


Figure 4: Bi-GRU Structure.

Source: Authors, (2026).

In Bi-GRU network consisting of two intermediate layers, where the output later overlaps and normalizes the outcomes of both forward and backward layers.

$$\overrightarrow{h}_t^1 = f(w_{xh^1} x_t + w_{h^1 h^1} \overrightarrow{h}_{t-1}^1 + b_{h^1}) \quad (3)$$

$$\overleftarrow{h}_t^1 = f(w_{xh^1} x_t + w_{h^1 h^1} \overleftarrow{h}_{t+1}^1 + b_{h^1}) \quad (4)$$

$$\overrightarrow{h}_t^2 = f(w_{h^1 h^2} \overrightarrow{h}_t^1 + w_{h^2 h^2} \overrightarrow{h}_{t+1}^2 + b_{h^2}) \quad (5)$$

$$\overleftarrow{h}_t^2 = f(w_{h^1 h^2} \overleftarrow{h}_t^1 + w_{h^2 h^2} \overleftarrow{h}_{t+1}^2 + b_{h^2}) \quad (6)$$

$$y_t = g(w_{h^2 y} \overrightarrow{h}_t^2 + w_{h^2 y} \overleftarrow{h}_t^2 + b_y) \quad (7)$$

Here, $\overrightarrow{h}_t^1$ and $\overleftarrow{h}_t^2$ represents the output vectors of the forward layers of first and second layer of Bi-GRU network. Significantly, $\overleftarrow{h}_t^1$ and $\overrightarrow{h}_t^2$ indicates the output vectors of backwards layers of the network. Consecutively, f implies the process of GRU network, g denotes the activation function and w and b refers to the weight and bias respectively. The output produced by both MLP and Bi-GRU are combined, which captures both non-sequential and sequential representations from MLP and Bi-GRU. These combined outputs are then passed via a fully connected layer, which generates the prediction.

WaOA

In addition to this, to further improvise the performance efficacy of MLP and Bi-GRU, it is essential to attain enhanced tuning performance, due to this reason, WaOA based optimization algorithm is deployed, which enables to achieve optimum tuning of network parameter, thereby, increasing the classification accuracy. WaOA is designed and developed based on the characteristics of walrus, WaOA is an intelligent approach derived based on the natural behaviour of walrus. WaOA is naturally a metaheuristic method based on population, where each walrus indicates an optimum solution to optimization problem and the position of each walrus represented using a vector. The population of the walrus is mathematically given as population matrix. This process begins with initialization, where the population of the walrus is randomly initialized, which is expressed using,

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix}_{N \times m} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,j} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & \cdots & x_{i,j} & \cdots & x_{i,m} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,j} & \cdots & x_{N,m} \end{bmatrix}_{N \times m} \quad (8)$$

Where, X implies the walrus population, X_i denotes the i_{th} walrus solution, $x_{i,j}$ indicates j_{th} decision variable, N and m refers to the number of walrus and decision variables. The objective function is determined using,

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix}_{N \times 1} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix}_{N \times 1} \quad (9)$$

Where, F denotes the objective function, F_i refers to the objective function value on the basis of i_{th} walrus. The process of WaOA is carried out in three different stages according to their behaviour.

Stage 1: Exploration Phase (Hunting Strategy)

The hunting process begins by grazing the sea floor, seeking and searching for food in flipper and susceptible motion. During, this process strongest walrus with longest tusk guides the other the walrus. Here, the length of the walrus is the objective function value quality. Henceforth, the best walrus with the optimum objective function is selected as the strongest walrus among the group. This hunting process of walrus leads to various scanning regions of search space, which in turn improves exploration process of WaOA in global search space. And the feeding mechanism of the walrus is evaluated is as,

$$x_{i,j}^{p_1} = x_{i,j} + rand_{i,j} \cdot (SW_j - I_{i,j} \cdot x_{i,j}), \quad (10)$$

$$X_i = \begin{cases} X_i^{p_1}, & F_i^{p_1} < F_i, \\ X_i, & \text{else} \end{cases} \quad (11)$$

Where, $X_i^{p_1}$ refers new position of i_{th} walrus, $x_{i,j}^{p_1}$ indicates j_{th} dimension, $F_i^{p_1}$ implies the objective function value, $rand_{i,j}$ refers to the random number between the interval $[0,1]$, SW denotes the strongest walrus and lastly, $I_{i,j}$ represents the randomly chosen integer.

