LEX LOCALIS-JOURNAL OF LOCAL SELF-GOVERNMENT
ISSN:1581-5374 E-ISSN:1855-363X
VOL. 24, NO. 2(2026)

OPTIMIZED DECISION TREE ALGORITHM WITH FRA FOR CROP YIELD PREDICTION IN IOT BASED AGRICULTURE

M. Sheerin Banu[1*]

[1*]Professor and Head, Department of Information Technology, R.M.K. Engineering College, India.

Corresponding Author: hod.it@rmkec.ac.in[1]

Abstract
In Internet of Things (IoT) based agricultural environment, crop yield prediction is one of the most desirable but difficult tasks for any country. Farmers are having troubles to achieve a good yield from crops because of unpredictable changes in the climate. To feed India's growing population, agriculture must incorporate cutting-edge technology and tools. Thus, machine learning techniques have been used widely to predict the crop yield. However, this study focuses to enhance the prediction accuracy using enhanced machine learning technique. Namely, for efficient crop yield prediction, an optimized decision tree (DT) classifier with fuzzy ranking algorithm (FRA) is used. At first, FRA is presented to choose the features set from the crop yield dataset. Using the selected features, crop yield is predicted by presenting optimized DT. Using a modified penguin search algorithm (MPSA), the decision node of the DT is optimally chosen. The proposed prediction model improves the prediction accuracy.

Keywords: Optimized DT, FRA, penguin search algorithm, crop yield prediction

1. Introduction

Agriculture has the potential to be a major source of income in the Indian economy. With India's population expected to rise to 1.5 billion by 2050, agricultural production optimization and food contribution chains are essential for more efficiently producing and delivering food, fibre, and fuel to meet rising demand. As a result, improved crop yield is required in the future. Generally, farmers must forecast crop yields in the future in order to maximise crop yield [1-3]. Analysis may also be important to help farmers in utilising full capacity in crop production. Nevertheless, because of climate change and urbanisation, this type of farmer's goal will be difficult to achieve. Risks and improbability have frequently occurred in agriculture as a result of erratic weather. Environmental conditions, input levels, lack of consistency in soil, mixture, and product prices have increased the importance of farmers using information and seeking assistance when making major farming decisions [5-7].

Accurate yield prediction for the various crops involved in the planning process can be a critical issue for agricultural future planning. To find practical and effective solutions to this problem, data mining methods were used [8] [9]. Sensor technologies or IoT, for example, were used to gather data from agriculture in order to calculate, forecast, or monitor damage [10] [11]. A forecasting evaluation, crop, integrated soil, and weather can evaluate implications such as crop yields and food uncertainty through using historical agricultural datasets. Predictive analytics is also used to improve the decision-making system for crop yield prediction. Currently, yield prediction can be a significant agricultural challenge.

AI and machine learning have an important effect on the quickly rising agricultural sector [12] [13]. Without human intervention, multiple supervised and unsupervised machine learning algorithms participate in crop analysis, analysis, and decision making. Thus, in this work, to enhance the prediction accuracy of crop yield, an enhanced machine learning technique is presented. Namely, the contributions are given as follows,

- To choose the efficient features set, FRA is used.

- To predict the crop yield, an optimized DT is presented.
- To improve the performance of DT, MPSA algorithm is used.

The upcoming sections are arranged as follows. Recent literatures which related to crop yield prediction are reviewed in section 2. The proposed crop yield prediction based on the optimized DT with FRA is proposed in section 3. Section 4 examines the findings. Section 5 brings the work to a close.

2. Related works

Recent literatures which focused research on crop yield prediction are reviewed in this section. The prediction data is clustered using a variety of machine learning algorithms in order to forecast crop productivity. However, the clustering is poorly accurate and of low grade. P. Suvitha Vani and S. Rathi [14] proposed a Proximity Likelihood Maximization Data Clustering (PLMDC) technique to increase clustering accuracy with less complexity and improve the accuracy of crop yield forecast for farmers. Using a logical linear regression model, superfluous data was removed from agricultural data in the approach. Following that, the suggested clustering algorithm was applied based on Manhattan distance weights and similarity. The genetic algorithm (GA), which has a good fitness function, was used to select the features from the clustered data. Lastly, the Apriori and Frequent Pattern (A-FP) growth algorithm computed the decision support system to forecast crop yields based on their chosen features. By presenting these schemes, the authors achieved better clustering accuracy.

