Multimedia Tools and Applications
https://doi.org/10.1007/s11042-020-09740-6

An efficient IoT based smart farming system using machine learning algorithms

Nermeen Gamal Rezk[1] **· Ezz El-Din Hemdan**[2] **· Abdel-Fattah Attia**[1] **· Ayman El-Sayed**[2] **· Mohamed A. El-Rashidy**[2]

Received: 12 May 2020 / Revised: 7 August 2020 / Accepted: 26 August 2020
Published online: 04 September 2020

Abstract
This paper suggests an IoT based smart farming system along with an efficient prediction method called WPART based on machine learning techniques to predict crop productivity and drought for proficient decision support making in IoT based smart farming systems. The crop productivity and drought predictions is very important to the farmers and agriculture's executives, which greatly help agriculture-affected countries around the world. Drought prediction plays a significant role in drought early warning to mitigate its impacts on crop productivity, drought prediction research aims to enhance our understanding of the physical mechanism of drought and improve predictability skill by taking full advantage of sources of predictability. In this work, an intelligent method based on the blend of a wrapper feature selection approach, and PART classification technique is proposed for crop productivity and drought predicting. Five datasets are used for estimating the proposed method. The results indicated that the projected method is robust, accurate, and precise to classify and predict crop productivity and drought in comparison with the existing techniques. From the results, the proposed method proved to be most accurate in providing drought prediction as well as the productivity of crops like Bajra, Soybean, Jowar, and Sugarcane. The WPART method attains the maximum accuracy compared to the existing supreme standard algorithms, it is obtained up to 92.51%, 96.77%, 98.04%, 96.12%, and 98.15% for the five datasets for drought classification, and crop productivity respectively. Likewise, the proposed method outperforms existing algorithms with precision, sensitivity, and F Score metrics.

Keywords Machine learning · Internet of things · Smart farming · Prediction · Drought · Crop productivity · And · Feature selection

✉ Ezz El-Din Hemdan
ezzvip@yahoo.com

Extended author information available on the last page of the article

1 Introduction

Recently, Agriculture is considered among the key strengths of the global and country's economy [17]. The practice of farming is one of the main occupations in the world and the product key the variety of crops. In recent times, agriculture is facing problems that may threaten its future, such as drought, crop quality, and productivity problems, Yield prediction problems [24]. The world's population is growing by about three people per second, equivalent to 250,000 people per day, and by 2025 the world's population will reach 8 billion, and the planet's population is expected to reach about 9.6 billion in 2050, according to figures from the Food and Agriculture Organization (FAO) of the United Nations to keep pace with this steady increase, farmers must upturn food production while conserving the environment and use natural resources rationally, but they cannot do it on their own, and customary farming techniques do not enable them to do. So, modern technologies play an essential role in helping to face the growing food needs of the world's population.

Agriculture is suffering a great transformation in the gathering and use of data to apprise useful farming decisions. Smart farming is the utilizing of modern Information and Communication Technology (ICT) like machine learning algorithms [10] in agriculture and the rationalization of the use of natural resources, as a capital-based system, advanced technology in food farming in sustainable and clean ways. Innovative technologies are bringing assistance to the majority of people around the world in various applications. Presently, the Internet of Things (IoT) [14], and data analysis such as big data analytics and data science [15, 18] have begun to play an important role in people's daily lives, extending their abilities to adjust the environs around them. Generally, the agro-industrial and environmental areas apply IoTs and data analysis in both diagnostics and control of the smart farming systems with providing vital information to the final farming, the consumer about the basis, and the properties of the agro-products and systems.

In past years, some methods have been established to solve the problems and impediments that appear with smart farming such as disease detection, yield prediction, species recognition, crop productivity problems and drought, and irrigation management. There is a few research work in using data analysis methods for a better decision support system for farming data. Most of them deal with customary techniques without considering the performance of the model, some of them neglect data preprocessing in the primitive stage. IoT provides solutions in various areas like healthcare, security, smart homes, retail, smart cities, and agriculture. The deployment of IoT in agriculture is the ideal solution due to there is a need for continuous monitoring and control in this area. In agriculture, IoT is used in agriculture applications that are micro-agriculture, livestock, and greenhouses, which are grouped into diverse control areas. All of these applications can be monitored with the help of different Internet-based information sensors/devices using wireless sensor networks (WSNs) that support farmers gather relevant data through IoT based sensors. Some IoT-based devices analyze and process data remotely through the application of cloud services, helping researchers and agriculturists to make better decisions.

In this research, an efficient prediction method called WPART based on the wrapper feature selection approach, and the PART classification technique are used for crop productivity and drought predicting. The Wrapper algorithm is essentially solving the optimizing the classifier performance problem. While, the PART is a partial decision tree method, which is the advanced version of C4.5 and RIPPER methods. The key specialty of the PART algorithm is that it does not need to achieve global optimization as C4.5 and RIPPER to produce the appropriate rules [2]. The entire proposed decision support system consists of four stages, the first stage collects the data of the environmental indicators based on IoT system, the second stage preprocesses data, the third stage uses the Wrapper feature selection approach to analyze the environmental indicators, and select the effective indicators of the farming system, and in the fourth stage, PART supervised machine learning algorithms are used to predict the drought depending on the selected features.

The proposed system was evaluated using five databases related to crop productivity and drought data that can be collected from IoT based smart farming system, and it achieved the highest accuracy, precision, sensitivity, and F-Score compared with traditional machine learning techniques. In conclusion, the contribution of this paper is as follows:

- Development of country with low cost, high-level support to speed up the efficacy of the cultivations, enhancing the environmental sustainability of the comprehensive process, and the production quality besides realization of low power consumption an ergonomic, diminished platform, and great performance.
- Develop an efficient decision support system and an effective trigger to connect farmers and experts as shown in Fig. 1 for agriculture applications such as detection at an early stage the presence of drought using machine learning. This illustrated in Section 3.
- Provide an IoT based farming system as shown in Fig. 2 that can support the deployment of the low-cost farming system. The IoT system explained in Section 2 as a part of the decision support system.
- Present a method called WPART that can classify agricultural data for crop productivity and drought prediction. This showed in Fig. 3 with the IoT based smart farming system in Section 3.
- Conducting a systematic evaluation of the proposed decision support system based on machine learning performances through different experiment consequences over different datasets using a 10-fold cross-validation method. The analysis of experimental results is presented in Section 4.

The remaining of this paper is prepared as follows: Section 2 provides related work while Section 3 describes the proposed decision support system for crop productivity and drought prediction in IoT based smart farming. The experimental study and results analysis are presented in Section 4 while the paper conclusion and future scope are delivered in Section 5.

2 Related work

In recent times, there is a large number of research in the use of machine learning techniques in the arena of smart farming. In [1], the author looks into the drought prediction problem using deep learning algorithms. It proposes a Deep Belief Network involving two Restricted Boltzmann Machines for long-term drought prediction using lagged values of Standardized Streamflow Index (SSI) as inputs. The study relates the efficiency of the proposed approach to that of classical methods like Support Vector Regression (SVR) and Multilayer Perceptron (MLP) for predicting the different time scale drought conditions. The proposed approach revealed an edge in performance over the classical methods. In [5], they proposed an ensemble method used to project crop production over some time. This method is compared to Naive