Stage 2: Migration Phase

Generally, one among the most natural behaviour of the walrus is their migration process which happens because of the warming of air during the summer season. This migration process usually guides the walrus in their search space to identify applicable regions in the search space, which is represented using,

$$x_{i,j}^{p_2} = x_{i,j} + rand_{i,j} \cdot (x_{k,j} - I_{i,j} \cdot x_{i,j}), \quad F_k < F_i \quad (12)$$

$$X_i = \begin{cases} X_i^{p_2}, & F_i^{p_2} < F_i, \\ X_i, & \text{else} \end{cases} \quad (13)$$

Where, $X_i^{p_2}$ indicates the new position of the walrus, $x_{i,j}^{p_2}$ implies the j_{th} dimension, $F_i^{p_2}$ denotes the objective function value, where $X_k, k \in \{1, 2, \dots, N\}$ respectively.

Stage 3: Exploitation Phase

Walrus are often attacked by polar bears and killer whales, thus, the escaping strategy and fighting against them leads to changed walrus position. The new updated position of the walrus is expressed using,

$$x_{i,j}^{p_3} = x_{i,j} + (lb_{local,j}^t + (ub_{local,j}^t - rand \cdot lb_{local,j}^t)) \quad (14)$$

$$Local\ bounds: \begin{cases} lb_{local,j}^t = \frac{lb_j}{t} \\ ub_{local,j}^t = \frac{ub_j}{t} \end{cases} \quad (15)$$

$$X_i = \begin{cases} X_i^{p_3}, & F_i^{p_3} < F_i, \\ X_i, & \text{else} \end{cases} \quad (16)$$

Where, $lb_{local,j}^t$ and $ub_{local,j}^t$ implies the lower and upper bounds of the variable respectively. Following algorithm demonstrates the Pseudo code of proposed WaOA tuning MLP-BiGRU.

Pseudo code of proposed algorithm

Step:1 Input the parameters of MLP-BiGRU

Step:2 Population size (N) and Maximum Iterations (T)