N. R. Prasad, N R Patel and Abhishek Danodia [15] employed the R package to predict cotton production in Maharashtra at three different intervals prior to the actual harvest using a machine learning-based random forest (RF) method. For the purpose of calibrating and validating the RF model, long-term agromet-spectral variables defined from multi-sensor satellites with actual crop output from 2001-2017 were used. With the main affecting variables, the RF model's performance was proven to be fast and trustworthy in predicting crop production. Results demonstrated that the RF algorithm is capable of integrating and processing a huge number of inputs generated from various satellite modalities.

A nutrient insufficiency analysis is required to ensure a high yield. Crop yield is determined by nutrient content, which has a significant impact on crop health. Thus, Sushila Shidnal, Mrityunjaya V. Latte and Ayush Kapoor [16] considered a paddy crop's nutrient deficiency. The authors used Tensor Flow to create a neural network that classified them as having nitrogen, potassium, or phosphorus deficiencies or being healthy. The content of nitrogen, potassium, and phosphorous were balanced optimally using this Tensor Flow. Tensor Flow's model identified the flaw by analysing a series of images. The outcome was fed into a "machine learning driven layer," which estimated the level of deficiency quantitatively. It employed the k means clustering algorithm in particular. It executed via the rule matrix to calculate the cropland yield. Due to the proposed scheme, the authors achieved prediction accuracy of 77%.

To feed India's growing population, the agricultural sector must incorporate cutting-edge technology and tools. Thus, Ekaansh Khosla, Ramesh Dharavath and Rashmi Priya [17] focused on the forecasting of major kharif crops. Because rainfall is the most significant feature to determine the amount of kharif crop production, using modular artificial neural networks (MANNs), the authors first forecast the proportion of monsoon rainfall. Then, using rainfall data and crop area, they used support vector regression to forecast the proportion of significant kharif crops. The article's findings showed that the proposed schemes improved prediction accuracy.

To solve the issue that the actual model of yield prediction for different crops such as wheat and rice has less accuracy, Li Tian et al [18] proposed the ecological distance algorithm based yield prediction model of different crops. The parameters such as weather, soil, and behavioural were acquired as significant parameters through the design process. To create a yield prediction model for both crops, the ecological distance algorithm was jointed with crop yield predictors. The proposed scheme achieved higher accuracy of prediction.

P.S. Maya Gopal and R. Bhargavi [19] presented hybrid Multiple Linear Regression and ANN (MLR-ANN) for the prediction of crop yield. The authors used the proposed model to examine the accuracy of prediction when MLR intercept and coefficients were used to initialise the weights and bias of input layer of ANN. For accurate paddy crop yield prediction, a Feed Forward ANN with Back Propagation training algorithm was applied. Instead of random weights and bias in the model, the weights and bias of input layer are initialised by the coefficients and bias of MLR. By presenting this proposed model, the authors attained better accuracy.

3. Optimized DT Algorithm with FRA for Crop Yield Prediction

3.1. Overview

Figure 1 depicts the overall structure of the proposed approach. The proposed approach consists of two stages namely, pre-processing, feature selection and classification. In this, the dataset records may be incomplete, noisy or duplicate. This will affect the classification accuracy. So, in pre-processing stage, we eliminate the duplicate, incomplete and redundant data. Then, the FRA is used to choose the best feature subset. The chosen features are then fed into the presented prediction model. An optimised DT structure is proposed for crop yield prediction. The MPSA algorithm is used to select the optimal decision node or upper level of the tree in the proposed structure. The optimised DT method determines crop yield prediction that uses the subset of features.