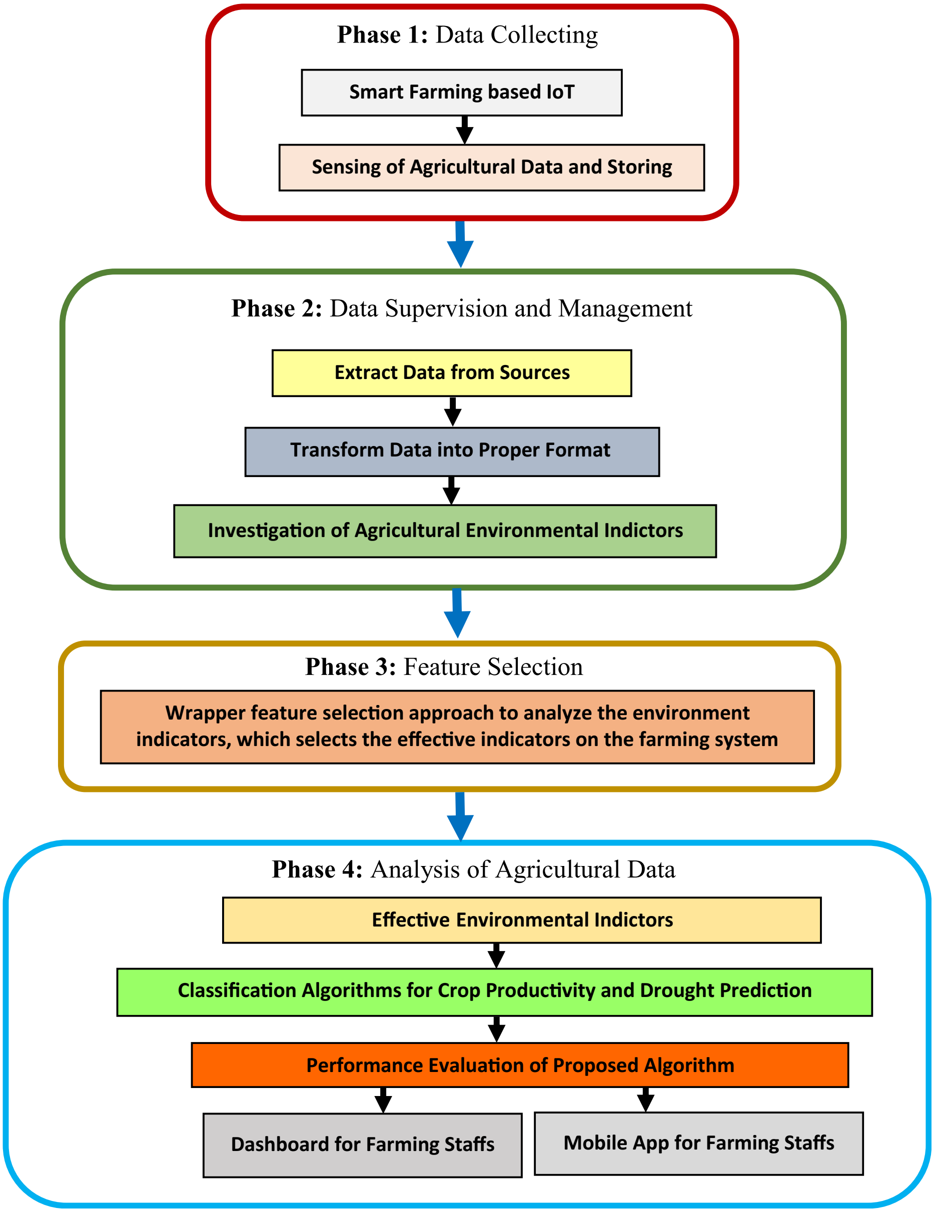

Fig. 1 Proposed Decision Support System for Smart Framing

Bayes and SVM methods. The two metrics used for the prediction of output are the accuracy and the classification error.

In [29], the IoT system was implemented to control the environmental issues in the crop fields, it involves three parts web application, and mobile application and finally hardware components. In [13] has offered a review on several IoT-based machine learning models to their particular architectures and contributions in various fields, such as agriculture, environment, and energy management using IoT applications, and machine learning techniques. In

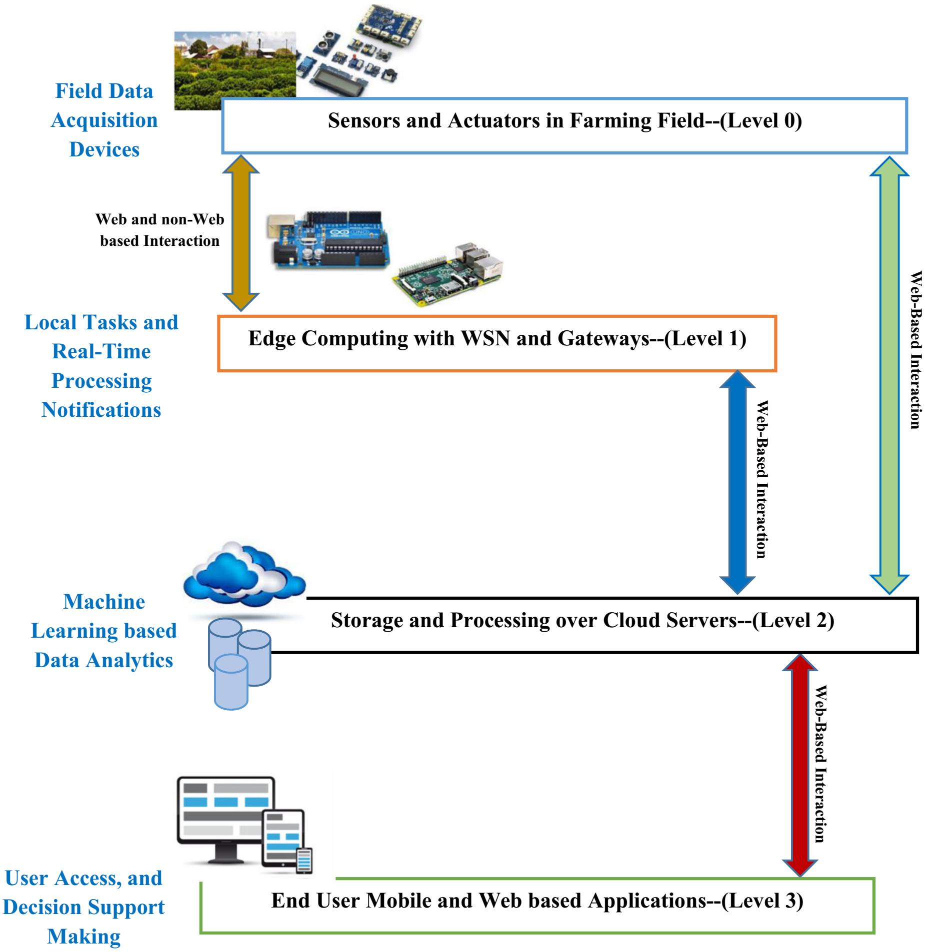

Fig. 2 Proposed IoT based Smart Farming System

[28], the proposed mySense system to organize data gaining actions to address common PA/PV issues. The system constructs over four layers: sensor and sensor nodes, crop field and sensor networks, cloud services, and support to front-end applications. In [26] a smart agriculture based monitoring system was presented to monitor temperature and soil humidity. The system handles the sensed data and applies the essential action based on the values of temperature and soil humidity deprived of human intervention.

In [27], the author presents different types of machine learning algorithms for predicting flood risk and classifying the results into three categories: normal, abnormal, and high-risk floods. Extensive research suggests that AI algorithms can result in enhancement when used in the pre-processing of flood data. These approaches have helped to gain better accuracy in classification techniques. Neural network engineering generally produces good results in many applications, however, the results of our experiments have shown that random forest work produces optimal results compared to reference models. In [23], the author presents an approach for water level forecasting in conjunction with the severity of the flood using the group model. This approach takes advantage of the latest developments in the Internet of

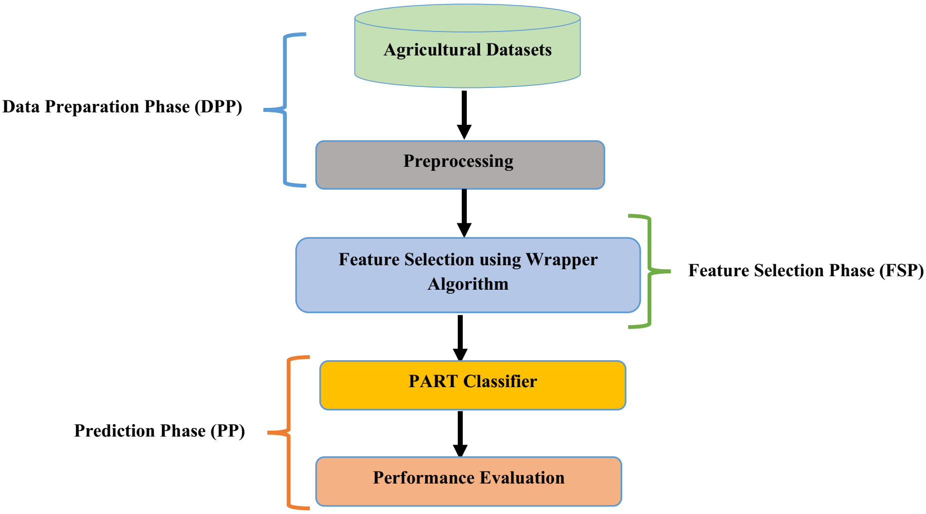

Fig. 3 Proposed Prediction Model

Things (IoT) and machine learning for automated analysis of flood data that may be useful for preventing natural disasters. Research results suggest that ensemble learning provides a more reliable tool for predicting flood severity levels.