Step:3 Initialize population (X) randomly

Step:4 Set Iteration count $t=1$

Step:5 Calculate Objective Function for each solution in initial population

Step:6 While ($t \leq T$) **Do**

Step:7 For $i=1$ to N **Do**

Step:8 Phase 1: Update the strongest walrus based on objective function

Step:9 Calculate function X^{p1} for phase 1

Step:10 Update X^i using X^{p1}

Step:11 Phase 2: Select X_k randomly as the migration destination for X_i

Step:12 Calculate function X^{p2} for phase 2

Step:13 Update X^i using X^{p2}

Step:14 Phase 3: Apply additional exploration and exploitation

Step:15 Calculate function X^{p3} for phase 3

Step:16: Update X^i using X^{p3}

Step:17: End

Step:18 Save the best solution found so far

Step:19 $t = t + 1$ // increment iteration counter

Step:20 End while

Step:21 Output the best optimal solution for objective function

Step:22 **End**

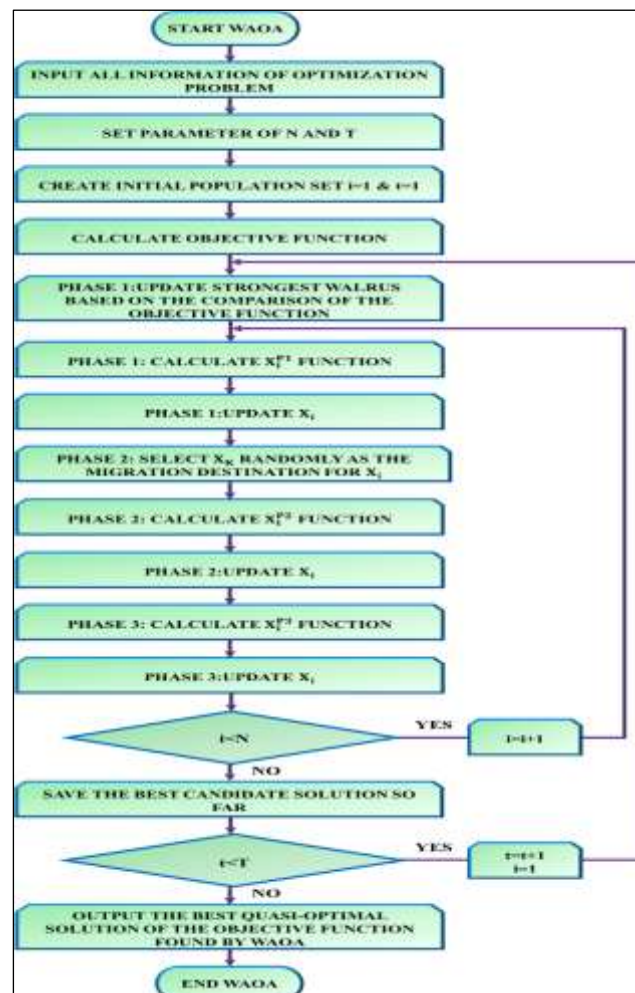


Figure 5: Flowchart of WAOA

Source: Authors, (2026).

Figure 5 displays the WAOA flowchart. Therefore, utilizing WAOA based MLP-Bi-GRU enables highly efficient and enhanced classification performance with increased accuracy and precision for selecting optimum crop on the basis of the environmental aspects. Thus, the overall system attains significantly improved performance with better functioning.

III. RESULTS AND DISCUSSIONS

The proposed system utilizing data pre-processing and enhanced WaOA optimized MLP-Bi-GRU is validated through Python Software and their obtained results are discussed and elaborated in this section. In addition to this, to define the performance efficacy of proposed model, comparative analysis against various conventional approaches are also depicted along with the results.

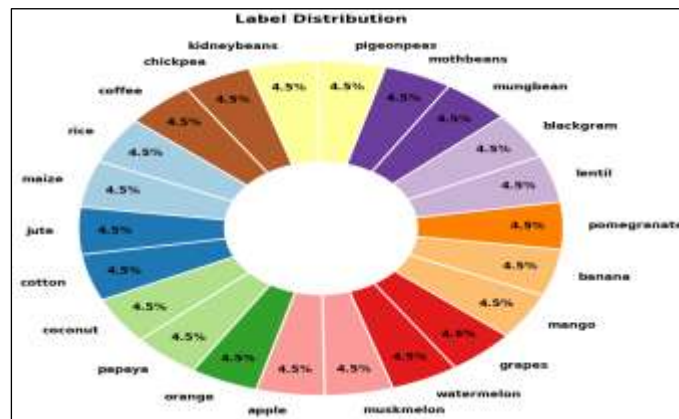


Figure 6: Label Distribution.

Source: Authors, (2026).

Figure 6 represents the Label Distribution chart comprising of various crops that are equally and evenly distributed with 4.5% each, indicating a balanced distribution of crops. Figure 7 indicates seven different histogram plots consisting of several agricultural aspects, where the first column depicts Nitrogen (N), Phosphorous (P) and Potassium (K), the second column of plots represents the temperature, humidity and pH in which temperature is at its highest around 25-30° C, humidity consists of various highest points and pH falls between 6-7 respectively. Finally, the last plot showcases the rainfall that varies against varying values.

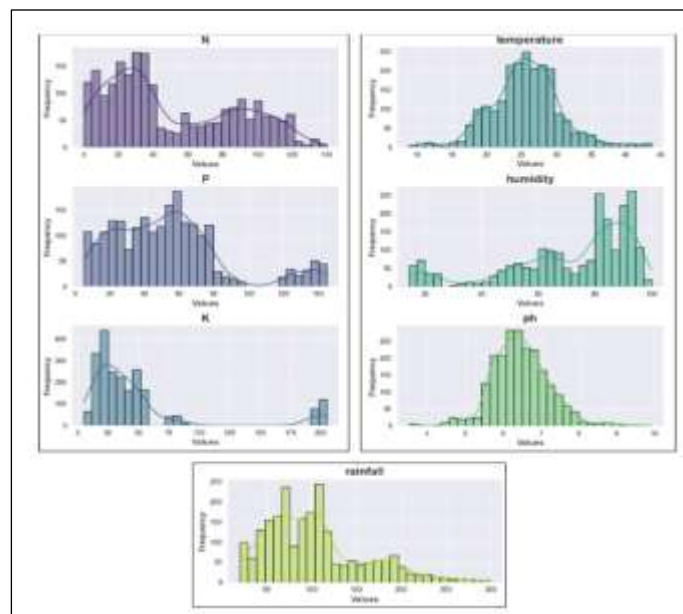


Figure 7: Waveform of Agricultural Parameters.