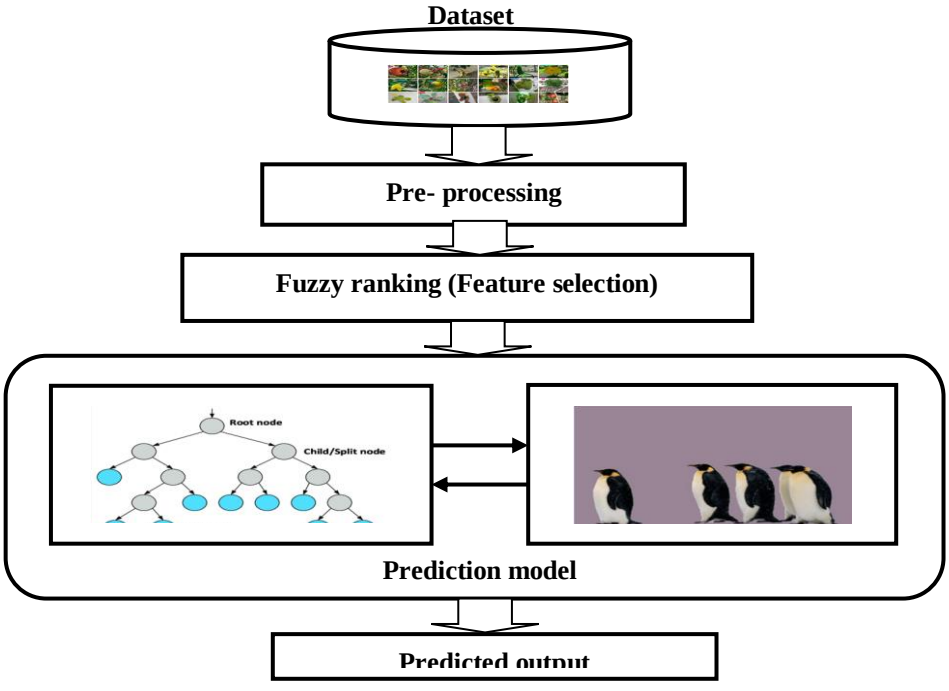


Figure 1: The overall structure of the proposed approach

3.2. Feature Selection using FRA

The FRA is employed in this approach to eliminate feature redundancy from the dataset and to decrease computational complexity. FRA selects the optimal features subset.

Assume that the dataset (D) has 'n' features and is defined as follows,

$$D = \{f_1, f_2, ..., f_n\} \quad (1)$$

Where, f_n denotes the n^{th} feature.

The major aim of this algorithm is to choose the best optimal feature subset (D') which includes the most appropriate features f_i'. Thus, to attain this best feature set, this algorithm employs the Maximum Relevance (maxR) and Minimum Redundancy (minR) methods.

To begin, the features with the maxR to the target class (T) are chosen by estimating the maxR of features to the T. Equation (2) defines the computation of max R (R).

$$R = \frac{1}{|D|} \sum_{f_i \in D} I(f_i; T) \quad (2)$$

Here, $I(f_i; T)$ is the mutual information between the T and the feature.

Besides, in this algorithm, fuzzy features are used to select the features. Equation (3) defines the computation of fuzzy mutual information among fuzzy features.

$$FI(X,Y) = E(X) + E(Y) - E(X,Y) \quad (3)$$

Here, X and Y denote the fuzzy variables; E (X) and E (Y) denote the fuzzy entropy values of X and Y respectively. E (X, Y) is the fuzzy joint entropy of X and Y.

The feature set chosen with maxR may include redundancy features. To lessen these redundancy features, minR is combined with maxR. Equation (4) defines the estimation of minR (mR) between the selected features.

$$mR = \frac{1}{|D|^2} \sum_{f_i, f_j \in D} FI(f_i, f_j) \quad (4)$$

The selected features are then given an mRMR score. This is the distinction between R and mR. This is how it is defined:

$$\max \Theta(R, mR), \quad \Theta(R - mR) \quad (5)$$

As per (5), a feature is selected as a candidate feature when it has the highest mRMR score.

At final, fuzzy rank is calculated and is assigned to the selected features. The estimation of fuzzy ranking is described as,

$$Rank_{Fuzzy} = \arg\max\left(\frac{R}{mR}\right) \quad (6)$$

3.3. Crop yield prediction using seagull optimization based DT

The proposed prediction model uses the ideally chosen features as input features. This approach presents an optimised DT for crop yield prediction. The MPSA is used to optimise DT performance.