The author in [4] described farming systems based on integrated systems, Internet systems, and wireless sensing networks for the agricultural field and livestock farms. This paper describes electronic circuit systems for systems, network protocols used, smart remote surveillance systems for computers, smartphones, etc. Later includes some suggestions. In [21] the authors propose a basic survey of various methods of crop selection, crop sowing, injunctive detection, system monitoring, and thus production. This work focused on the entire system based on image processing, the Internet of Things, machine learning, and artificial intelligence.

The authors in [20] promote optimal use of water with higher yield by assessing the relationship between different physical parameters like soil temperature, soil moisture, air temperature, UV density, and air humidity with plant growth. With the use of machine learning methods, predicting soil moisture gives the benefit to conserve water sources. In [33] the system will integrate IoT sensors such as PH sensor, humidity, precipitation, temperature, and humidity to monitor data from these sensors and apply machine learning algorithms: Random Forest and GDBoost. The most suitable crops are predicted according to the current environment. In [19], they provided an outline about IoMT and future research directions along with Opportunities, Challenges, and Solutions in this domain. A description of a comparison of some related schemes by highlighting the aim, proposed solution, pros, and cons are tabulated in Table 1.

This study proposes an efficient prediction method consists of three phases precisely: (1) Data Preparation Phase (DPP), (2) Feature Selection Phase (FSP), and (3) Prediction Phase (PP). This research aims to build an innovative intelligent and robust prediction method based on machine learning techniques that can classify drought and crop productivity in smart agriculture applications. So, it is required to employ a feature selection technique that can specify important features successfully. The feature selection process is achieved by a Wrapper algorithm that can effectively choose the most essential features. Then, we use the PART classifier to classify and predict drought, and crop productivity in the smart farming system. This efficient prediction method is developed to help farmers in their decision making and

Table 1 A comparative analysis between the proposed work and existing schemes

	aim	proposed solution	pros	cons
Proposed Method	Provide an IoT based smart farming system along with an efficient prediction method called WPART based on machine learning techniques to predict crop productivity and drought	an intelligent method based on the blend of a wrapper feature selection approach, and PART classification technique is proposed for crop productivity and drought predicting	The proposed method outperforms existing algorithms for the five datasets for drought classification, and crop productivity respectively.	Don't focus on using the time-series analysis to predict future values based on previously observed values.
[6]	Object-oriented expert system for Tea Crop	A total of 65 real field cases were taken from the system which evaluated the relevant parameters and the results were almost satisfactory. The system was stationed in the tea fields for two years	An expert system for help farmers in improving the quality of agriculture system for Tea crop	The accuracy of the system was only 90%. The remaining inaccuracy was the result of inappropriate diagnosis
[30]	Image Processing for Wheat Crop	SVM and neural networks were used for classification purposes. The accuracy of SVM was 86.8% and the neural network was 94.5% accurate. The algorithm used was more accurate than manually developed an algorithm	Using image processing tools for the classification of Wheat in an efficient manner using SVM and NN	More precisely algorithms can be designed which will have less computational cost. Additionally, various varieties of wheat can be classified using different feature sets

obtain better predictive abilities than other existing techniques. The proposed method compared with most popular algorithms such as the Logistic Regression (LR), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (NB), Random Forest (RF) and Artificial Neural Network (ANN) [3, 12, 16, 22].

3 Proposed decision support system for smart framing

In this section, methodical phases and block diagram of the proposed system are clarified. The methodical phases of the proposed system consist of four stages like shown in Fig. 1 as the following:

Phase 1. Data Collecting

- Raw data can be gathered from smart farming based IoT sensors, the IoT system consists of different levels as shown in Fig. 2 similar to the following:

- **Level 0:** In this level, IoT sensors and actuators are responsible for gathering and monitoring different environmental parameters. For sensing or collecting the parameters, various types of sensors are deployed over the farming field. Generally, three types of the sensor can be used such as pressure sensor to observe air humidity level, temperature sensor to observe temperature level, and rainfall sensor to observe the amount of rain. These sensors are attached to the IoT based controller such as Raspberry pi3 [32] and Ardunio.
- **Level 1:** In this level, the gateway and edge devices operate between the sensor network and cloud servers. These devices such as IoT based controller platform.
- **Level 2:** In this level, cloud computing is utilized, which is an emergent technology and is used efficiently in smart farming. The proposed system uses cloud computing servers for the storage and processing of different agricultural field data. The IoT based controller sends the gathered data to the respective channel periodically via a communication protocol.
- **Level 3:** To develop mobile and web applications that communicate with cloud servers to get the analytics results from applying machine learning methods on the farming stored data to help in making the best decision by farmers and official staff in agriculture organization.
- Sensing of specific Agricultural data and Storing in databases.
- After gathering all data, it is used to create an agricultural dataset that will be submitted to ML algorithms for classification investigation.

Phase 2. Data Supervision and Management

- Extract data from dataset sources
- Transform and convert agriculture data into the proper format

The dataset that collected from the IoT system in any smart farming application needs to be accumulated before it can be put to use, it may have missing data values or noise data. These missing values may be produced from faulty sensors or lack of communication among the components in a data collection system. These missing values affect the performance of the system and need to be addressed accordingly. In this research, missing values are replaced with their appropriate values using a linear regression technique.

Phase 3. Feature Selection

In this stage, the Filter and Wrapper feature selection approach is used to analyze the environmental indicators, which selects the effective indicators on the farming system. Furthermost Filter approaches calculate a score for each feature and then select the features with the highest scores [25], this score depends on the similarity and dissimilarity data of each feature with the output labels, while Wrapper methods are based on the greedy search algorithm, it evaluates all possible combinations of the features and chooses the combination that produces the best result for a specific machine learning technique, Wrapper method finds the best set of features for a specific algorithm [25]. In this research Wrapper selection technique is used to select the effective environment indictors, which is efficient rather than filter selection techniques in the case of the high similarity data of each attribute and the smallest number of dataset attributes.

Phase 4. Analysis of Agricultural Data

- Preparing the data of effective environment indicators relating to drought prediction.
- Applying the supervised machine learning algorithms for agro data classification.

In this stage, the PART algorithm is used to construct the classifier for prediction purpose, it generates the distinctive rules that can distinguish the similarity data of each subset of dataset features with the output labels. When the similarity of the data of each dataset feature is high, this algorithm is accurate in the classification process, which tries to find the distinctive rules of these data with the output labels. The PART algorithm is trained by the effective indicators that selected by the Wrapper technique, and it uses the separate-and-conquer approach to develop a partial decision tree in each iteration, each tree uses subsets of cases extracted from the complete training set to generate rules (http://www.indiawaterportal.org/articles/district-wise-monthly-rainfall-data-list rain gauge stations-India-meteorological-department/). Given training vectors $X_i \in R^n$, i = 1… l and a label vector Y$\in R^l$, a decision tree recursively partitions the space such that the samples with the same labels are grouped. Let the data at node m be represented by Q. For each candidate split$\theta = (j, t_m)$ consisting of a feature j and thresholdt_m, partition the data into and Subsets $Q_{LEFT(\theta)}$and $Q_{RIGHT(\theta)}$

$$Q_{LEFT(\theta)} = \left((X, Y)|X_j < t_m\right) \tag{1}$$

$$Q_{RIGHT(\theta)} = \left(Q|Q_{LEFT(\theta)}\right) \tag{2}$$

The impurity at m is computed using an impurity function H (), the choice of which depends on the task being solved (classification or regression).

$$\mathrm{G}\ (\mathrm{Q}, \theta) = \frac{n_{LEFT}}{N_m}\mathrm{H}\left(Q_{LEFT(\theta)}\right) + \frac{n_{RIGHT}}{N_m}\mathrm{H}\left(Q_{RIGHT(\theta)}\right) \tag{3}$$

Select the parameters that minimize the impurity

$$\theta^* = argmin_\theta\ \mathrm{G}\ (\mathrm{Q}, \theta) \tag{4}$$

Recurse for subsets $Q_{LEFT\left(\theta^*\right)}andQ_{RIGHT\left(\theta^*\right)}$ until the maximum allowable depth is reached, $N_m < min_{samples}$ or $N_m = 1$.