Source: Authors, (2026).

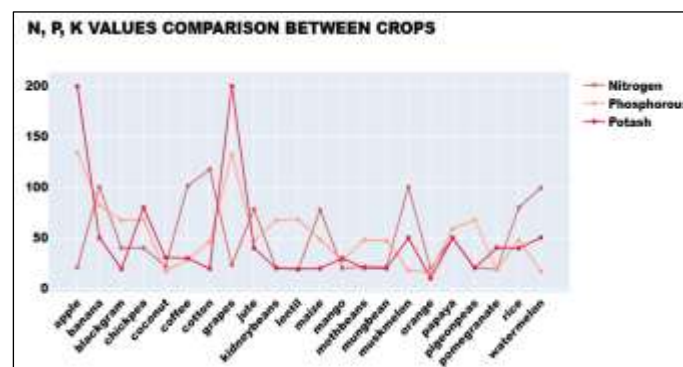


Figure 8: N, P and K values comparison.

Source: Authors, (2026).

Figure 8 denotes the comparison of N, P and K for various crops, where x-axis depicts various crop types and y-axis depicts their corresponding values. From the above chart it is understandable that, K showcases high value of around 200 for apple and graphs, Similarly, Apple and graph also depicts highest P value of around 140. Significantly, banana, muskmelon and cotton comprises of highest N value which falls under 100 and below respectively.

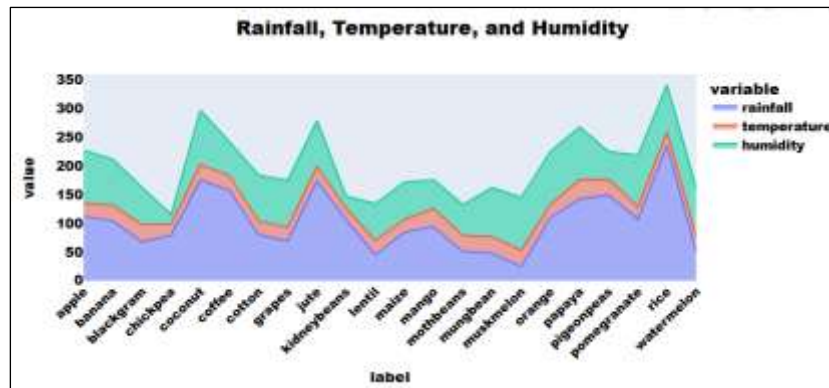


Figure 9: Waveform of Temperature, rainfall and humidity.

Source: Authors, (2026).

Figure 9 implies the rainfall, temperature and humidity for various listed crops, where x-axis represents the crop category and y-axis depicts their specific values ranging between 0-350 respectively. Here, rice showcases the highest value of rainfall, temperature and humidity, while muskmelon depicts the lowest value of rainfall, temperature and humidity. Each variable showcases varying values based on the crop type, referring to easy and enhanced crop selection process for suitable environmental parameters.

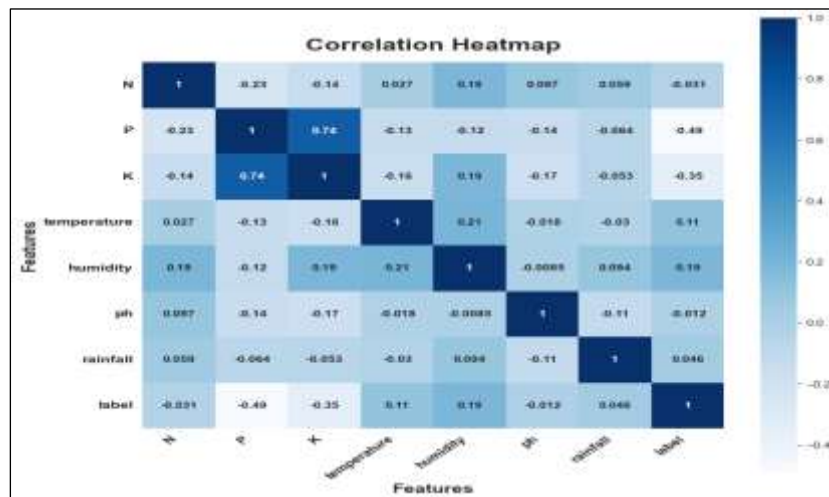


Figure 10: Correlation Heatmap.