***DT*:** It is a structured supervised classifier. The main advantages of this DT are that it requires less data to train the structure and offers the best performance evaluation. A tree is built using a divide and conquers strategy in this classifier. DT builds a tree based on the training samples (TSs). The TS includes the features with the values obtained from the class. The classifier calculates the frequency of each class in the TS. When all features have the same class, a decision node is created with that class. Otherwise, if the TS contains features from more than one class, splitting criteria are used to select the best feature. Furthermore, to construct a tree structure, this DT algorithm employs a recursive scheme.

Construction of DT

Build_Tree (TS, D')

1. Consider a decision node (dn).
2. If all of the data from the TSs contain the same class C_i, make dn as a leaf node.
3. If the feature set D' is empty, think of dn as a leaf node with the most common class C_i.
4. Select feature f_j from feature set D' with the greatest amount of information gain and label dn with feature f_j.
5. For every value (v) of feature f_j:
    a. Make a division with the state f_j=v from dn.
    b. Consider the TS subset with f_j=v..
    c. If the sample subset is empty, assign f_j to the leaf node which has the most common class. Or else, associate node produced by Build Tree. (TS, D')

Although the DT produces better results, its accuracy and size could be improved further. Furthermore, over-fitting by creating globally optimised DTs should be avoided in order to improve DT performance. As a result, the MPSA algorithm is presented to address these issues. In DT, this algorithm chooses the best features or non-leaf nodes. As explained below, the MPSA algorithm is used to select a set of nodes or features:

***MPSA*:** PSA was inspired by the penguin's foraging behaviour. Penguins are seabirds that cannot fly because they have adapted to aquatic life. PSA is a new demographic and memory-based metamorphic method. PSA was created in 2013 by Youcef Gheraibia and Abdelouahab Moussaoui [20]. Penguins can stay underwater for up to twenty minutes in order to dive deeper. Furthermore, penguins can dive over 520 meters to search water for food. While swimming under water is higher performance and less fatigue than skiing, they should come up consistently every two minutes for air. When they lower their heart rate and keep their eyes open to look for food, they can breathe faster. Their retinas are capable of distinguishing shapes and colors. Penguins eat krill, fish, squid, and crustaceans. To dive deeper and longer, it takes more energy, so they have to eat more food. The penguin population detects early stages, with each penguin swimming submerged to hunt fish when ingesting its oxygen reserves. There are various types of contact between penguins are performed from time to time and the amount of fish they eat increases. Each penguin group searches for fish until their oxygen reserves are depleted, and each group forms into a certain number of groups. The penguins try to hunt the maximum number of fish around the allotted area during the foraging phase. To enhance the performance of PSA algorithm, OBL method is added with PSA. OBL improves the searching ability and decreases the processing time

Summarizing the observations from the penguin's foraging behaviour, the below listed rules are provided;

Rule 1: There are various groups of penguins within a population. Every group has several penguins, which vary depending on the availability of food in the relevant food area.

Rule 2: A group of penguins begins searching for food at a certain depth underneath the water based on information about the energy gain and the cost of acquiring it.

Rule 3: As a group, they feed and follow their guide, which has more food on the last dive. Penguins will hunt for food until their oxygen reserves run out.

Rule 4: Following several dives, the penguins return to the surface to inform their local subsidiaries of location, food resources, and intra-group communication.

Rule 5: If food maintain is low for a given group of penguins to survive, one piece of the group will move to a new place through interaction between the groups.

The overall process steps of the MPSA algorithm is outlined below figure (2);

Listed below steps are involved in the proposed MPSA;

Step 1: Initialization: Here, The solutions are the set of dns or features. These are initially generated at random. Each non-leaf node is referred to as a penguin, and solutions are referred to as agents. Every penguin's position is updated utilising fitness function. Equation (7) provides the initial solution format.

$$S_M = \{S_{i1}, S_{i2}, ..., S_{iD}\} \tag{7}$$

$$S_{iD} = \{dn_1, dn_2, ..., dn_n\}_{iD} \tag{8}$$

Here, $S_{i,D}$ is the position of i^{th} penguin in D^{th} dimension.

