

## Research Article

# Crop Yield Maximization Using an IoT-Based Smart Decision

**Amna Ikram,<sup>1</sup> Waqar Aslam<sup>2</sup>,<sup>3</sup> Roza Hikmat Hama Aziz,<sup>3</sup> Fazal Noor<sup>4</sup>,  
Ghulam Ali Mallah<sup>5</sup>, Sunnia Ikram,<sup>2</sup> Muhammad Saeed Ahmad,<sup>1</sup>  
Ako Muhammad Abdullah<sup>6</sup>,<sup>7</sup> and Insaf Ullah<sup>7</sup>**

<sup>1</sup>Department of Computer Science & IT, Government Sadiq College Women University, Bahawalpur, Pakistan

<sup>2</sup>Department of Computer Science & IT, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

<sup>3</sup>Department of Computer Science, College of Basic Education, University of Sulaimani, Kurdistan Region, Iraq

<sup>4</sup>Department of Computer and Information Systems, Islamic University of Madinah, Madinah 400411, Saudi Arabia

<sup>5</sup>Department of Computer Science, Shah Abdul Latif University, Khairpur Mir's, Pakistan

<sup>6</sup>University of Sulaimani, College of Basic Education, Computer Science Department, Sulaimaniyah, Kurdistan Region, Iraq

<sup>7</sup>Hamdard Institute of Engineering & Technology, Islamabad 44000, Pakistan

Correspondence should be addressed to Waqar Aslam; waqar.aslam@iub.edu.pk and Insaf Ullah; insafktk@gmail.com

Received 2 March 2022; Revised 9 April 2022; Accepted 22 April 2022; Published 17 May 2022

Academic Editor: Waliullah Khan

Copyright © 2022 Amna Ikram et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Today, farmers are suffering from the low yield of crops. Though right crop selection is the main boosting key to maximize crop yield by doing soil analysis and considering metrological factors, the lack of knowledge about soil fertility and crop selection is the main reason for low crop production. In the changed current climate, the farmers having primitive knowledge about conventional farming are facing challenges about making sagacious decisions on crop selection. The selection of the same crop in every seasonal cycle makes the low soil fertility. This study is aimed at making an efficient and accurate system using IoT devices and machine learning (ML) algorithms that can correctly select a crop for maximal yield. Such a system is reliable as compared to the old laboratory testing manual systems, which bear the chances of human errors. Correct selection of a crop is predominantly a priority in agricultural arena. As a contribution, we propose an ML-based model, Smart Crop Selection (SCS), which is based on data of metrological and soil factors. These factors include nitrogen, phosphorus, potassium, CO<sub>2</sub>, pH, EC, temperature, humidity of soil, and rainfall. Existing IoT-based systems are not efficient as compared to our proposed model due to limited consideration of these factors. In the proposed model, real-time sensory data is sent to Firebase cloud for analysis. Its results are also visualized on the Android app. SCS ensembles the following five ML algorithms to increase performance and accuracy: Decision tree, SVM, KNN, Random Forest, and Gaussian Naïve Bayes. For rainfall prediction, a dataset containing historical data of the last fifteen years is acquired from Bahawalpur Agricultural Department. This dataset and an ML algorithm, Multiple Linear Regression leverages prediction of the rainfall in future, a much-desired information for the health of any crop. The Root Mean Square Error of the rain fall prediction model is 0.3%, which is quite promising. The SCS model is trained for 11 crops' prediction, while its accuracy is 97% to 98%.

## 1. Introduction

The agriculture sector is the lifeline of mankind and plays a vital role in economy. In conventional farming technique, crop selection was done by primitive knowledge of farmers. Mostly, farmers prefer to select the trendiest crop in their areas or the crop in their neighborhood. Due to a lack of scientific knowledge about farming and no rotation of crops,

fertility of lands is affected adversely. Major factors contributing to the crop quality are soil nutrients, ground water level, and type of fertilizer used. A traditional farmer faces recurrent challenges. Soil acidity may increase due to selection of wrong crops and inadequate soil nutrients [1, 2]. The unpredictable climate is the main factor for effecting crop's quality and yield. Soil fertility is an important factor for right crop selection and its health.

The motivation behind our research is to find problems that farmers face to get good and healthy crops. To overcome the mentioned issues of agriculture, we propose an ML algorithm-based model, Smart Crop Selection (SCS) that uses IoT. It is aimed at overcoming few current farming issues that arise due to inefficient approaches. SCS considers metrological factors such as temperature, humidity, rainfall, CO<sub>2</sub> level in air, soil pH, EC, and soil type, as depicted in Figure 1. The metrological factors directly affect the plant growth and production [3–5]. To check the soil fertility, soil analysis is performed. For soil analysis, the macro nutrients of soil, nitrogen, phosphorus, and potassium are considered. These three primary nutrients are very helpful for the well-being of plants and to prevent diseases. The soil pH shows the alkalinity of soil and regulates the soil nutrient availability for crops by controlling the chemical reactions and forms of nutrients. The higher and lower values of soil EC affect the plant growth. It also indicates the soil fertility, water quality, and salinity of soil. The CO<sub>2</sub> level in air plays an important role for crop health. It is used in photosynthesis process of plants. The proposed model is applicable for two types of soil: loamy and clay. These types of soil have good humidity and moisture level required for most of the crops. Rainfall is also an important factor for crop health [6, 7]. Each crop may have different water requirement. It is highly useful to know the average amount of rainfall of the season before sowing the crop. Its prediction is difficult but machine learning algorithms show promising results. Moving towards precision agricultural techniques allows growth rate of crop yield from 50% to 90%. Precision agriculture is a systematic way to make reasonable decisions and optimal utilization of resources [2]. By such an approach, soil fertility is maintainable.

Towards precision agriculture, IoT can be a key enabler. An IoT-based farming system can create effective decision-making and avoid undesirable situations. An automation system in smart agriculture is not very expensive but more precise than traditional farming system. The structure of IoT systems is based on three layers: perception, network, and application. On perception layer, physical devices like sensors, RFID tags, and cameras are used for data collection. Network layer is used for data communication and forwarding. Application layer is used to combine IoT with specific domain of usage [3].

ML is an exciting application of Artificial Intelligence. It provides the ability to learn by experiences without any explicit program [8]. The proposed model is based on simple and cost-effective hardware that can be used by agriculture officers and farmers to get good productivity of crops. SCS model is trained by classifying dataset and tested subsequently. The accuracy and performance of an ML classifier depend only on the type and size of the dataset [9]. Our dataset used for training the model has 2200 instances for 11 crops. For dataset classification, five supervised ML algorithms (DT, SVM, KNN, RF, and NB) are used. To overcome the weaknesses of individual ML algorithm, they are ensembled for improved accuracy. Our experimental results show 97% to 98% accuracy on real-time testing. In contrast to the previous studies, SCS model has the following novelty aspects:

- (i) In previous studies, limited parameters were used but in our proposed model, additional parameters related to soil and metrological factors are considered
- (ii) Existing smart agriculture systems are based on costly and complex sensors, whereas we use inexpensive sensors
- (iii) The laboratory system for soil analysis is time-consuming and expensive as compared to SCS. In existing research, soil samples are mostly collected manually and tested for fertility in laboratories. There is a risk of human error. For accurate results, an automated system should be developed, as done in our case
- (iv) SCS estimates crop yields in totality

The rest of the paper is structured as follows. In Section 2, we summarize related works. In Section 3, we describe the proposed system. In Section 4, we have presented and discussed the results. In Section 5, conclusions are drawn.