Classification outcome taking on values 0, 1… K-1, for node m, representing a region $R_m$with observations, let

$$P_{mk} = {}^1\!/_{N_m} \sum_{X_i \in R_m} I(y_i = k) \tag{5}$$

Be the proportion of class k observations in node m.

Common measures of impurity are Gini.

$$\mathrm{H}\ (X_m) = \sum_k P_{mk}(1 - P_{mk}) \tag{6}$$

Table 2 Datasets used

Datasets	Attributes Names	Number Attributes	Instances
Dataset 1	District, year, month, rainfall, Average Rainfall, Temperature, average Temperature, Pressure, Average Pressure	9	1537
Dataset 2	District, Year, Season, Crop, Area, Production, Productivity, Rainfall, Temperature	9	105
Dataset 3	District, Year, Season, Crop, Area, Production, Productivity, Rainfall, Temperature	9	112
Dataset 4	District, Year, Season, Crop, Area, Production, Productivity, Rainfall, Temperature	9	104
Dataset 5	District, Year, Season, Crop, Area, Production, Productivity, Rainfall, Temperature	9	96

Entropy.

$$\mathrm{H}\ (X_m) = -P_{mk}\ \log P_{mk} \tag{7}$$

And Misclassification.

$$\mathrm{H}\ (X_m) = 1-MAX_{(p_{mk})} \tag{8}$$

Where X_m is the training data in node m.

4 Experimental study and analysis

This section provides the performance evaluation and analysis of the efficient proposed WPART method.

4.1 Datasets used and experimental environment

The proposed analysis based machine learning model evaluates the experimental agriculture crop productivity and drought using the agriculture data obtained from data sources in (http://mahaagri.gov.in/cropwatch/asp/rpt1.asp/) [11]. The comprehensive information of these datasets like the number of attributes and instances present, instances, and the distribution of dataset for training and testing based on 10 fold cross-validation is highlighted in Table 2. The cross-validation helps us better use our data, and it gives us much more information about the proposed algorithm performance.

Table 3 Statistical analysis of Bajer Dataset features

Feature name	Minimum	Maximum	Mean	StdDev
Year	2001	2014	2007.183	3.909
Area	2	324,200	60,211.57	85,146.677
Production	1	258,500	48,568.452	69,024.676
Productivity	0.28	1.409	0.68	0.253
Rainfall	0	4	1.419	0.913
Temperature	0	4	2.011	0.938

Table 4 Statistical analysis of Sugarcane Dataset features

Feature name	Minimum	Maximum	Mean	StdDev
Year	2001	2014	2007.537	4.008
Area	3	205,500	30,835.731	42,930.41
Production	229	20,049,700	2,539,993.694	4,006,062.424
Productivity	37.692	101.29	71.251	14.489
Rainfall	0	4	1.454	0.911
Temperature	0	4	2.222	0.95

The agriculture datasets are applied to the proposed machine learning algorithm for crop productivity and drought predict purpose. The experiments are implemented using machine learning algorithms written in python and Waikato Environment for Knowledge Analysis (Weka) running on Windows 8 with Intel® Core™ i5-4288U CPU @ 2.60 GHz processor and 8.00GB RAM.

Before instigating the experimental study, it is necessary to prepare the information perceptively before using machine learning models. The existing real-world farming data obtained from sensors are distributed non-uniformly, and hereafter, the data cannot be exploited validly amid training and testing machine learning models. In this manner, the input features are normalized, and numeric data are converted to categorical variables through the data label encoder for proficient processing. The results achieved are perceived from the perspective of multiclass classification, as well as for the individual class basis.

As shown in Tables 3, 4, 5, 6 and 7, StdDev and mean values indicated to the highest inter similarity of the feature values in all datasets, these similarities have been negative effects on the feature selection process using filter methods compared to Wrapper method as shown in Fig. 5 and Table 9. Filter methods select the features based on calculating a score for each feature, which this score relates to the inter similarity and dissimilarity data of each feature with the output labels, while Wrapper methods are based on the greedy search algorithm, it estimates all possible combinations of the feature subsets and selects the subset that achieves the highest accuracy for a specific machine learning technique. Therefore, the Wrapper selection method is used to select the effective environment indictors for the PART classification algorithm, which is an efficient classification algorithm compared with the other algorithms as shown in Tables 10, 11, 12 and 13 in the results analysis division. Table 8 shown the label classification datasets results.

In the experiments, various classification techniques have been applied to five datasets that collecting from the farming system to predict the drought as well as crop productivity. These techniques such as RF, LR, KNN, ANN, and NB respectively. The proposed prediction 'WPART' method using a wrapper algorithm for Feature selection and PART algorithm for

Table 5 Statistical analysis of Jowar Dataset features

Feature name	Minimum	Maximum	Mean	StdDev
Year	2001	2014	2007.631	4.114
Area	2	162,800	37,458.854	47,719.697
Production	2	250,200	43,028.825	60,654.71
Productivity	0.25	2.085	1.033	0.411
Rainfall	0	4	1.408	0.923
Temperature	0	4	2.117	0.983

Table 6 Statistical analysis of soybean Dataset features

Feature name	Minimum	Maximum	Mean	StdDev
Year	2001	2014	2007.765	4.059
Area	28	373,000	113,326.814	123,508.237
Production	39	670,400	118,562.745	147,825.558
Productivity	0.279	2.392	1.265	0.504
Rainfall	0	4	1.51	0.972
Temperature	0	4	2.255	1.021

classification. Wrapper methods are based on a greedy search algorithm, it evaluates all possible combinations of the feature subsets and selects the subset that achieves the highest accuracy for a specific machine learning algorithm. Therefore, the Wrapper selection method is used to select the effective environment indicators for the PART classification algorithm, which is an efficient classification algorithm compared with the other algorithms. The WPART method composed of three phases as shown in Fig. 3 exactly as follows:

1. **Data Preparation Phase (DPP):** In this phase, the agriculture data collected from the IoT based smart farming system needs to be preprocessed and handled before it can be put to use in the next phases.
2. **Feature Selection Phase (FSP):** In this phase, the Wrapper feature selection approach is used to analyze the environmental indicators, which selects the effective indicators on the IoT based farming system.
3. **Prediction Phase (PP):** In this phase, the PART algorithm is used to construct the classifier for drought and crop productivity prediction, then the proposed prediction method is evaluated and compared with existing ones.

4.2 Performance metrics

The performance evaluation of machine learning model in a precise classification and prediction task is measured by several statistical and mathematical models are used. These performance evaluation metrics such as accuracy, precision, recall, and f-score are used to compare the performances of the proposed algorithm compared to existing ones. The remarks are achieved by considering true negatives (TN), true positives (TP), false positives (FP), and false negatives (FN). The accuracy of the classification model on a specified test is the percentage of the test set that is correctly classified by the classifier. Precision is the measure

Table 7 Statistical analysis of drought Dataset features

Feature name	Minimum	Maximum	Mean	StdDev
Rainfall	1	10	1.89	1.54
Average Rainfall	1	6	1.844	1.278
Average Temperature	2	9	4.917	2.071
Year	2001	2016	2008.5	4.611
Temperature	1	6	3.628	1.431
Pressure	1	5	2.577	0.918
Average Pressure	1	5	2.698	0.915

Table 8 The label classification datasets results

Dataset name	Class Label	Description
Bajra Crop	Crop Label	1,2,3,4 where 1 count 34 in Bajra dataset,2 count 43, 3 count 13, and 4 count 3
Sugarcane Crop	Crop Label	1,2,3 where 1 count 60, 2 count 2, and 3 count 27
Jower Crop	Crop Label	1,2,3,4,5 where 1 count 17, 2 count 42, 3 count 31, 4 count 4, and 5 count 9
Seabean Crop	Crop Label	1,2,3 where 1 count 65, 2 count 21, and 3 count 16
Drought	Drought Classification	1,2 where 1 count 1232, and 2 counts 304

of the correctness of positive labeled examples. The recall is the measure of completeness or accuracy of positive examples that how many examples of the positive class are labeled correctly. Accuracy, precision, sensitivity (recall), and F-score are calculated as per in Eqs. (9)–(12) respectively [6–9]:

$$Accuray = \frac{TP + TN}{TP + FP + FN + TN} \tag{9}$$

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$\text{Senstivity} = \frac{TP}{TP + FN} \tag{11}$$