Source: Authors, (2026).

Figure 10 represents the correlation heatmap where x and y axis depict various features including N, P, K, temperature, humidity, pH, rainfall and Label. From the above matrix it is visible that, P and K contains highest positive correlation of 0.74, N depicts negative correlation. Significantly, temperature and humidity depict slight correlation of (0.21) and pH remains uncorrelated. Consecutively, rainfall also showcases slight correlations. From this correlation heatmap, selecting suitable crop type depending on the environmental factors is easy to understand and analyse.



Figure 11: Training and Testing Accuracy.

Source: Authors, (2026).

Figure 11 refers training and validation accuracy and loss, where the first graph showcases training and testing accuracy in which blue line depicts the training accuracy and red depicts the validation accuracy. The proposed model shows training accuracy of 97% and validation accuracy of 98% respectively. Whereas, the second graph shows the training and testing loss which is significantly low, indicating enhanced classification performance, enabling to select optimum crop suitable for specific environmental conditions.

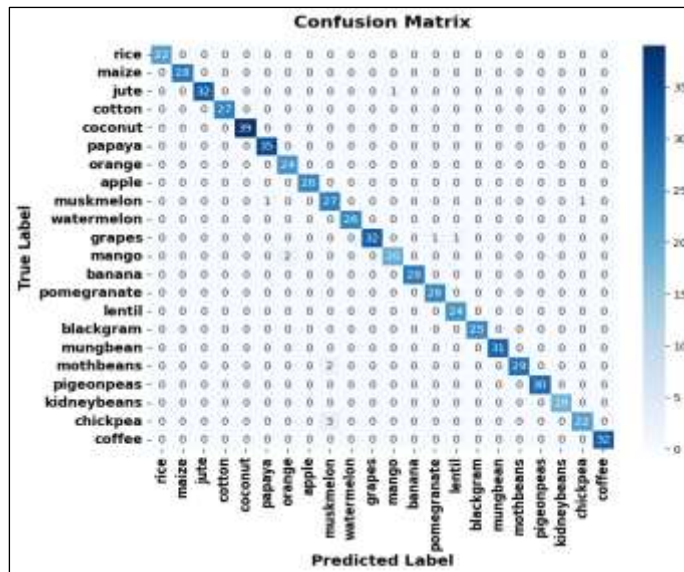


Figure 12: Confusion Matrix.

Source: Authors, (2026).

Figure 12 represents confusion matrix, where x-axis depicts the predicted labels and y-axis indicates the true labels, where coconut showcases the highest value, referring to accurate classification. Significantly, the matrix showcases that most of the listed crop are precisely classified indicating improved classification performance with better system performance.

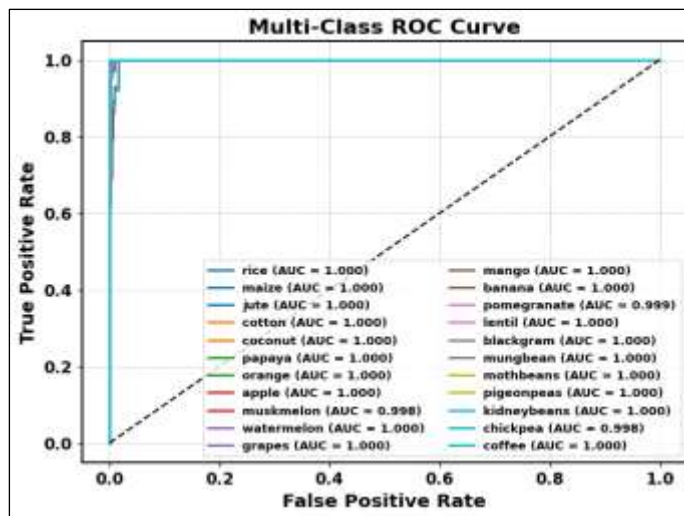


Figure 13: Multi-Class ROC Curve.

Source: Authors, (2026).

Figure 13 denotes the multi-class ROC-Curve, where x-axis depicts the False Positive Rate (FPR) and y-axis indicates the True Positive Rate (TPR), the ROC-Curve for each crop is at its highest, indicating enhanced classifier performance. In addition to this, AUC score depicted by all the other crops is 1.000 except for pomegranate which possesses slightly reduced AUC of 0.9999 respectively. Table 2 and 3 showcases the proposed model parameters and performance matrices, where the sample dataset is split into 70% and 30 % for training and testing. A proposed model attains highest accuracy of 98%, indicating highly efficient model functioning.