## 2. Related Works

Crop selection by real-time sensing data and soil analysis attributes is a big contribution in research of smart agriculture. Bhojwani et al. proposed a model based on three modules: crop selection, crop management, and crop maturity. They used parameters soil moisture, temperature, humidity, air pressure, and air quality with weather conditions for better crop selection and health monitoring. A real-time sensory data was used for analysis on ThingSpeak application with KNN algorithm [10]. Patil et al. proposed a scientific approach for crop selection by using various sensors of temperature, soil, humidity, and infrared with microcontroller for collecting real-time data. Some data mining techniques are applied for data preprocessing and comparing real-time data with trained data for crop prediction. They also considered crop prices for crop prediction, listed on National Commodity and Derivative Exchange. The KNN classifier is applied for data analysis [11].

Majumdar et al. have focused on IoT-oriented agricultural methods for weather monitoring. The prediction methods are investigated for commercial and scientific perspectives, cost of IoT components, security threats, and dependency of weather parameters on irrigation of crops [12]. Imran suggested a smart irrigation and crop selection system based on the parameters like temperature, humidity, light intensity, and moisture level of soils. Experiments were performed on five types of soils (loamy, black, laterite, alluvial, and silt soil). Experimental results show that soil's characteristics of different lands can be used for crop selection. ThingSpeak application is used for data analysis. An Android application is also designed to intimate the farmers about required water level of fields [13].

Rekha et al. proposed an IoT framework to improve farming methods for best use of land to increase crop production and profit maximization. A wireless sensor network

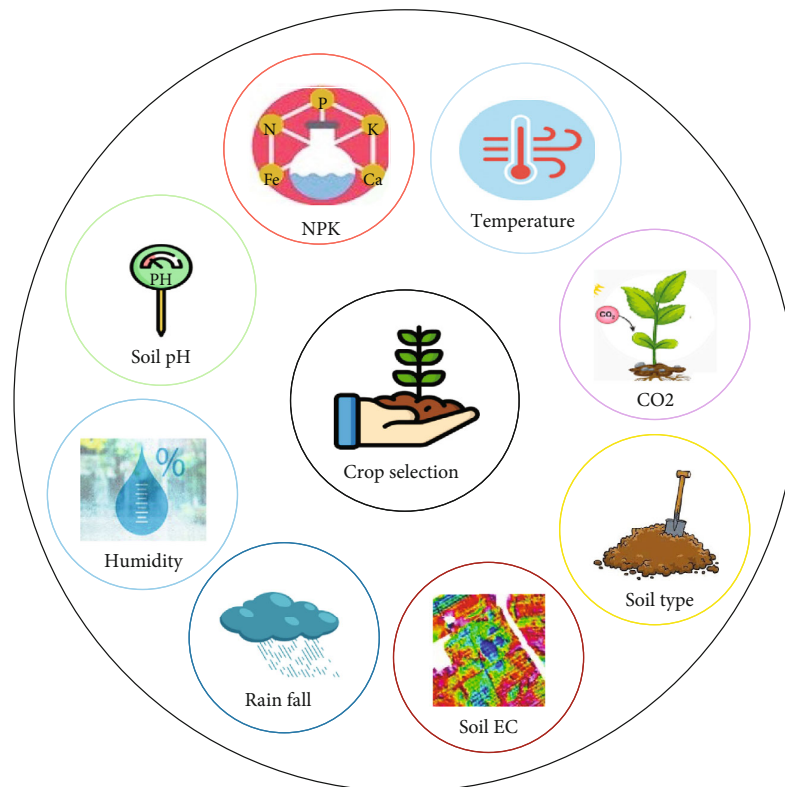


FIGURE 1: Real-time crop selection features.

TABLE 1: A brief comparison of SCS with contemporary works.

Citations	Objectives	Parameters	Methodology	Shortcomings
G Sai Pravallika et al., 2020 [31]	Crop selection	Temperature, moisture, humidity, and pH	No ML algorithm is applied. Simply compare the sensor data with the static data store in data base by SQL query to predict desirable crop.	Important parameters of soil fertility (NPK) are missing. An important factor rainfall is missing.
Bakthavatchalam et al., 2022 [21]	Crop prediction	Temperature, humidity, NPK, pH, and rainfall	Multilayer perceptron, JRip, Decision table	The performance metric results in Python are more reliable than in WEKA.
Waikar et al., 2020 [32]	Crop prediction	NPK, pH, and EC	SVM, Naïve Bayes, Artificial Neural Network, AdaBoost; Bagged Tree (Ensemble technique)	Results show less accuracy as compared to SCS. Important parameters, temperature, humidity, and rainfall, are missing
Jain and Ramesh, 2020 [33]	Crop selection	pH, NPK, and drainage capacity	RNN for weather prediction and Random Forest algorithm for crop selection	Random Forest is slow and inefficient for real-time predictions. Important parameters are missing.
A. Chlingaryan et al., 2020 [34]	Crop selection and fertilizer recommendation	Temperature, humidity, and water level	Linear Regression, Decision tree, K-Nearest Neighbors, and XGBoost	Limited parameters are considered for crop selection.

was deployed in the field to sense data for different parameters and for proper monitoring of field. The pH sensor was used to find the soil nutrients that helps to select the required fertilizer. An Android app was also developed to take proper farming decisions relating to irrigation with

the help of weather conditions [14]. Mulge et al. proposed a crop prediction method for crop yield maximization and quality of crops by considering real-time data of metrological factors using ML algorithms: precipitation, temperature, humidity, and solar light [15].

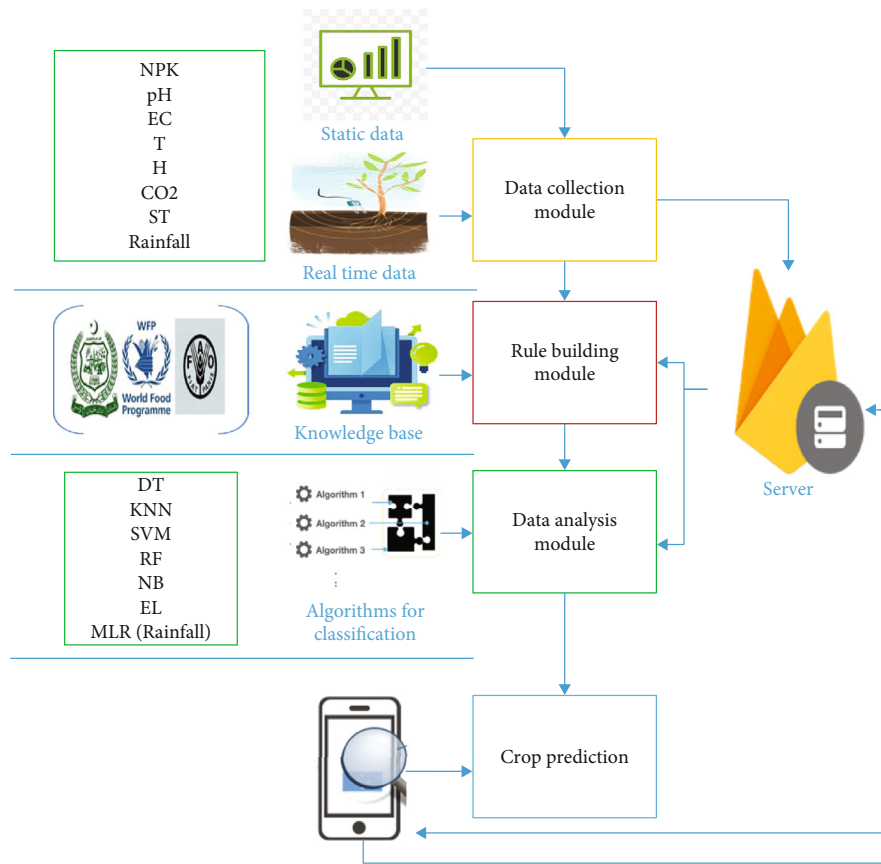


FIGURE 2: Process flow of SCS.

TABLE 2: Summary of the used symbols.