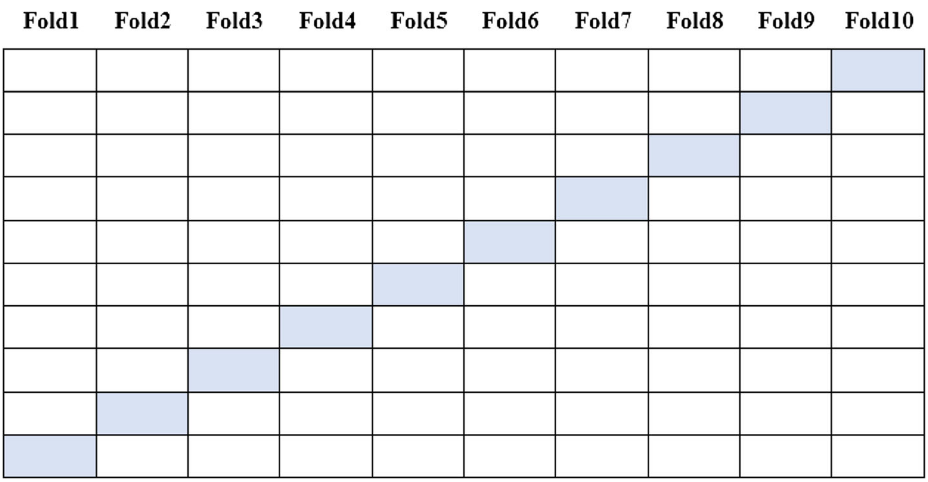

Fig. 4 10 Fold-Cross Validation Diagram

Table 9 Selected features of each feature selection algorithm

Feature selection algorithm	Number of selected features	Selected features
PCA	6	Year, Rainfall, Average Rainfall, Average Temperature, Pressure, and Average Pressure.
Info Gain	8	District, Year, Month, Rainfall, Temperature Average Temperature, Pressure, and Average Pressure.
Gain Ratio	7	District, Year, Month, Rainfall, Temperature Average Temperature, and Average Pressure.
Correlation Attribute	7	District, Year, Month, Pressure, Temperature Average Temperature, and Average Pressure.
Cfs	7	District, Month, Rainfall, Average Rainfall, Pressure, Average Temperature, and Average Pressure.
Wrapper	5	District, Year, Rainfall, Average Rainfall, and Average Temperature.

$$\text{F1 Score} = \frac{2*(Recall*Precision)}{\text{Recall} + \text{Precision}} \tag{12}$$

4.3 10-fold cross-validation method

Commonly, this method is utilized to evaluate prediction methods. This method has the advantage of reducing the bias that happens in the training and testing datasets using random sampling. First, the dataset is separated into k equal subsets. Then, (k-1) subsets are used to train the model and the k^{th} subset is used for testing the model. This method is recurrent k times using k different training and testing datasets.

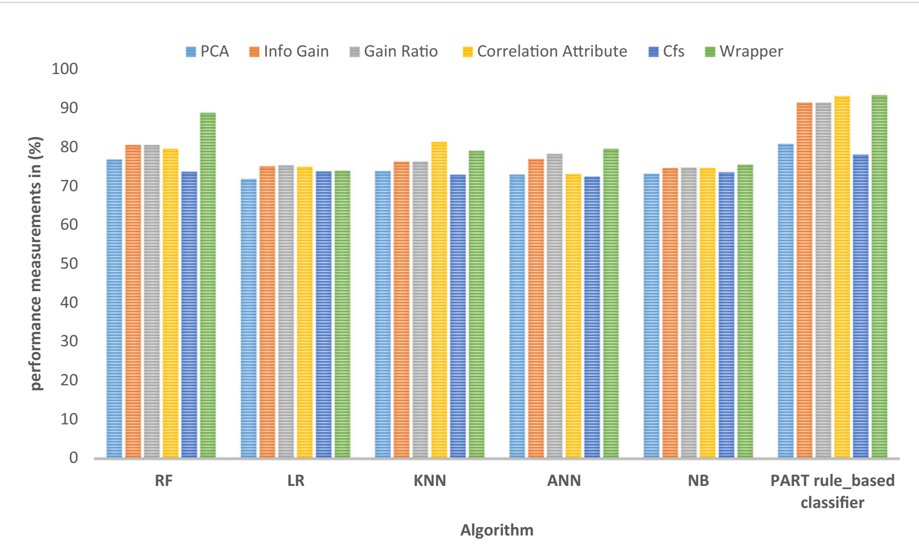

Fig. 5 Accuracy of the various feature selection techniques for different classification algorithm to predict drought

The resolve of using cross-validation is to test the ability of the proposed prediction method to predict the result and to overcome overfitting or selection bias. Stratified 10-fold cross-validation was applied to re-organize the dataset to validate that each fold is a suitable reordering of the entire dataset. 10-fold cross-validation is the best choice method, as it offers a less biased estimate of the accuracy. The dataset is separated into ten equal subsets. Nine subsets are used for training and the remaining subset is used for testing the prediction technique. This process is repeated ten times using different subsets. The performance metrics results are the average result of the ten stages to achieve more accurate results. Fig. 4 signifies a graph that describes the 10-fold validation in a rich form. In each fold, the nine unshaded rectangles exemplify the training set and the tenth shaded rectangle represents the testing set.

4.4 Results analysis

Some feature selection techniques are applied to select the effective features that contributing to improving crop productivity and drought prediction in the future. The time feature appeared in the feature selection results, but the goal of the proposed system is to predict the drought in the future. Therefore, the time feature was ignored in the training and testing phases. As shown in Table 9, the Wrapper feature selection technique identifies the lowest number of the effected features compared with the various techniques of the filter selection approach. The Wrapper feature selection technique is likewise evaluated to realize the efficiency of the selected features compared with the other techniques, the results show that the Wrapper feature selection is higher accuracy than Filter feature selection techniques as shown in Fig. 5.

Feature selection methods depend on evaluating each feature through its similarity data with the output labels. In the case of the high similarity data of each attribute, the Wrapper selection method is efficient rather than filter selection techniques as shown in Fig. 5, because it estimates all possible combinations of the features and chooses the combination set that produces the highest accuracy. Therefore, the proposed system based on the Wrapper selection method to select the least number of effective environment indicators.