Table 2: Proposed Model Parameters.

Dataset	Crop Recommendation dataset
Samples	(2200,8)
Train Test Split	70%-30%
Training	1320
Testing	616
Validation	264

Source: Authors, (2026).

Table 3: Performance Matrices of the proposed model.

Accuracy	0.9805
Precision	0.9805
Recall	0.9803
F1-Score	0.9800
AUC-ROC	0.9998
Log Loss	0.0590
Cohen's Kappa	0.9796
Matthews Corrocoef	0.9796
Hamming Loss	0.0195
Jaccard Score	0.9625
Balanced Accuracy	0.9803

Source: Authors, (2026).

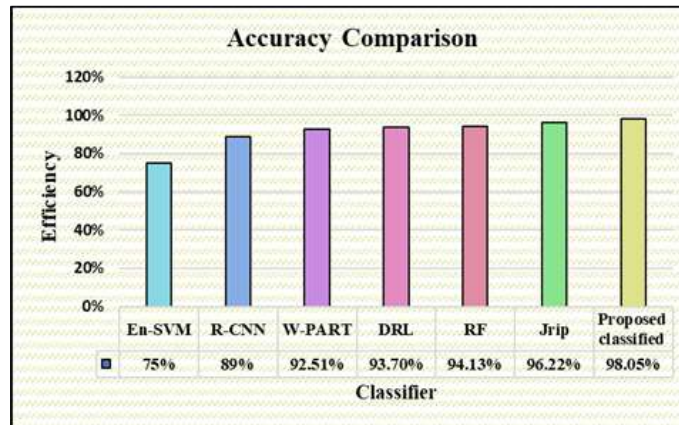


Figure 14: Accuracy Comparison.

Source: Authors, (2026).

Figure 14 represents the comparison accuracy between proposed classifier and other conventional classifiers, to determine the performance efficacy [25] of the proposed work. The above graph comprises of conventional classifiers with accuracy slightly lower such as En-SVM (75%) [26], R-CNN (89%) [27], W-PART (92.51%) [28], DRL (93.70%) [1], RF (94.13%) [29] and JRip (96.22%) [8] When compared to propose WaOA optimized MLP-Bi-GRU classifier which attained higher accuracy of 98.05%, indicating improved crop selection process.

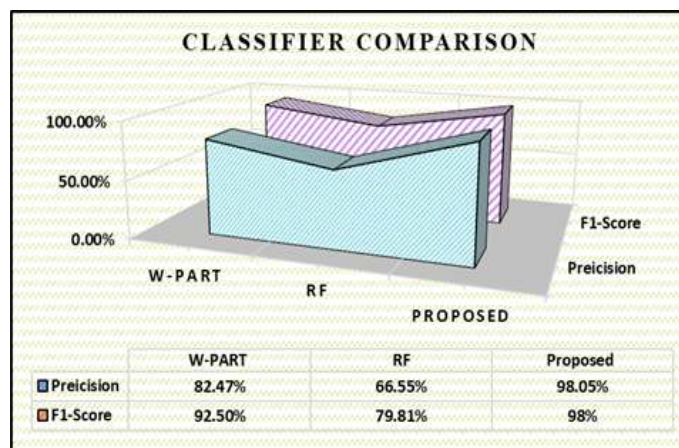


Figure 15: Classifier Comparison.

Source: Authors, (2026).

Figure 15 indicates the precision and F1-score comparison, here, it is evident that proposed classifier showcases higher precision (98.05%) and F1-Score (98%) than that of the other two conventional approaches, implying improved and enhanced classifier performance, which in turn improves the overall efficiency of the proposed system. Table 4 displays the error rate comparison between conventional MLP and proposed WaOA optimized MLP-Bi-GRU, which showcases that proposed classifier outdraws conventional MLP by acquiring reduced error rate respectively.

Table 4: Comparison of classifier error rate.

Classifier Type	MAE	RMSE	RAE	RRSE
MLP [11]	0.0082	0.0545	9.47	26.14
Proposed Classifier	0.0055	0.0402	6.85	19.75

Source: Authors, (2026).