Symbol	Description
N	Nitrogen in soil
P	Phosphorous in soil
K	Potassium in soil
pH	Alkalinity of soil
T	Temperature of soil
H	Humidity of soil
EC	Electrical conductivity
DT	Decision tree
KNN	K-Nearest Neighbor
SVM	Support Vector Machine
RF	Random Forest
NB	Gaussian Naïve Bayes
MLR	Multiple Linear Regression
EL	Ensemble Learning

Nagasubramanian et al. proposed crop growth monitoring and disease detection by using cameras and IoT. Another work, ECPRC proposed an ensemble classification using SVM and CNN for data analysis and result prediction [16]. Paravalika et al. effectively predicted crops suitable to soil

type. The used parameters are temperature, humidity, moisture, and pH [17]. Ram and Kumar predicted best crop by considering metrological and soil factors and using Ensemble Learning on Decision tree and Linear Regression [18]. Colombo-Mendoza et al. presented a design of smart farming system using IoT sensors for data collection and ML algorithms. A new data mining approach is used to combine two types of datasets: climate data and crop production data for crop yield prediction [19]. Khongdet et al. proposed a model for smart crop tracking and monitoring by storing real-time data from IoT devices. SVM is used for crop disease detection. Fertilizer recommendation is also performed on the basis of previous land data [20]. Bakthavatchalam et al. presented a smart module to recommend a suitable crop for farming that can maximize crop yield. WEKA tool is used for data analysis by ML algorithms. A decision table classifier and multilayer perceptron rule-based classifier JRip are used for classification [21].

Gupta and Nahar also proposed a two-tier ML model for crop yield prediction. In the first tier, a classifier, Adaptive K-Nearest Centroid Neighbor (aKNCN) is proposed to analyze the soil quality and classify the input soil samples into different classes based on soil properties. In the second tier, Extreme Learning Machine algorithm model is used for crop yield prediction [22]. A crop is predicted by using performance comparisons of three ML algorithms; KNN, SVM,

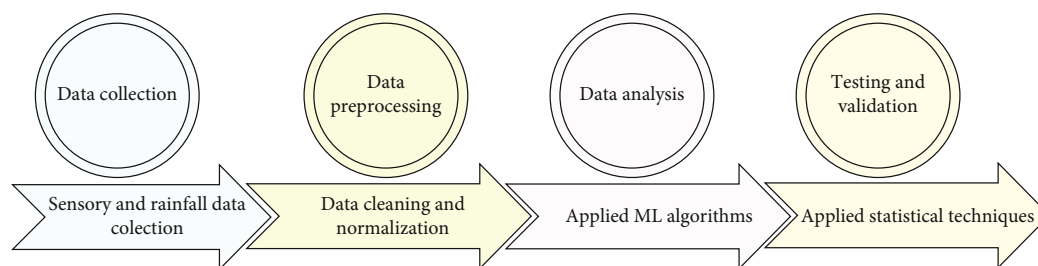


FIGURE 3: The architecture of SCS system.

1. Import libraries
2. Define variables and terms
3. Import datasets
4. Input rainfall data by user
5. Merge sensory dataset with rainfall prediction
6. Applied normalization
7. Applied split data
8. Applied classification algorithms
9. Evaluate results
10. Applied voting ensemble technique
11. Results output (selected crop)

ALGORITHM 1: Proposed algorithm for crop Selection Machine Learning (SCS) model.

TABLE 3: Sensors used in SCS.

Soil sensors	Specification	Description
Integrated soil sensor	Power supply	12 V-24 V
	Temperature measuring range	-45°C-115°C
	PH measuring range	3-9PH
	NPK measuring range	0-1999 mg/kg
	EC measuring range	0-10000 $\mu$ S/cm
DHT11	Power supply	3.5 V to 5.5 V
	Measuring range	20-95%RH
	Accuracy	$\pm 1^\circ\text{C}$ and $\pm 1\%$
MQ135	Power supply	5 V
	Power consumption	150 mA
	Detect/measure	NH3, Nox, CO2, alcohol, benzene, smoke

and Decision tree [23]. Another model is trained for suitable crop selection and monitoring conditions in field for disease detection and weather analysis. CNN is used for disease detection [24]. On the network layer of IoT system, security issues and information leakages are pointed out [25–27]. Paul et al. have proposed a model for yield prediction by selecting suitable crop for sowing. An external dataset is acquired for soil analysis. Different micro and macro nutrients of the soil are considered. Rapid minor application is used for applying ML classification algorithms (KNN and Naïve Bayes) for training the dataset [28]. There is also proposal for rainfall forecasting. A metrological data including monthly rainfall details is used for validation [29]. In this work, four parameters are used: temperature, humidity, pH, and rainfall for crop selection. It also guides about the

number of nutrients required for a particular crop. For dataset classification, Decision tree is applied and for rainfall prediction SVM is applied [30]. A comparison of existing research with SCS model is given in Table 1.

### 3. Proposed Model

The proposed system is based on real-time sensing of the soil parameters by sensors and the rain fall prediction on the basis of external dataset. The real-time data is saved in database on cloud, and ML algorithms are applied for further analysis and prediction as given in Figure 2. An Ensemble Learning (EL) technique is applied on some distinct ML algorithms, i.e., Decision tree, Naïve Bayes, Support Vector Machine, K-Nearest Neighbor, and Random Forest. For

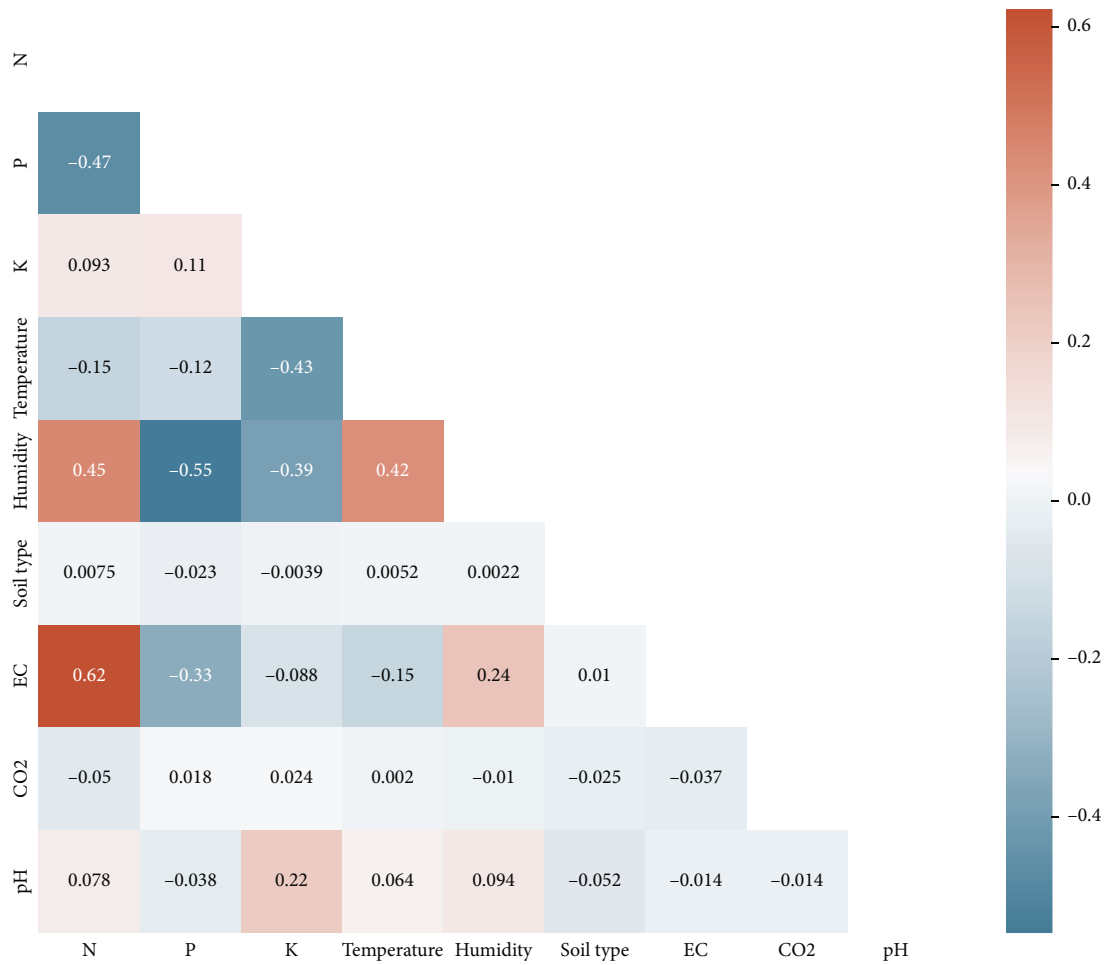


FIGURE 4: The correlation matrix of the data used.