Step 2: OBL: An opposite solution is estimated for each solution in this phase. It is described as follows,

$$\overline{S} = a + b - S \tag{9}$$

Where, $S \in [a,b]$ is a real number

Step 3: Fitness evaluation: After initialization, the fitness value for each solution is computed. The gain ratio is used to calculate fitness for every solution or feature from D'. The information gain is used to calculate the gain ratio of every feature. The TS's information gain is given by the equation (10):

$$\inf(TS) = -\sum_{i=1}^{n} pr_i * \log_2(pr_i) \tag{10}$$

Here, pr_i is the feature probability belongs to class C_i and is described by (11).

$$pr_i = \frac{frq(TS, C_i)}{|TSs|} \tag{11}$$

Here, $|TSs|$ is the whole amount of features from the TSs and $frq(TSs, C_i)$ is the whole amount of features belongs to class C_i.

The TSs are divided into q partitions based on the domain values of a non-class attribute f_i. The information gain of the splitting process is calculated using Equation (12).

$$\inf\left(f_i, TSs\right) = -\sum_{j=1}^{q} \frac{|TSs(j)|}{|TSs|} \inf\left(TSs(j)\right) \quad (12)$$

The information gain of feature f_i is determined by (10) and (12) and is described in equations (13).

$$gain(f_i) = \inf\left(TS\right) - \inf\left(f_i, TSs\right) \quad (13)$$

The feature f_i gain ratio is determined as follows:

$$gain\ ratio(f_i) = \frac{gain(f_i)}{split\ \inf\left(f_i\right)} \quad (14)$$

Here, $split\ \inf\left(f_i\right)$ is the information gain obtained by dividing the TSs into q subset on test feature f_i and is described in (15).

$$split\ \inf\left(f_i\right) = \sum_{j=1}^{q} \frac{|TSs(j)|}{|TSs|} * \log_2\left(\frac{|TSs(j)|}{|TSs|}\right) \quad (15)$$

Thus, the fitness of each solution is estimated as follows,

$$F_i = \frac{1}{1 + gain\ ratio(f_i)} \quad (16)$$

$$Fitness = \min(F_i) \quad (17)$$

According to equation (17), the optimal node for data splitting is the node or feature with the lowest fitness. If the best solution is not found, the solution is updated as follow.

Step 4: Update the solution: The location of the penguin is updated using the below foraging behaviour.

Swimming course update: The penguin swims to a new location at time $t+1$ in the whole solution space is expressed in the following equation (18);

$$S_j^i(t+1) = S_j^i(t) + O_j^i(t) \times Rand() \times \left(S_{localbest}^i - S_j^i(t)\right) \quad (18)$$

Where, $S_j^i(t+1)$ is the updated location of the penguin, $S_j^i(t)$ is the old location of the penguin, $S_{localbest}^i$ is the local best location, $O_j^i(t)$ is the oxygen reserve of the j[th] penguin of the i[th] group and $Rand()$ is represented is a random number [0, 1].

Oxygen reserve update: The oxygen reserve of each penguin is updated after each dive. It is expressed in the following equation (19);

$$\left| O_j^i(t+1) = O_j^i(t) + \left(F\left(S_j^i(t+1)\right) - F\left(S_j^i(t)\right)\right) \times \left| S_j^i(t+1) + S_j^i(t) \right| \right|$$

(19)

Where, $O_j^i(t+1)$ is the new oxygen reserve and F is represented as the objective function. If the new solution is superior to the previous one, the oxygen balance improves. If the new solution is bad, the oxygen balance decreases.

Food abundance update: In addition, the quantity of eaten fish (QEF) and the penguin group membership were also updated and it is shown in the following equation (20);

$$QEF^i(t+1) = QEF^i(t) + \sum_{j=1}^{d_i} \left(O_j^i(t+1) - O_j^i(t)\right)$$

(20)

Where, $QEF^i(t)$ is represents the amount of eaten fish.

A large QEF value means that the area provides adequate food for the entire group and demands migratory penguins from other groups.