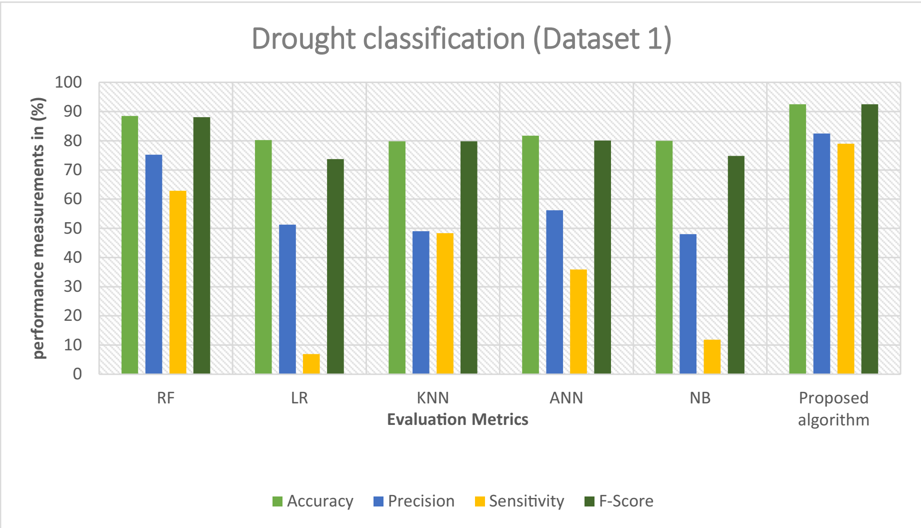

Fig. 6 Results of Drought classification

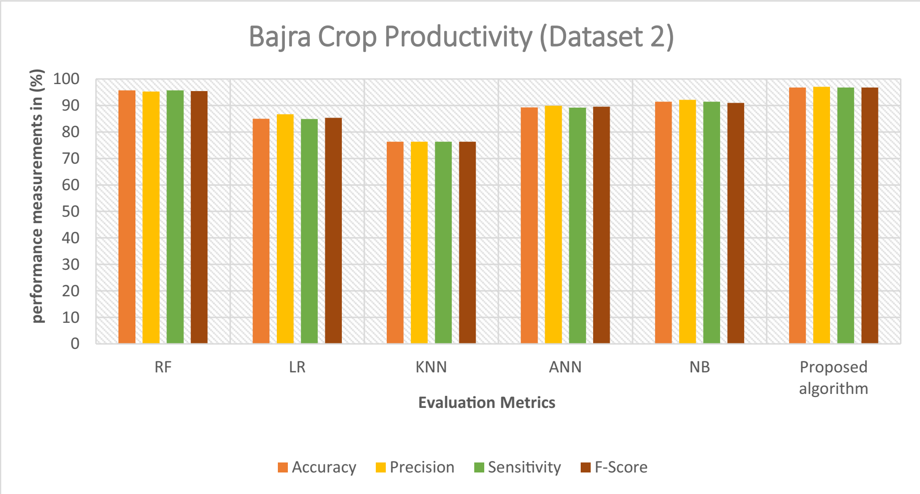

Fig. 7 Results of Bajra Crop Productivity

From Fig. 5, it is clear that the proposed system that depending on the Wrapper feature selection and PART algorithm achieved the highest accuracy. The proposed system has been evaluated by using accuracy and F-Score metrics, which F-Score is an appropriate metric in the case of imbalanced data. And the results in Fig. 5 also show the proposed system obtained the highest accuracy values compared with the other classification approaches using the Wrapper algorithm.

The system comprises of extraction of data related to parameters required for drought classification, which are, Rainfall, Average Rainfall, Average Temperature, Temperature and Pressure, Average Pressure, Year, Month, District. This data will then be standardized to have a consistent format and loaded into the database. The analysis of this data will produce trends and correlations with other parameters, which can be used to calculate the probability of drought and classifying it as low, medium, or high. Likewise, the system works on the features

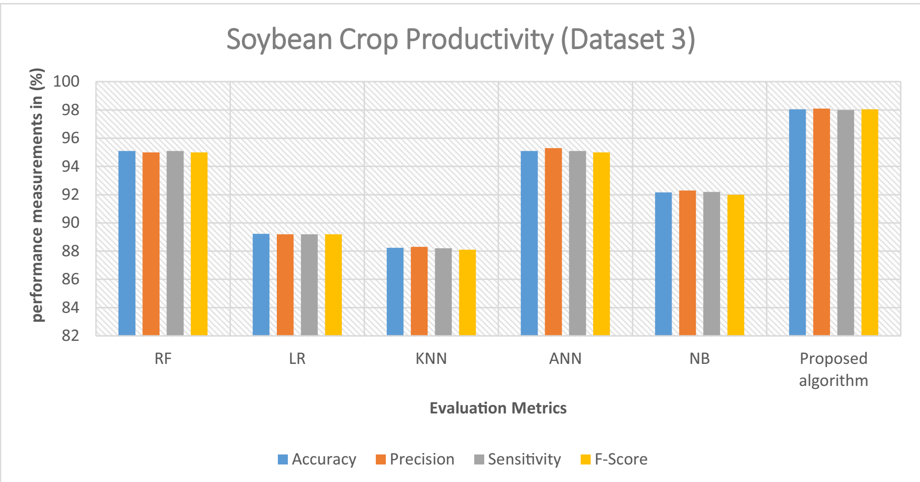

Fig. 8 Results of Soybean Crop Productivity

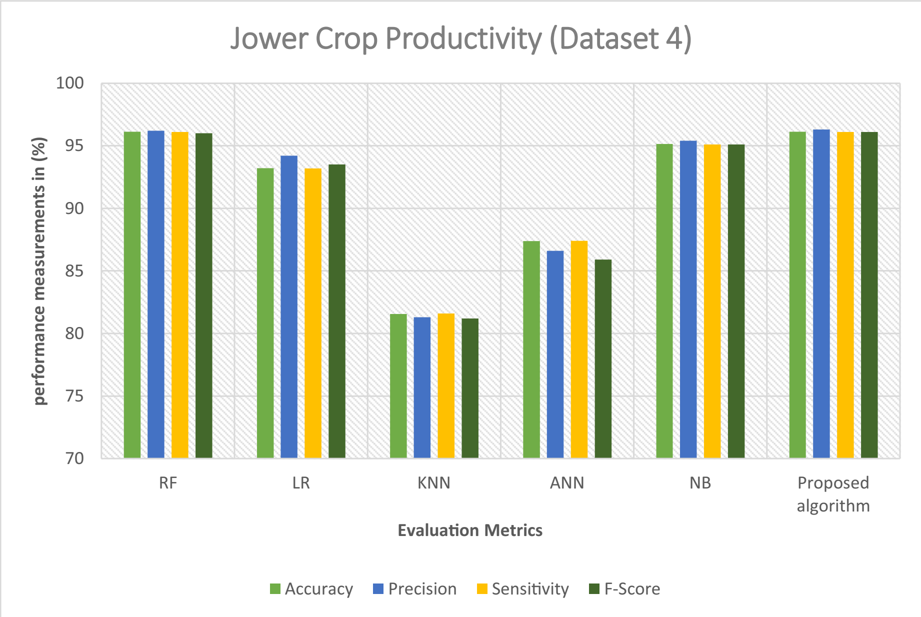

Fig. 9 Results of Jower Crop Productivity

of District, Year, Season, Crop, Area, Production, Productivity, Rainfall, and Temperature for crop productivity classification.

Figs. 6, 7, 8, 9 and 10 shows the performance of the proposed algorithm for crop productivity and drought predictions where the proposed algorithm gives high performance in comparison with existing and most popular classifiers using ten-fold cross-validation based on precision, sensitivity, and F Score metrics.

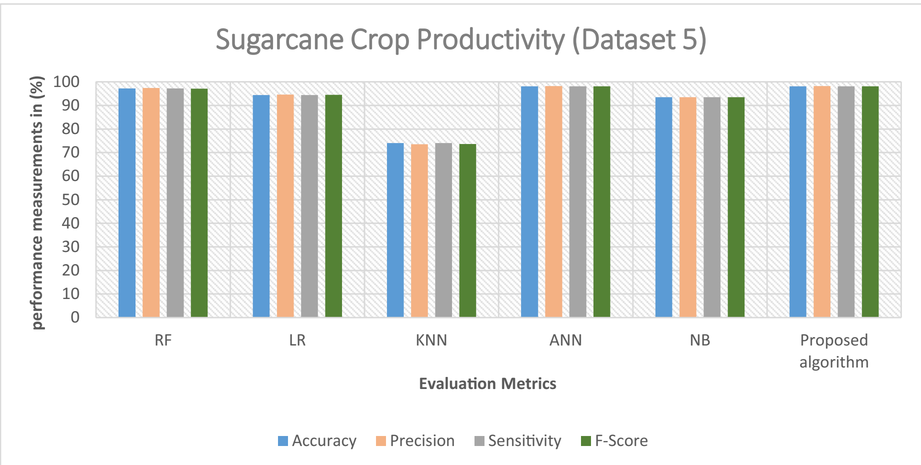

Fig. 10 Results of Sugarcane Crop Productivity

Table 10 Accuracy results of five datasets for Drought and Crops Productivity

Algorithm	Accuracy				
	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
RF	88.54	95.70	95.10	96.12	97.22
LR	80.27	84.95	89.22	93.21	94.44
KNN	79.82	76.34	88.24	81.55	74.07
ANN	81.77	89.25	95.10	87.38	98.15
NB	80.01	91.40	92.16	95.15	93.52
Proposed algorithm	**92.51**	**96.77**	**98.04**	**96.12**	**98.15**

4.4.1 Comparative study and analysis

In this part, we provide a comparison between the proposed method and existing algorithms over the five used datasets. The results are contained from Table 10, 11, 12 and 13 of our pilot study, listing the results of supervised learning techniques, through the application of validation and testing for the five datasets, respectively, and The outcomes gained from our experiments indicate that accuracy, precision, F-score, and sensitivity of our model outperformed all other models, thus the performance of the proposed algorithm for the five datasets for both crop productivity and drought predictions gives high performance is compared with existing algorithms. In our experiments, single classifiers produced accuracy values of 88.54, 80.27, 79.82, 81.77,80,01 and 92.51 for RF, LR, KNN, ANN, and NB respectively, during the validation/training for Bajra Dataset .while, the combination of PART classifier algorithm with wrapper feature selection produce accuracy values 92.51 for Bajra dataset.