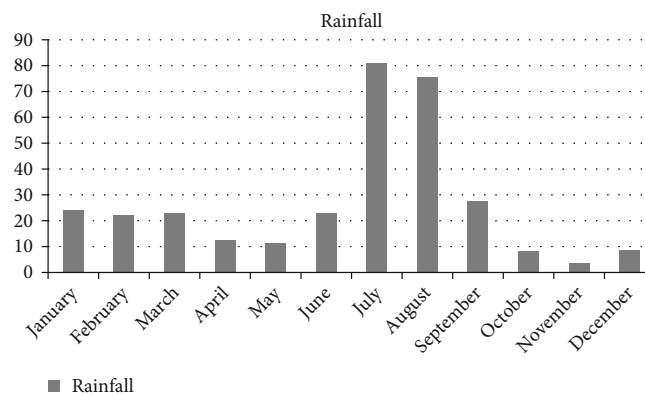


FIGURE 5: Average rainfall in Punjab.

rainfall prediction, a ML algorithm Multiple Linear Regression model is used. Python is used for implementation, as it is flexible. The proposed solution is based on real-time data sensing of major soil nutrient parameters such as

NPK (nitrogen, phosphorus, and potassium) and other factors (T, H, pH, EC, CO<sub>2</sub>, soil type, and rainfall). In SCS, five ML algorithms are used due to their high performance and accuracy for results. We have shown the description of the

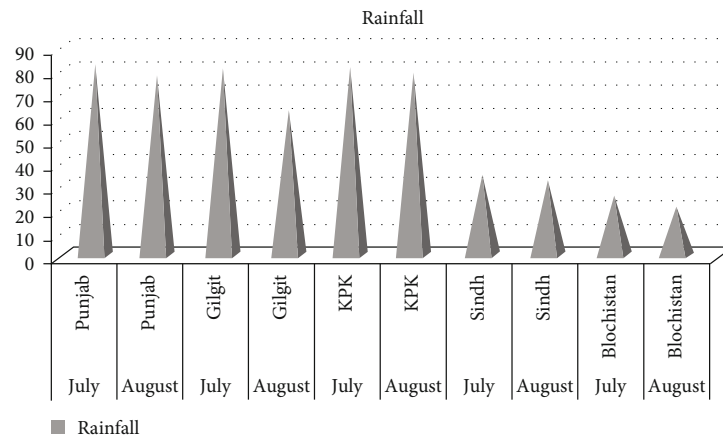


FIGURE 6: Average rainfall in different regions.

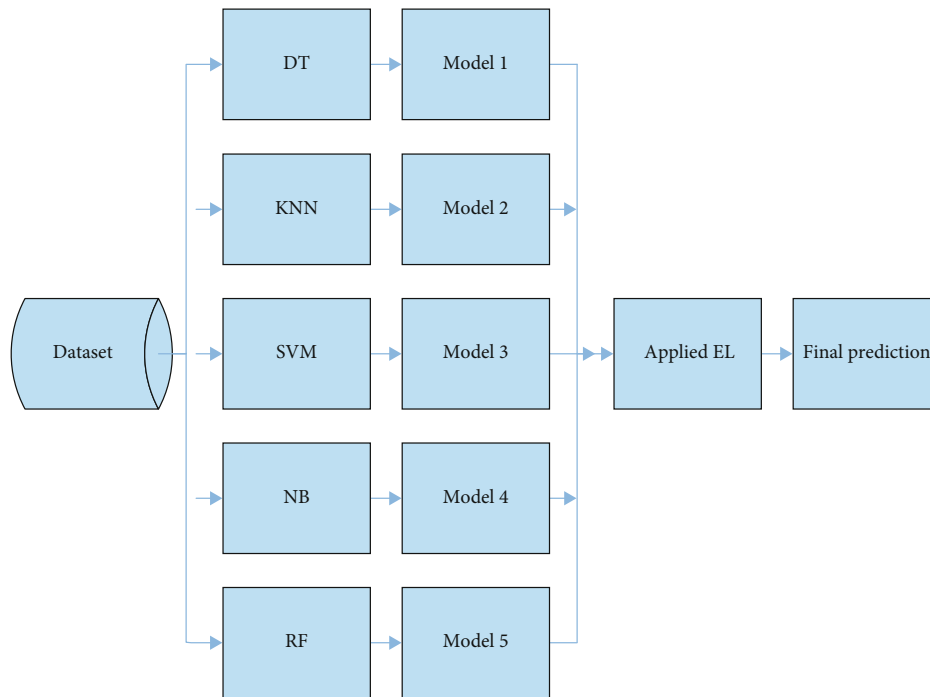


FIGURE 7: Layout used for the Ensemble technique.

used symbols in Table 2. The process flow of our proposed model comprises of different phases, which are discussed in Figure 3.

**Data collection:** two dataset are used in proposed model. One dataset is used for training the model, and the second one is used for testing and validation of the model. Real-time dataset is collected for NPK, pH, EC, T, and H by connecting NPK, pH, and EC sensors with Arduino microcontroller. The rainfall data of last 15 years is obtained from government website <https://bakhabarkissan.com/>. Experiments are performed on two types of soils: loamy and clay.

**Data preprocess:** the real-time sensory data is usually in raw form. The data mining techniques are applied for data preprocessing. As real-time data is coming from different sensors so there could be chance of errors. For rainfall data, the same data mining techniques are applied as for sensory data. The preprocessing techniques are applied on dataset are given next.

**Filling missing entries (cleaning) and feature scaling (normalization):** in rainfall dataset, the user input is string based so data transformation technique is applied to convert string data into numeric form. **Data analysis:** after