Group membership update: Each penguin updates its membership according to the degree of food abundance in various groups. The value of the membership process for joining the group i is a probability given as the following equation (21);

$$R_i(t+1) = \frac{QEF^i(t)}{\sum_{j=1}^{k} QEF^j(t)}$$

(21)

Step 5: Termination criteria: The MPSA is terminated when the best decision nodes or features are obtained for the satisfaction of a termination criterion. Once the solution is obtained, the algorithm will be terminated. The overall algorithm of the MPSA is depicted in figure 2;

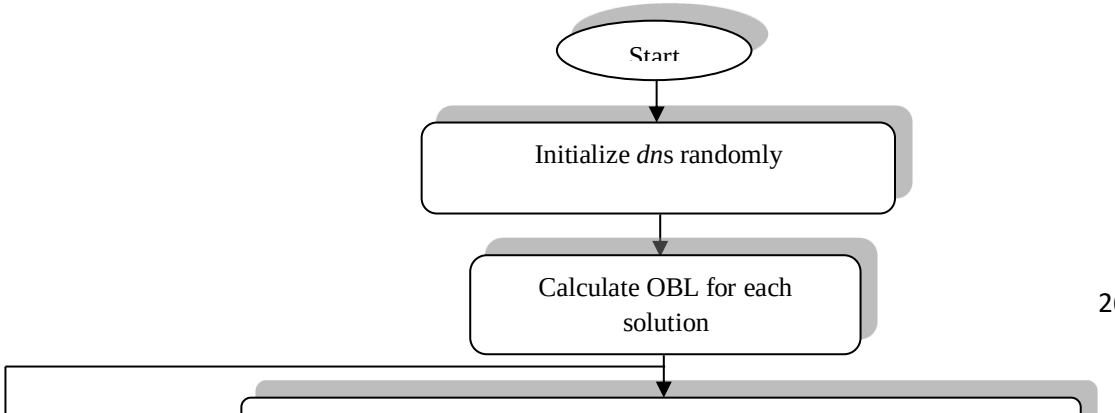


Figure 2: Flowchart of proposed MPSA method

3.4. Crop yield prediction

Following DT optimization, the model is trained using features selected from the 80% training dataset. The trained model is fed 20% of the test dataset's optimal features. The trained model generates the output score value according to the selected features (O^{Score}). The yield is predicted according to the score value.

4. Results and discussion

This section analyses the proposed prediction model's results. This model is powered by an Intel Core i5 processor and a PC with 6GB of RAM that runs Windows 10. The proposed method is being simulated using Phython.

4.1. Dataset description

Data is collected from various agricultural lands in Tamil Nadu for experimentation. A total of 7844 data points are collected, with each data point containing twelve features. This dataset contains information on seventeen different types of crops. For yield prediction also, we are collect the data from agricultural land. The major features such as season, area, temperature, pH, average rainfall, nitrogen, phosphorous and potash are selected to predict crop yield. Figure 3 illustrates the yield prediction of the crop maize.

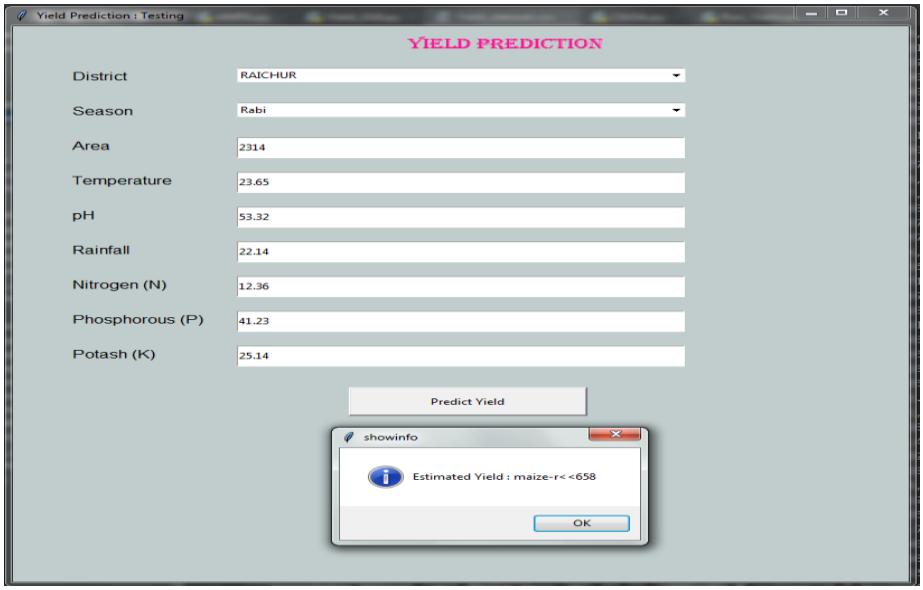


Figure 3: Screen shot for crop prediction (recommended for maize)

4.2. Performance analysis

The performance of different crop yield prediction models such as MPSA-DT, DT, SVM and NB is analysed in this section. Besides, the execution is analysed using accuracy, precision, recall and F-measure. The following sections evaluate the performance of different yield prediction models for the crops sunflower, soybean, mustard and maize.