IV. CONCLUSION

The developed CRS system utilizing the novel approaches offers highly efficient and improves system performance by optimizing crop selection on the basis of the agricultural and environmental conditions. The implementation of pre-processing stage where data cleaning and label encoding is performed enables improved data quality with enhanced conversion of numerical data into the form suitable for ML based classifiers. The deployment of EDA, offers improved visualization and analysis of the data patterns within the dataset. Moreover, the utilization of WaOA optimized MLP and Bi-GRU ensures, better extraction of complex and intricate patterns and sequential dependencies within the agricultural data. Furthermore, the Python software results validate that, the proposed model assures enhanced crop selection process with increased stability and reliability, contributing greater advancement in the farming field. Therefore, the development of CRS offers optimum solution with improved performance in overcoming the agricultural shortcomings, by ensuring optimum decision-making ability, improved efficiency (98%), Precision (98%), Recall (98%), F1 Score (98%) and achieving improved food productivity and quality.

V. AUTHOR'S CONTRIBUTION

Conceptualization: C. Pandi and M. Vedaraj.

Methodology: C. Pandi and M. Vedaraj.

Investigation: C. Pandi and M. Vedaraj.

Discussion of results: C. Pandi and M. Vedaraj.

Writing – Original Draft: C. Pandi.

Writing – Review and Editing: C. Pandi and M. Vedaraj.

Resources: M. Vedaraj.

Supervision: M. Vedaraj.

Approval of the final text: C. Pandi and M. Vedaraj.

VI. REFERENCES

- [1] D. Elavarasan, and P.D. Vincent, "Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications", IEEE access, vol. 8, pp. 86886-86901, May. 2020.
- [2] D. Gosai, C. Raval, R. Nayak, H. Jayswal, and A. Patel, "Crop recommendation system using machine learning, International Journal of Scientific Research in Computer Science", Engineering and Information Technology, vol. 7, no. 3, pp. 558-569, Jun. 2021.
- [3] J.P. Bharadiya, N.T. Tzenios, and M. Reddy, "Forecasting of crop yield using remote sensing data, agrarian factors and machine learning approaches", Journal of Engineering Research and Reports, vol. 24, no. 12, pp. 29-44, 2023.
- [4] S. Iniyan, and R. Jebakumar, "Mutual information feature selection (MIFS) based crop yield prediction on corn and soybean crops using multilayer stacked ensemble regression (MSER)", Wireless Personal Communications, vol. 126, no. 3, pp. 1935-1964, Oct. 2022.
- [5] P. Patil, V. Panpatil, and S. Kokate, "Crop prediction system using machine learning algorithms", Int. Res. J. Eng. Technol. (IRJET), vol. 7, no. 02, Feb. 2020.
- [6] S.K.S. Durai, and M.D. Shamili, "Smart farming using machine learning and deep learning techniques", Decision Analytics Journal, vol. 3, pp. 100041, Jun. 2022.
- [7] J. Pant, R.P. Pant, M.K. Singh, D.P. Singh, and H. Pant, "Analysis of agricultural crop yield prediction using statistical techniques of machine learning", Materials Today: Proceedings, vol. 46, pp. 10922-10926, Jan. 2021.
- [8] K. Bakthavatchalam, B. Karthik, V. Thiruvengadam, S. Muthal, D. Jose, K. Kotecha, and V. Varadarajan, "IoT framework for measurement and precision agriculture: predicting the crop using machine learning algorithms", Technologies, vol. 10, no. 1, pp. 13, Jan. 2022.
- [9] S. Rani, A.K. Mishra, A. Kataria, S. Mallik, and H. Qin, "Machine learning-based optimal crop selection system in smart agriculture", Scientific Reports, vol. 13, no. 1, pp. 15997, Sep. 2023.
- [10] F. Abbas, H. Afzaal, A.A. Farooque, and S. Tang, "Crop yield prediction through proximal sensing and machine learning algorithms", Agronomy, vol. 10, no. 7, pp. 1046, Jul. 2020.
- [11] E. Elbasi, C. Zaki, A.E. Topcu, W. Abdelbaki, A.I. Zreikat, E. Cina, A. Shdefat, and L. Saker, "Crop prediction model using machine learning algorithms", Applied Sciences, vol. 13, no. 16, pp. 9288, Aug. 2023.
- [12] M.K. Senapaty, A. Ray, and N. Padhy, "IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture", Computers, vol. 12, no. 3, pp. 61, Mar. 2023.
- [13] S.Á. Raja, B. Sawicka, Z. Stamenkovic, and G. Mariammal, "Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers", IEEE Access, vol. 10, pp. 23625-23641, Feb. 2022.
- [14] A. Ali, T. Hussain, N. Tantashutikun, N. Hussain, and G. Cocetta, "Application of smart techniques, internet of things and data mining for resource use efficient and sustainable crop production", Agriculture, vol. 13, no. 2, pp. 397, Feb. 2023.
- [15] S.S. Yamaç, and M. Todorovic, "Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data", Agricultural Water Management, vol. 228, pp. 105875, Feb. 2020.
- [16] S.K. Sharma, D.P. Sharma, and J.K. Verma, "Study on machine-learning algorithms in crop yield predictions specific to Indian agricultural contexts", In 2021 international conference on computational performance evaluation (ComPE), pp. 155-166, Dec. 2021.
- [17] H.M.B.P. Ranaweera, R.M.G.H.N. Rathnayake, and A.S.G.J.K. Ananda, "Crop Price Prediction Using Machine Learning Approaches", Reference to the Sri Lankan Vegetable Market. J. Manag. Matters, vol. 10, no. 1, pp. 19–34, 2023.