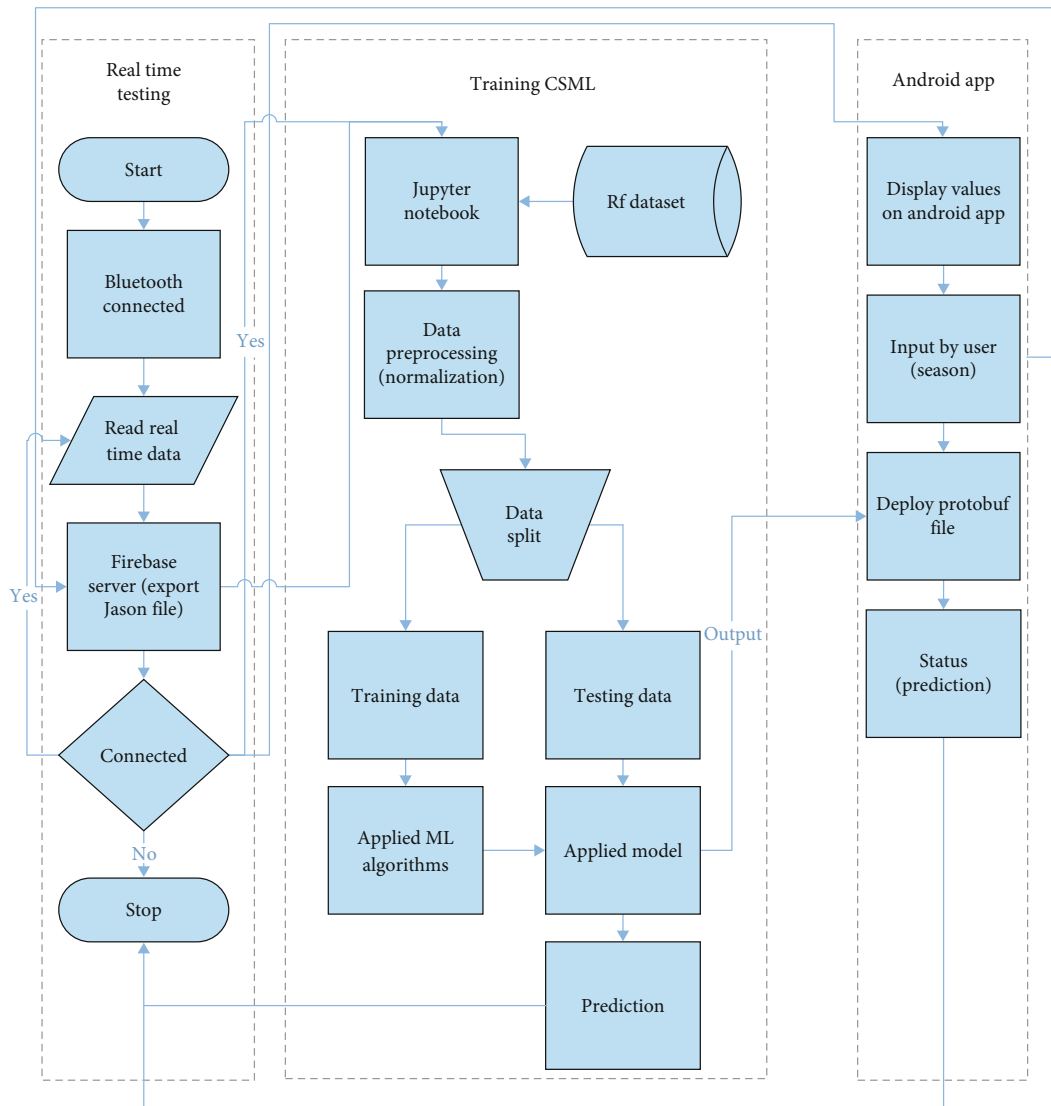


FIGURE 8: Workflow of SCS.

preprocessing the data, decision rules are applied on dataset. In decision rules, standard ranges of parameters are defined for each crop. ML algorithms are used for dataset training. The performance of each ML algorithms is compared and then applied voting Ensemble technique on applied algorithms for better performance and good accuracy. For testing and validation of the system, a real-time sensory dataset is used. For rain fall and soil type attributes, a user input acquired by Android app. For crop prediction, the results are displayed on an Android application. This is very efficient way to assist farmers. The proposed SCS model is described in Algorithm 1.

The architecture of SCS is shown in Figure 3. The SCS is based on four modules. In sensing module, real-time data is collected by sensors and send to Firebase cloud for storage. The external rainfall dataset is also used to get the predicted rainfall value on the basis of historical data. In the second

module, working rules are applied to define the standard ranges for data classification. In the third module, ML algorithms are applied for classification and prediction of output. On the last step, final output is displayed on Android application.

We have considered various features for our proposed SCS including NPK, pH, temperature, humidity, EC, soil type, CO<sub>2</sub>, and rainfall. In proposed SCS, a 7 in 1 soil integrated sensor is used to perform data collection for major soil features like NPK, pH, T, M, and EC. To check CO<sub>2</sub> level in air, MQ135 sensor is used. To calculate temperature and humidity, DHT11 sensor is used. For Bluetooth connectivity, HC-05 module is used. By integrating all these devices with Arduino Nano, a real-time data is send to cloud server [4]. The specifications of sensors are given in Table 3.

For prediction of right crop by our SCS model, the real-time dataset is collected by sensors. The dataset consists of



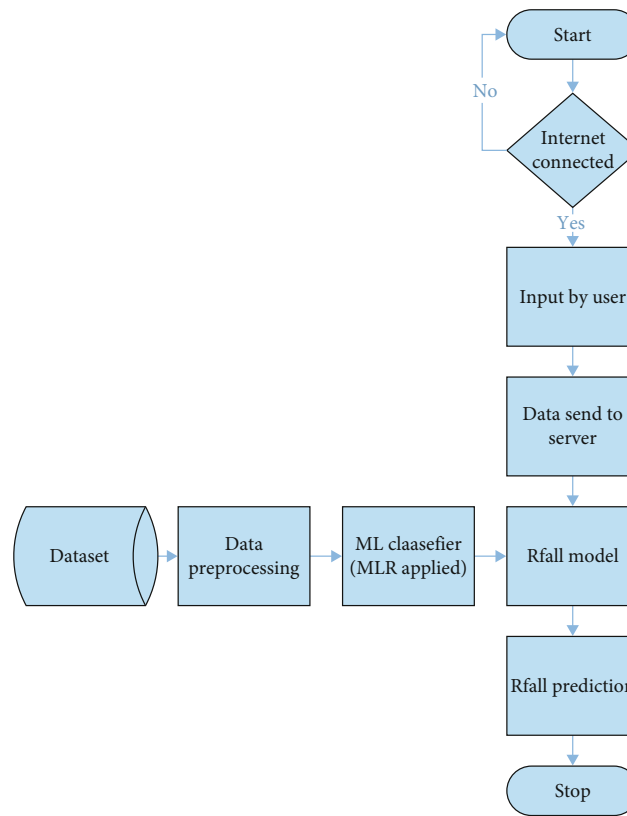


FIGURE 9: Workflow of the rainfall model.

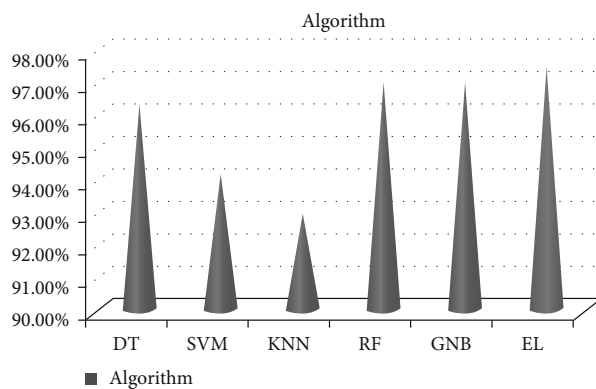


FIGURE 10: Accuracy comparisons of the ML algorithms with SCS.

ten features that have primary role in yield maximization. Each crop has different requirement of nutrients, temperature, humidity, and water. To train SCS model, we have selected 11 crops that are commonly used in different regions of Pakistan. For training the model, 2200 observations are used. The crop feature is used as a dependent variable in our model.

In Figure 4, correlation matrix of training dataset is given. A heat map is a good way of representing data visually as it is difficult to show large data into tabular form. In cor-

relation matrix, the density and intensity of the variables are shown. Nitrogen and EC are showing higher correlation. All data is symmetrical from left to bottom. The dark red shows higher values, and dark blue shows low values. For heat map implementation, Python libraries are used. Rainfall data is collected from the local agricultural department of Bahawalpur region and government website of agriculture department. The collected data was in raw form. This data is based on the average rainfall of each month of each season of the last 15 years. It has 300 observations. The average rainfall in Punjab is shown in Figure 5. The average rainfall in the month of July and August is higher than remaining months in Punjab. In Figure 6, we also have juxtaposed the rainfall data of different regions of the country. The rainfall rate is higher in Punjab and KPK in the month of July and August than other provinces. These regions are more fertile than other regions of Pakistan. The soil fertility is also an important factor for crop maximization but it mainly depends on availability of water resources.