As revealed in the results, the proposed algorithm-generated high sensitivity and F-Score for all categories while single classifier performed well-using sensitivity and F-Score. This is due to the variety of representations of datasets. In general, the proposed algorithm retains attractive features of the decision list, such as handling unrelated/ redundant descriptors. In terms of the training procedure, this model was much faster compared to the single classifier. One of the main reasons why the proposed algorithm resulted in higher accuracy was that it was able to generalize better using the aggregated evidence of member works. Although some samples in data sets are misleadingly labeled, the WPART can easily evaluation lost values, and work efficiently with unbalanced data, challenging another model.

Table 11 Precision results of five datasets for Drought and Crops Productivity

Algorithm	Precision				
	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
RF	75.20	95.00	95.00	96.20	97.40
LR	51.22	89.20	89.20	94.20	94.60
KNN	49.00	88.30	88.30	81.30	73.50
ANN	56.19	95.30	95.30	86.60	98.20
NB	48.00	92.30	92.30	95.40	93.50
Proposed algorithm	**82.47**	**98.10**	**98.10**	**96.30**	**98.20**

Table 12 Sensitivity results of five datasets for Drought and Crops Productivity

Algorithm	Sensitivity				
	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
RF	62.83	95.70	95.10	96.10	97.20
LR	6.91	84.90	89.20	93.20	94.40
KNN	48.36	76.30	88.20	81.60	74.10
ANN	35.86	89.20	95.10	87.40	98.10
NB	11.84	91.40	92.20	95.10	93.50
Proposed algorithm	**78.95**	**96.80**	**98.00**	**96.10**	**98.10**

Experimental results present comparative analysis with detailed results for better validation and analysis of the proposed method performance with most common existing classifiers. The results reveal that the proposed method enhances the classification precision, sensitivity, and F-Score, as well as enhancing feature selection outcomes. Likewise, From Table 14, the importance of the proposed system is quite evident. The proposed system is introduced to develop decision support making for drought and crop productivity prediction for smart IoT based smart farming using an efficient algorithm called WPART. Therefore, the proposed system can support to achieve the following:

- A systematic assessment of the proposed decision support system based on machine learning performance through different results of the experiment across different data sets using the 10-fold cross-data validation method. The proposed method achieves maximum accuracy comparing existing algorithms, obtained up to 92.51%, 96.77%, 98.04%, 96.12%, and 98.15% for the five data sets for drought classification, and crop productivity in return. Similarly, the superior performance of the proposed method was assessment measures when applied to available data.
- Recognize appropriate seasons for each crop, and the corresponding trend for that crop in a given area and suggest season crops such as Rice, and Bajra if the rainfall in a given area in the Kharif Season.
- Prediction of rainfall to alert the farmers of forthcoming drought so that they can prepare and take the right action.

Table 13 F-Score results of five datasets for Drought and Crops Productivity

Algorithm	F-Score				
	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
RF	88.10	95.40	95.00	96.00	97.10
LR	73.70	85.30	89.20	93.50	94.50
KNN	79.80	76.30	88.10	81.20	73.60
ANN	80.10	89.50	95.00	85.90	98.10
NB	74.80	91.00	92.00	95.10	93.50
Proposed algorithm	**92.50**	**96.80**	**98.04**	**96.10**	**98.10**

Table 14 Analysis Study with the previously proposed systems

Work	Agriculture Uses	Propose	Processing Data
[3]	A Smart IoT based monitoring system for field environment and collection of useful data and stored in ThingSpeak for future analysis	Provide a smart farming method based on Internet of Things (IoT) to deal with the adverse situations	Data of soil humidity and temperature
[2]	A Detection system of cherry branches with full foliage	Colored digital images depicting leaves, branches, cherry fruits, and the background	cherry tree branches with full foliage images
[32]	An expert system by integrating sensor networks with Artificial Intelligence systems such as neural networks and Multi-Layer Perceptron (MLP) for the assessment of agriculture land suitability	help the farmers to assess the agriculture land for cultivation in terms of four decision classes, namely more suitable, suitable, moderately suitable, and unsuitable	Data from pH sensor, soil moisture sensor, salinity sensor, and an electromagnetic sensor
[31]	A system which when deployed will give an insight into the real-time condition of the crop. The system leverages the Internet of Things (IoT) and Machine learning to produce an affordable smart farming module.	Predict the future condition of the crops based on its past data	Data of temperature and humidity, moisture soil, air quality, sunlight the plant
Proposed system	Decision support making for drought and crop productivity prediction in smart farming	Provide an IoT based smart farming system along with an efficient prediction method called WPART based on machine learning techniques to predict crop productivity and drought for proficient decision support making in IoT based smart farming systems	Data of District, year, month, rainfall, Average Rainfall, Temperature, average Temperature, Pressure, Average Pressure, Season, Crop, Area, Production, Productivity

5 Conclusion and future scope

In this paper, an efficient prediction method named WPART is presented that can be employed in the decision-making system in a smart farming environment that depending on IoT based system. The proposed method used both of Wrapper feature selection, and PART algorithm. The Wrapper technique is used to analyze collected data of the environment indicators to select the effective indicators on drought and crop production problems. The PART algorithm is used to construct the drought and crop productivity predictor. The highest rate of Accuracy and F-Score values obtained by using the proposed system into the real dataset collecting from the smart farming system. Accuracy is 93.42% and F-Score is 93.40%, which these rates are significant compared with the current methods. The results validate that the proposed method is effective in improving smart farming systems with decision support through the prediction of drought and crop productivity. Likewise, from the experimental results, the proposed method proved to be most accurate in providing predictions as well as the productivity of crops like Bajra, Soybean, Jowar, and Sugarcane. The proposed system accuracy obtained up to 92.12% for drought classification, and up to 98.15% for crop productivity.

For future improvement, we consider using time-series analysis to predict future values based on previously observed values. Also, we can broaden our scope by also adding other parameters such as soil quality, agricultural inputs, soil nutrients, irrigated area. These parameters should account for anomalies in the data, as well as improve the accuracy by multi-fold. Unsupervised clustering to label data for classifiers will also improve accuracy. Also, we planned to use IoT based computer vision system using deep learning models to improve the quality of production in the smart farming field.

References

1. Agana, NA and Homaifar, A (2017). A deep learning based approach for long-term drought prediction. In SoutheastCon 2017 (pp. 1-8). IEEE
2. Amatya S, Karkee M, Gongal A, Zhang Q, Whiting MD (2016) Detection of cherry tree branches with full foliage in planar architecture for automated sweet-cherry harvesting. Biosyst Eng 146:3–15
3. AshifuddinMondal M, Rehena Z (2018, January) Iot based intelligent agriculture field monitoring system. In 2018 8th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 625-629). IEEE
4. Ashok T, Varma PS (2020) Crop prediction based on environmental factors using machine learning ensemble algorithms, In *Intelligent Computing and Innovation on Data Science* (pp. 581–594). Springer, Singapore
5. Balakrishnan N, Muthukumarasamy G (2016) Crop production-ensemble machine learning model for prediction. International Journal of Computer Science and Software Engineering 5(7):148
6. Banerjee G, Sarkar U, Ghosh I (2017) A radial basis function network based classifier for detection of selected tea pests. International Journal of Advanced Research in Computer Science and Software Engineering 7(5):665–669
7. Ben-Bassat M (1982) Pattern recognition and reduction of dimensionality. Handbook of Statistics 2(1982): 773–910
8. Blum AL, Langley P (1997) Selection of relevant features and examples in machine learning. Artif Intell 97(1–2):245–271
9. Breiman L, Friedman J, Stone CJ, Olshen RA (1984) Classification and regression trees. CRC press
10. Brier ME, Ray PC, Klein JB (2003) Prediction of delayed renal allograft function using an artificial neural network. Nephrol Dial Transplant 18(12):2655–2659
11. Brown TS, Elster EA, Stevens K, Graybill JC, Gillern S, Phinney S, Salifu MO, Jindal RM (2012) Bayesian modeling of pretransplant variables accurately predicts kidney graft survival. Am J Nephrol 36(6):561–569
12. Chen, M, Narwal, N and Schultz, M (2019). Predicting price changes in Ethereum. International Journal on Computer Science and Engineering (IJCSE) ISSN, 0975-3397
13. Din IU, Guizani M, Rodrigues JJ, Hassan S, Korotaev VV (2019) Machine learning in the internet of things: designed techniques for smart cities. Futur Gener Comput Syst 100:826–843
14. El-Din, HE and Manjaiah, DH (2017). Internet of things in cloud computing. In Internet of Things: Novel Advances and Envisioned Applications (pp. 299–311). Springer, Cham
15. Essa YM, Hemdan EED, El-Mahalawy A, Attiya G, El-Sayed A (2019) IFHDS: intelligent framework for securing healthcare BigData. J Med Syst 43(5):124
16. Greaves, A and Au, B (2015). Using the bitcoin transaction graph to predict the price of bitcoin. No Data
17. Hemdan, EED and Manjaiah, DH (2018). Cybercrimes investigation and intrusion detection in internet of things based on data science methods. In Cognitive Computing for Big Data Systems Over IoT (pp. 39–62). Springer, Cham
18. Hemdan, EED and Manjaiah, DH (2020). Digital investigation of cybercrimes based on big data analytics using deep learning. In Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications (pp. 615-632). IGI global
19. Iorkyase ET, Tachtatzis C, Glover IA, Lazaridis P, Upton D, Saeed B, Atkinson RC (2019) Improving RF-based partial discharge localization via machine learning ensemble method. IEEE Transactions on Power Delivery 34(4):1478–1489