- [18] H. Singh, S. Kumar, and A. Arya, “Ethno-dermatological relevance of medicinal plants from the Indian Himalayan region and its implications on cosmeceuticals”, A review, *Journal of Drug Research in Ayurvedic Sciences*, vol. 8, no. 2, pp. 97-112, Apr. 2023.
- [19] N.N. Thilakarathne, M.S.A. Bakar, P.E. Abas, and H. Yassin, “A cloud enabled crop recommendation platform for machine learning driven precision farming”, *Sensors*, vol. 22, no. 16, pp. 6299, Aug. 2022.
- [20] B. Charbuty, and A. Abdulazeez, “Classification based on decision tree algorithm for machine learning”, *J.Appl.Sci.Technol.Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021.
- [21] A. Priyadharshini, S. Chakraborty, A. Kumar, and O.R. Pooniwalla, “Intelligent crop recommendation system using machine learning”, In *Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 843–848, Apr. 2021.
- [22] A. Shehadeh, O. Alshboul, R. E. Al Mamlook, and O. Hamedat, “Machine learning models for predicting the residual value of heavy construction equipment”, An evaluation of modified decision tree, LightGBM, and XGBoost regression, *Automation Construction*, vol. 129, pp. 103827, Sep. 2021.
- [23] M.Y. Shams, S.A. Gamel, and F.M. Talaat, “Enhancing crop recommendation systems with explainable artificial intelligence”, A study on agricultural decision-making, *Neural Comput. Appl*, vol. 36, no. 11, pp. 5695–5714, Apr. 2024.
- [24] J. Shook, T. Gangopadhyay, L. Wu, B. Ganapathysubramanian, S. Sarkar, and A.K. Singh, “Crop yield prediction integrating genotype and weather variables using deep learning”, *PLoS ONE*, vol. 16, no. 16, pp. e0252402, Jun. 2021.
- [25] M. Tabar, D. Lee, D.P. Hughes, and A. Yadav, “Mitigating Low Agricultural Productivity of Smallholder Farms in Africa: Time-Series Forecasting for Environmental Stressors”, In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 11, pp. 12608–12614, Jun. 2022.
- [26] G. Nagasubramanian, R. K. Sakthivel, R. Patan, M. Sankayya, M. Daneshmand and A. H. Gandomi, “Ensemble Classification and IoT-Based Pattern Recognition for Crop Disease Monitoring System”, in *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 12847-12854, Apr. 2021.
- [27] K. Osorio, A. Puerto, C. Pedraza, D. Jamaica, and L. Rodríguez, “A deep learning approach for weed detection in lettuce crops using multispectral images, *AgriEngineering*”, vol. 2, no. 3, pp. 471-488, Aug. 2020.
- [28] 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, pp. 773-797, Jan. 2021.
- [29] V. Pandith, H. Kour, S. Singh, J. Manhas, and V. Sharma, “Performance evaluation of machine learning techniques for mustard crop yield prediction from soil analysis”, *Journal of scientific research*, vol. 64, no. 2, pp. 394-398, 2020.