**3.1. Soil Sampling.** Soil samples are randomly chosen on the basis of availability. The soil samples were taken 4 to 6 inches below the earth surface. The loamy soil is considered best for sowing due to its structure having combination of soil particles of different soils. The clay soil has high capacity to absorb water. It is cold in winter and wet in summer. The

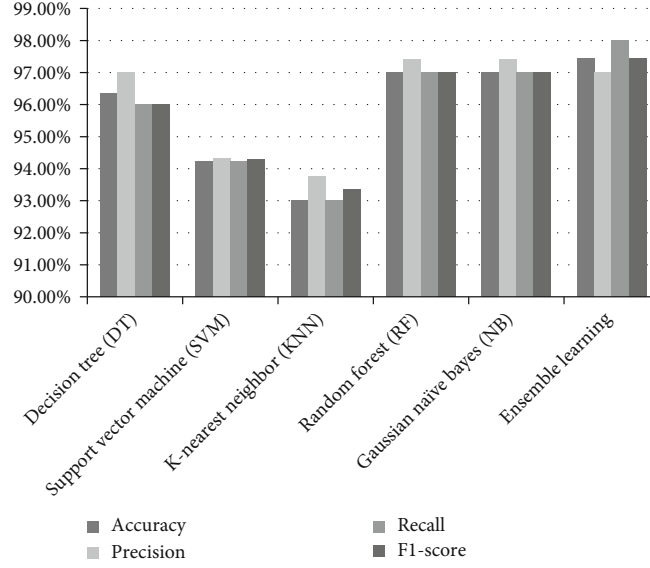


FIGURE 11: Statistical analysis of the ML algorithms with SCS.

purpose of getting soil samples of different soils is to check if the soil has adequate nutrients.

**3.2. Used ML Algorithms.** We used various ML algorithms including Decision tree (DT), Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forest (RF), and Ensemble Learning (EL) technique for crop prediction and Multiple Linear Regression (MLR) for rainfall forecasting. In the proposed SCS, these ML algorithms are used as the base models for dataset given in Figure 7.

**3.3. Decision Rules.** In our proposed SCS, decision rules are applied for data analysis and classification by ML algorithms to train the model. The ranges for each feature are defined to classify crops. We have designed working rules by getting the standard values of each factor for each crop. The first module of SCS is designed by connecting sensors with Arduino Nano. Every sensor is connected with separate Arduino. The real-time values are displayed on Android app by Bluetooth connectivity (HC-05 module) and sent to Firebase cloud database. The rainfall values are acquired from external dataset. The real-time data collected by sensors usually have some noise, missing values, and errors that can affect the decision by ML classifier algorithms. For accurate results and good accuracy, data preprocessing techniques should be applied. To remove missing entries, we have filled the fields with null entries and with the most frequent values. As all the features have different units to measure the magnitudes, so there is a need for data scaling. Two types of data scaling techniques used are normalization and standardization [35]. In SCS, normalization (Min-Max) technique is applied for data scaling. The algorithms using Euclidean distance (KNN and SVM) for data classification require data scaling. By this technique, all the feature data is converted into (0-1) range from the original range. The Min-MaxScaler method

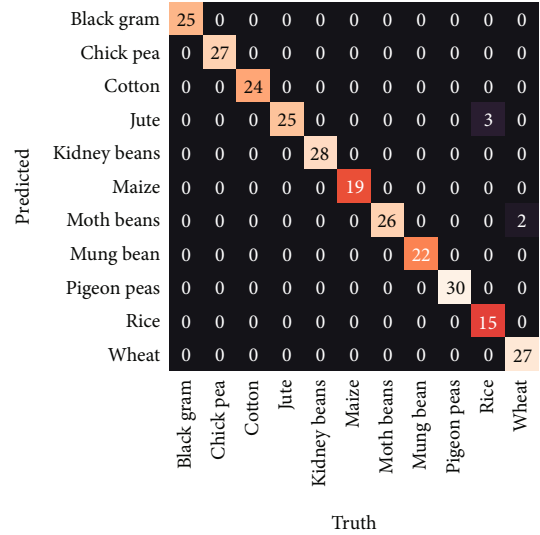


FIGURE 12: Confusion matrix of SCS.

is implemented from Scikit-learn library of Python [34, 36]. The normalized data is stored in variable  $x_s$  as follows:

$$x_s = \frac{(x - x_{\min})}{x_{\max} - x_{\min}}, \quad (1)$$

where  $x_s$  denotes scaled value of  $X$  and  $x_{\min}$  and  $x_{\max}$  refer to minimum and maximum values of  $X$ , respectively.

An Ensemble technique is applied on ML algorithms, which are used as base learners. The final result is based on the prediction of these algorithms. All applied base learner algorithms are appended together in an array.

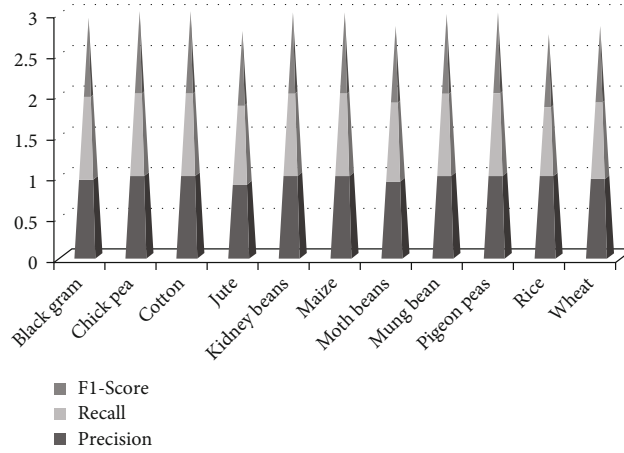


FIGURE 13: Performance metric of the classes.

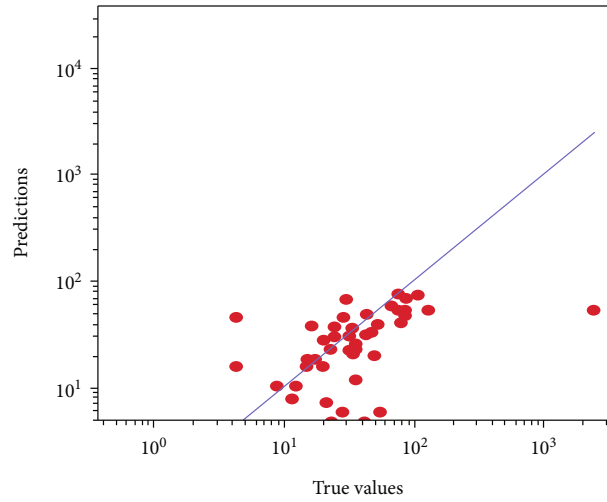


FIGURE 14: Rainfall: actual data vs. predicted data.

**3.4. Workflow.** In Figure 8, the whole workflow of model of SCS is shown. The data acquisition is performed by SCS hardware module by collecting real-time values of NPK, pH, EC and T, H, and CO<sub>2</sub>. SCS database is created in Firebase. When Bluetooth is connected with Android app, then real-time data is collected by sensors and sent to Firebase database. The real-time database exported a JSON file which we have converted into xsl and import into android studio. To train the SCS model, all working is performed in Jupyter Notebook using Python. An Android app is created to show SCS prediction values. A Protobuf file is exported to Android studio.

The workflow model for rainfall prediction is shown in Figure 9. ML algorithm MLR (Multiple Regression Model) [37, 38] is applied on rainfall data for the prediction of future values of rainfall. The dataset is based on average monthly rainfall of last 15 years. The parameters of dataset are year, month, season, and previous rain fall data. The calculated Root Mean Square Error is 0.3, and  $r^2$  is 0.610132.

The  $r^2$  is used to show the percentage of variance of dependent variable to independent variable.