20. Khalaf M, Alaskar H, Hussain AJ, Baker T, Maamar Z, Buyya R, … Al-Jumeily D (2020) IoT-enabled flood severity prediction via ensemble machine learning models. IEEE Access 8:70375–70386
21. Khalaf, M, Hussain, AJ, Al-Jumeily, D, Baker, T, Keight, R, Lisboa, P, ... and Al Kafri, AS (2018). A data science methodology based on machine learning algorithms for flood severity prediction. In 2018 IEEE Congress on Evolutionary Computation (CEC) (pp. 1–8). IEEE
22. Kinderis, M, Bezbradica, M and Crane, M (2018). Bitcoin currency fluctuation
23. Kishore KVK, Kumar BY, Venkatramaphanikumar S (2020) Optimized water scheduling using IoT sensor data in smart farming, In *Smart Technologies in Data Science and Communication* (pp. 205–219). Springer, Singapore
24. Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D (2018) Machine learning in agriculture: a review. Sensors 18(8):2674
25. Mafarja MM, Mirjalili S (2019) Hybrid binary ant lion optimizer with rough set and approximate entropy reducts for feature selection. Soft Comput 23(15):6249–6265
26. Mahbub M (2020) A smart farming concept based on smart embedded electronics, internet of things and wireless sensor network. Internet of Things 9:100161
27. Moore, J, Chase, JS and Ranganathan, P (2006). Weatherman: automated, online and predictive thermal mapping and management for data centers. In 2006 IEEE international conference on Autonomic Computing (pp. 155-164). IEEE
28. Morais R, Silva N, Mendes J, Adão T, Pádua L, López-Riquelme JA, Pavón-Pulido N, Sousa JJ, Peres E (2019) Mysense: A comprehensive data management environment to improve precision agriculture practices. Comput Electron Agric 162:882–894
29. Muangprathub J, Boonnam N, Kajornkasirat S, Lekbangpong N, Wanichsombat A, Nillaor P (2019) IoT and agriculture data analysis for smart farm. Comput Electron Agric 156:467–474
30. Punn M, Bhalla N (2013) Classification of wheat grains using machine algorithms. International Journal of Science and Research (IJSR) 2(8):363–366
31. Varghese, R and Sharma, S (2018). Affordable smart farming using IoT and machine learning. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 645-650). IEEE
32. Vincent DR, Deepa N, Elavarasan D, Srinivasan K, Chauhdary SH, Iwendi C (2019) Sensors driven AI-based agriculture recommendation model for assessing land suitability. Sensors 19(17):3667
33. Zikria, YB, Afzal, MK and Kim, SW (2020). Internet of multimedia things (IoMT): opportunities, Challenges and Solutions

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Nermeen Gamal Rezk has received her B.Sc from the Department of Computer Science and Engineering, Faculty of Engineering, Kafrelsheikh University, Egypt, in 2009. She received her M.Sc. From the Department of Computer Science and Engineering, Faculty of Engineering, Tanta University, Egypt, in 2017. She is working towards her Ph.D. degree in the Department of Computer Science & Engineering, Faculty of Electronic Engineering, Menoufia University, Egypt. Her research area of interest includes; Data Science and Big Data Analytics, Image Processing, Internet of Things/Nano-Things, Artificial intelligence.

Ezz El-Din Hemdan has received his B.Sc from the Department of Computer Science & Engineering, Faculty of Electronic Engineering, Menoufia University, Egypt, in 2009. He received his M.Sc. From the Department of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufia University, Egypt, in 2013. He received his Ph.D. degree in the Department of Computer Science, Mangalore University, India in 2018. He has several publications in national/international conferences and journals. His research area of interest includes; Canacelable Biometric, Blockchain, Digital Twins, Image Processing, Virtualization, Cloud Computing, Internet of Things/Nano-Things, Cryptography, Data Hiding, Digital Forensics, Cloud Forensics, Big Data Forensics, Data Science and Big Data Analytics.

Abdel-Fattah Attia graduated from Mansoura University in 1991 and obtained his M. Sc. from Ain Shams University, 1997, Cairo, Egypt and the Ph.D. degree in Control Engineering and Robotics from Czech Technical University in Prague, Czech Republic, 2002. He is the Founder and Head of Intelligent Systems Research Group (ISRG) Group & Vice Dean for Graduate Studies & Research at Faculty of Engineering, Kafrelsheikh University. He has more than 50 scientific research papers published in prestigious international journals. He co-supervised 3 Ph.D. and 6 M.Sc. theses at Ain Shams University. His research interests include the development, modeling and identification of nonlinear systems, fuzzy models, neural networks, neuro-fuzzy, optimization and AI, and its application to nonlinear systems

Ayman El-Sayed a Professor and Dean of Electronic Engineering, Menoufia University. He received the B.Sc. degree in computer science and engineering in 1994, the master's degree in computer networks in 2000 from the Menoufia University, Egypt, and the Ph.D. degree in computer network in 2004 from (Institute National De Poly technique De Grenoble) INPG, France. He is specialized in soft computing, algorithms, data structure, and cloud computing and big data analysis. In addition, his interests include multicast routing protocols, application-level multicast techniques, multicast on both mobile network and mobile IP, and image processing techniques. In addition, there are other interesting topics such as bioinformatics, Biocomputing, and bio computer. He is an approved supervisor for M.Sc. and Ph.D. programs in various University. He has completed various project in government and private organization. He has published more than 105 research papers in international Journals and two books about OSPF protocol and multicast protocols. Currently, he is serving as an editorial board member in various international Journals and conferences. He is a senior member of the IEEE.

Mohamed A. El-Rashidy obtained his B.Sc. degree in computer science and engineering in 2005 and M.Sc. degree in computer science and engineering in 2008 from the Menoufia University, Egypt. He received Ph.D. degree in data science and machine learning, 2012. Currently, he is working as assistant professor in the Dept. of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufia University, Egypt. His main research interest's Data science and machine learning, pattern recognition, web mining, business intelligent, image processing, data privacy, big data analysis and cloud computing.

Affiliations

Nermeen Gamal Rezk[1] **· Ezz El-Din Hemdan**[2] **· Abdel-Fattah Attia**[1] **· Ayman El-Sayed**[2] **· Mohamed A. El-Rashidy**[2]

Nermeen Gamal Rezk
nermeenbarakaa@yahoo.com

Abdel-Fattah Attia
aheliel@eng.kfs.edu.eg

Ayman El-Sayed
ayman.elsayed@el-eng.menofia.edu.eg

Mohamed A. El-Rashidy
mohamed.elrashedy@el-eng.menofia.edu.eg

[1] Department of Computer Science and Engineering, Faculty of Engineering, Kafrelsheikh University, Kafrelsheikh, Egypt

[2] Department of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufia University, Menoufia, Egypt