#### 4. Performance Evaluations

By applying EL, we have improved the performance of our model by increasing the accuracy as compared to individual models. The accuracies are calculated by dividing all the true predictions calculated by algorithm with total number of dataset values.

$$\text{Accuracy} = \frac{\text{Tp} + \text{Tn}}{\text{Tp} + \text{Fp} + \text{Tn} + \text{Fn}}, \quad (2)$$

where Tp is the true positive, Tn is the true negative, Fp is the false positive, and Fn is the false negative. The classification error rate is calculated as follows:

$$\text{Error rate} = (1 - \text{accuracy}) \times 100. \quad (3)$$

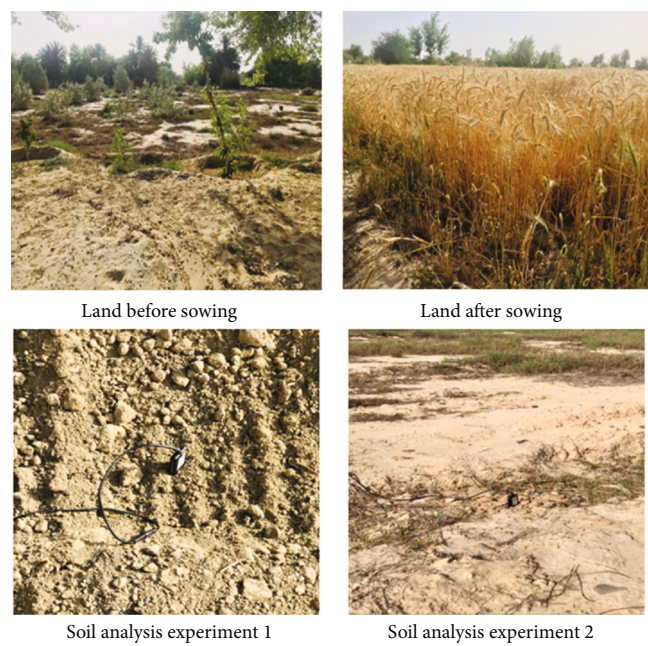


FIGURE 15: Agronomic measurement for SCS.



FIGURE 16: Android application for SCS.

TABLE 4: An accuracy comparison with existing approaches.

Citations	Year	Dataset size	Features selected	Approach	Accuracy
S. Parween et al. [39]	2021	Not mentioned	5	Naïve Bayes	96.00%
M. H. Kishan Das et al. [24]	2021	300	4	Decision tree	97.00%
R. Kumar and V. Singhal [40]	2022	298	3	XGBoost	92%
SCS (proposed model)	2022	2200	10	Ensemble Learning (DT, NB, SVM, KNN, and RF)	97.45%

The precision, recall, and F1-score are calculated to check the reliability of our model. Precision is calculated to show the accurate predictions in positive classes.

$$\text{Precision} = \frac{Tp}{Tp + Fp} \quad (4)$$

Recall is calculated to predict how many positive cases our model can predict.

$$\text{Recall} = \frac{Tp}{Tp + Fn} \quad (5)$$

For calculating F1-score, both precision and recall are combined,

$$\text{F1-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

We have compared the performances of ML algorithms. In Figure 10, we observe that performance of RF and NB algorithms has higher classification accuracy as compared to other algorithms KNN, DT, and SVM. This is because in Random Forest, multiple DTs are involved that make the decision boundary more stable and accurate. SVM is more suitable for binary classification.

In Figure 11, we have analyzed the accuracy, precision, recall, and F1-score of implemented algorithms. We observe RF, NB, and EL algorithms DT, SVM, and KNN in terms of accuracy, precision, recall, and F1-score in SCS system. To check class wise performance of SCS, a confusion metrics is given in Figure 12.

Confusion matrix is showing count of truly predicted instances and wrongly predicted instances of each class. By Equation (2), the accuracy of the whole model is calculated as

$$\text{Accuracy} = \frac{\text{sum of the truly predicted values}}{\text{sum of the total values}} = \frac{268}{275} = .9745. \quad (7)$$

The performance metrics of one output label class “black gram” is calculated by getting values from confusion matrix. By setting values  $Tp = 25$ ,  $Tn = 250$ ,  $Fp = 0$ , and  $Fn = 1$  and applying formulas in Equations (2), (4), (5), and (6), the following values of performance metrics are determined: accuracy = 100, precision = 96%, recall = 100, and F1-score = 98%.

Similarly, values of other classes are given in Figure 13.

The rainfall predicted data is given in Figure 14. We see that the values closer to diagonal line show correct prediction. We observe that most of our predicted values are closer to diagonal line. The diagonal line shows the actual values of rainfall data. The cross-validation technique is also performed for validity of the results. For k-fold, 5 splits are set to divide the data, and accuracy of each fold of model is displayed and validated.

Experiments are performed on two types of soil samples: loamy and clay. SCS recommended “wheat” crop for loamy soil and “lentils” crop for clay soil. The performance of SCS was also measured by direct observation of wheat crop using agronomic measurement given in Figure 15. The experiment was performed on 1 acre land of Bahawalpur region. By using SCS system, wheat crop was selected. The NPK value is maintained by adding required amount of fertilizer in the soil for yield maximization. A good yield and healthy crop are observed in the control field.

An interface is also designed on Android application to show the real-time sensory data and for input of rainfall data and soil type from user. Final crop prediction is displayed on the Android app. By connecting Bluetooth, the real-time values are received and sent to Firebase cloud for analysis. The decision support system displays final output of crop prediction on Android application given in Figure 16.

**4.1. Comparison with Existing Approaches.** In precision agriculture, a lot of work has been performed on crop selection using different features and dataset sizes. An accuracy comparison of our proposed SCS model with few representative works [24, 39, 40] is shown in Table 4.

**4.2. Limitations.** SCS seeks crop yield maximization through selection of the correct crop. Some limitations of proposed model are as follows:

- (i) EL is computationally expensive
- (ii) The performance of SCS varies with size of the dataset
- (iii) Air pressure and light intensity can be considered as influential factors
- (iv) The system is limited for few crops only

## 5. Conclusions

Farmers using traditional methods in agriculture face problems such as low crop yield due to unpredicted weather, wrong amount of water and nutrients, and wrong selection of crop. In previous research work, limited parameters were



used that are insufficient for high yield of crops. Our research is aimed at maximizing the crop yield by selecting suitable crop. We tackle this issue by applying technology methodically and evidence-based analysis. For instance, adding required amount of nutrients gives improved yields. Our work is based on selection of the influential parameters. The ML algorithms used in our proposed research give improved accuracy with less computational cost as compared to previous research. To facilitate farmers, an Android app is developed. The cost of our system is very low, and all the used sensors are easily available and easy to use.

In future, more parameters and crops can be added to this system. The more accurate and efficient ML algorithms like CNN and LSTM can also be studied. SCS model can be integrated with security to protect crop data. For crop monitoring, drone cameras can also be used. Fertilizer recommendation system can also be developed on the basis of real-time sensory data of soil nutrients.

## Data Availability

All data generated or analyzed during this study are included in this published article.

## Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] P. Kanchan and N. Shardoor, "Krashignyan: a farmer support system," *Asian Journal For Convergence In Technology (AJCT)* ISSN-2350-1146, vol. 7, no. 3, pp. 1–7, 2021.
- [2] F. Javed, M. K. Afzal, M. Sharif, and B. S. Kim, "Internet of Things (IoT) operating systems support, networking technologies, applications, and challenges: a comparative review," *IEEE Communication Surveys and Tutorials*, vol. 20, no. 3, pp. 2062–2100, 2018.
- [3] H. A. Issad, R. Aoudjit, and J. J. Rodrigues, "A comprehensive review of data mining techniques in smart agriculture," *Engineering in Agriculture, Environment and Food*, vol. 12, no. 4, pp. 511–525, 2019.
- [4] J. Flak, "Technologies for sustainable biomass supply-overview of market offering," *Agronomy*, vol. 10, no. 6, 2020.
- [5] P. M. Kavita, "Satellite-based crop yield prediction using machine learning algorithm," in *2021 Asian Conference on Innovation in Technology (ASIANCON)*, pp. 1466–1470, PUNE, India, August 2021.
- [6] W. Mupangwa, R. Makanza, L. Chipindu et al., "Temporal rainfall trend analysis in different agro-ecological regions of southern Africa," *Water SA*, vol. 47, no. 4, pp. 466–479, 2021.
- [7] A. G. Djibo, H. Karambiri, O. Seidou et al., "Linear and non-linear approaches for statistical seasonal rainfall forecast in the Sirba watershed region (SAHEL)," *Climate*, vol. 3, no. 3, pp. 727–752, 2015.
- [8] N. Jain, A. Kumar, S. Garud, V. Pradhan, and P. Kulkarni, "Crop selection method based on various environmental factors using machine learning," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 2, pp. 1530–1533, 2017.
- [9] D. S. Wankhede, "Analysis and prediction of soil nutrients pH, N, P, K for crop using machine learning classifier: a review," in *International Conference on Mobile Computing and Sustainable Informatics*, pp. 111–121, Springer, Cham, 2020, January.
- [10] Y. Bhojwani, R. Singh, R. Reddy, and B. Perumal, "Crop selection and IoT based monitoring system for precision agriculture," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 2, 2017.
- [11] J. S. Raj, *International Conference on Mobile Computing and Sustainable Informatics*, EAI/Springer Innovations in Communication and Computing, Springer Cham, 2020.
- [12] P. Majumdar, S. Mitra, and D. Bhattacharya, "IoT for promoting agriculture 4.0: a review from the perspective of weather monitoring, yield prediction, security of WSN protocols, and hardware cost analysis," *Journal of Biosystems Engineering*, vol. 46, no. 4, pp. 440–461, 2021.
- [13] S. Imran, "Effective crop selection and conservative irrigation using IoT," *International Journal of Science and Research (IJSR)*, 2016.
- [14] P. Rekha, V. P. Rangan, M. V. Ramesh, and K. V. Nibi, "High yield groundnut agronomy: an IoT based precision farming framework," in *2017 IEEE Global Humanitarian Technology Conference (GHTC)*, vol. 2017, pp. 1–5, San Jose, CA, USA, Dec. 2017.
- [15] M. Mulge, M. Sharnappa, A. Sultanpure, D. Sajjan, and M. Kamani, "An invitation to subscribe: the international journal of analytical and experimental modal analysis," vol. 10, no. 1, pp. 1112–1117, 2020.
- [16] 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," *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 12847–12854, 2021.
- [17] G. S. Pravalika, L. Kundana, K. S. Thanvi, G. Sirisha, and C. Rupa, "Proficient smart soil based IoT system for crop prediction," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 752–757, July 2020.
- [18] A. Ram and R. Kumar, *Prediction of the crop cultivating using resembling and IoT techniques in agricultural fields for increasing productivity 1*, European Journal of Molecular & Clinical Medicine, 2020.
- [19] L. O. Colombo-Mendoza, M. A. Paredes-Valverde, M. D. P. Salas-Zarate, and R. Valencia-Garcia, "Internet of Things-driven data mining for smart crop production prediction in the peasant farming domain," *Applied Sciences*, vol. 12, no. 4, 2022.
- [20] K. Phasinam, T. Kassanuk, and M. Shabaz, "Applicability of Internet of Things in Smart Farming," *Journal of Food Quality*, Article ID 7692922, p. 7, 2022.
- [21] K. Bakthavatchalam, B. Karthik, V. Thiruvengadam et al., "IoT framework for measurement and precision agriculture: predicting the crop using machine learning algorithms," *Technologies*, vol. 10, no. 1, p. 13, 2022.
- [22] A. Gupta and P. Nahar, "Classification and yield prediction in smart agriculture system using IoT," *Journal of Ambient Intelligence and Humanized Computing*, 2022.
- [23] A. Gupta, D. Nagda, P. Nikhare, and A. Sandbhor, "Smart crop prediction using IoT and machine learning," *International Journal of Engineering Research & Technology (IJERT)*, pp. 18–21, 2021.

- [24] M. H. Kishan Das, D. Mishra, and D. Deepa, "Automation and integration of growth monitoring in plants (with disease prediction) and crop prediction," *Materials Today: Proceedings*, vol. 43, no. 1, pp. 3922–3927, 2021.
- [25] W. U. Khan, X. Li, A. Ihsan, M. A. Khan, V. G. Menon, and M. Ahmed, "2021. NOMA-enabled optimization framework for next-generation small-cell IoV networks under imperfect SIC decoding," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2021.
- [26] W. U. Khan, Z. Ali, M. Waqas, and G. A. S. Sidhu, "Efficient power allocation with individual QoS guarantees in future small-cell networks," *AEU-International Journal of Electronics and Communications*, vol. 105, pp. 36–41, 2019.
- [27] W. U. Khan, J. Liu, F. Jameel, M. T. R. Khan, S. H. Ahmed, and R. Jantti, "Secure backscatter communications in multi-cell NOMA Networks: enabling link security for massive IoT networks," in *In IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, vol. 2020, pp. 213–218, July 2020.
- [28] M. Paul, S. K. Vishwakarma, and A. Verma, "Analysis of soil behaviour and prediction of crop yield using data mining approach," in *Proceedings -2015 International Conference on Computational Intelligence and Communication Networks, CICN*, vol. 2016, pp. 766–771, Jabalpur, India, 2015, August.
- [29] N. Gnanasankaran and E. Ramaraj, "A multiple linear regression model to predict rainfall using Indian meteorological data," *International Journal of Advanced Science and Technology*, vol. 29, no. 8, pp. 746–758, 2020.
- [30] M. Mulge, M. Sharnappa, A. Sultanpure, D. Sajjan, and M. Kamani, "Agricultural crop recommendation system using IoT and M.L," *The International Journal of Analytical and Experimental Modal Analysis*, vol. 12, no. 6, 2020.
- [31] G. S. Pravallika and G. Sirisha, "Proficient smart soil based IoT system for crop prediction," *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 752–757, 2020.
- [32] V. C. Waikar, S. Y. Thorat, A. A. Ghute, P. P. Rajput, and M. S. Shinde, "Crop prediction based on soil classification using machine learning with classifier ensembling," *Int. Res. J. Eng. Technol.*, vol. 7, no. 5, 2020.
- [33] S. Jain and D. Ramesh, "Machine learning convergence for weather based crop selection," in *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science*, vol. 2020, pp. 1–6, SCEECS, 2020, Feb.
- [34] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review," *Computers and Electronics in Agriculture*, vol. 151, pp. 61–69, 2018.
- [35] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*, Elsevier, 2011.
- [36] K. Palanivel and C. Surianarayanan, "An approach for prediction of crop yield using machine learning and big data techniques," *International Journal of Computer Engineering and Technology*, vol. 10, no. 3, pp. 110–118, 2019.
- [37] L. A. Yousif, A. A. Khatir, F. M. El-Hag et al., "Rainfall variability and its implications for agricultural production in Gedarf State, Eastern Sudan," *African Journal of Agricultural Research*, vol. 13, no. 31, pp. 1577–1590, 2018.
- [38] W. M. Ridwan, M. Sapitang, A. Aziz, K. F. Kushiar, A. N. Ahmed, and A. El-Shafie, "Rainfall forecasting model using machine learning methods: case study Terengganu, Malaysia," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1651–1663, 2021.
- [39] S. Parween, A. Pal, I. Snigdh, and V. Kumar, "An IoT and machine learning-based crop prediction system for precision agriculture," in *In Emerging Technologies for Smart Cities*, pp. 9–16, Springer, Singapore, 2021.
- [40] R. Kumar and V. Singhal, "IoT enabled crop prediction and irrigation automation system using machine learning," *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, vol. 15, no. 1, pp. 88–97, 2022.