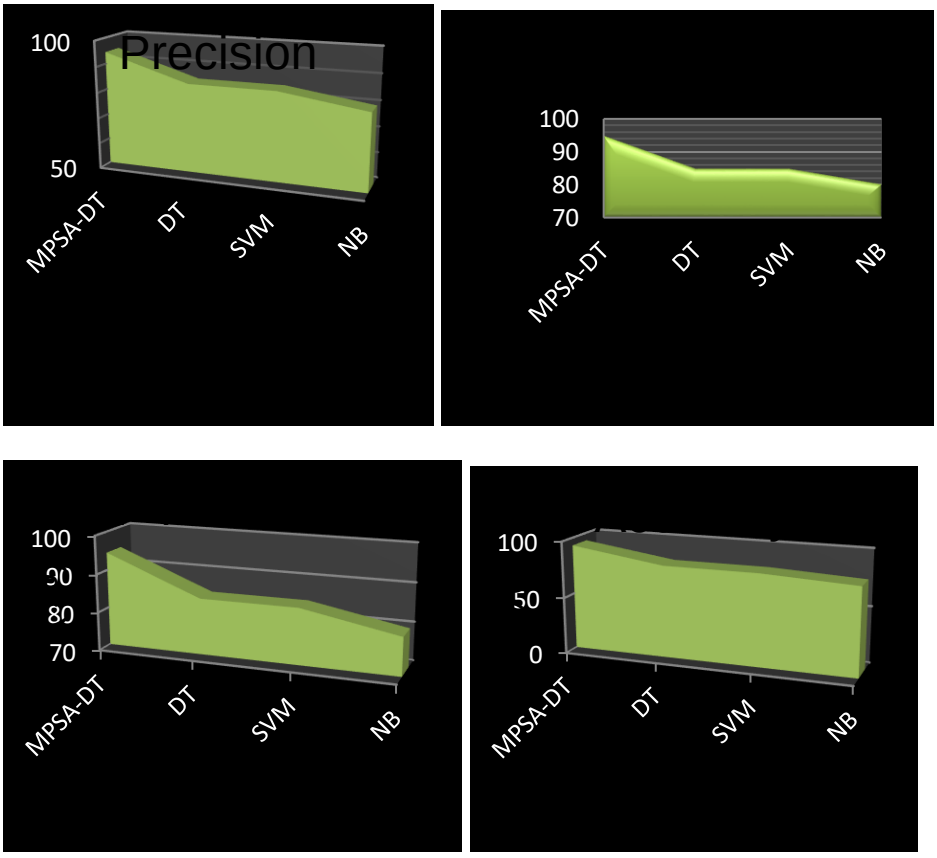
4.2.1. Yield prediction of sunflower

In this section, yield prediction of sunflower is analysed. Table 1 and figure 4 illustrate the evaluation of different prediction models for sunflower yield prediction. As illustrated in the table and figure, the conventional DT and SVM attained similar accuracy of 81.3% while the NB obtained the accuracy of 77.4%. As the performance of DT is improved by optimizing the decision nodes of it using MPSA algorithm, the accuracy of the MPSA-DT is increased to 93.2% than the conventional prediction models.

Table 1: The evaluation of different prediction models for sunflower yield prediction

Metrics / Models	Precision	Recall	F-measure	Accuracy
MPSA-DT	94.5	94.63	94.57	93.2
DT	84.57	84.62	84.58	81.3
SVM	84.57	84.62	84.58	81.3
NB	79.8	79.92	79.87	77.4

Figure 4: Graphical representation of the evaluation of different prediction models for sunflower yield prediction



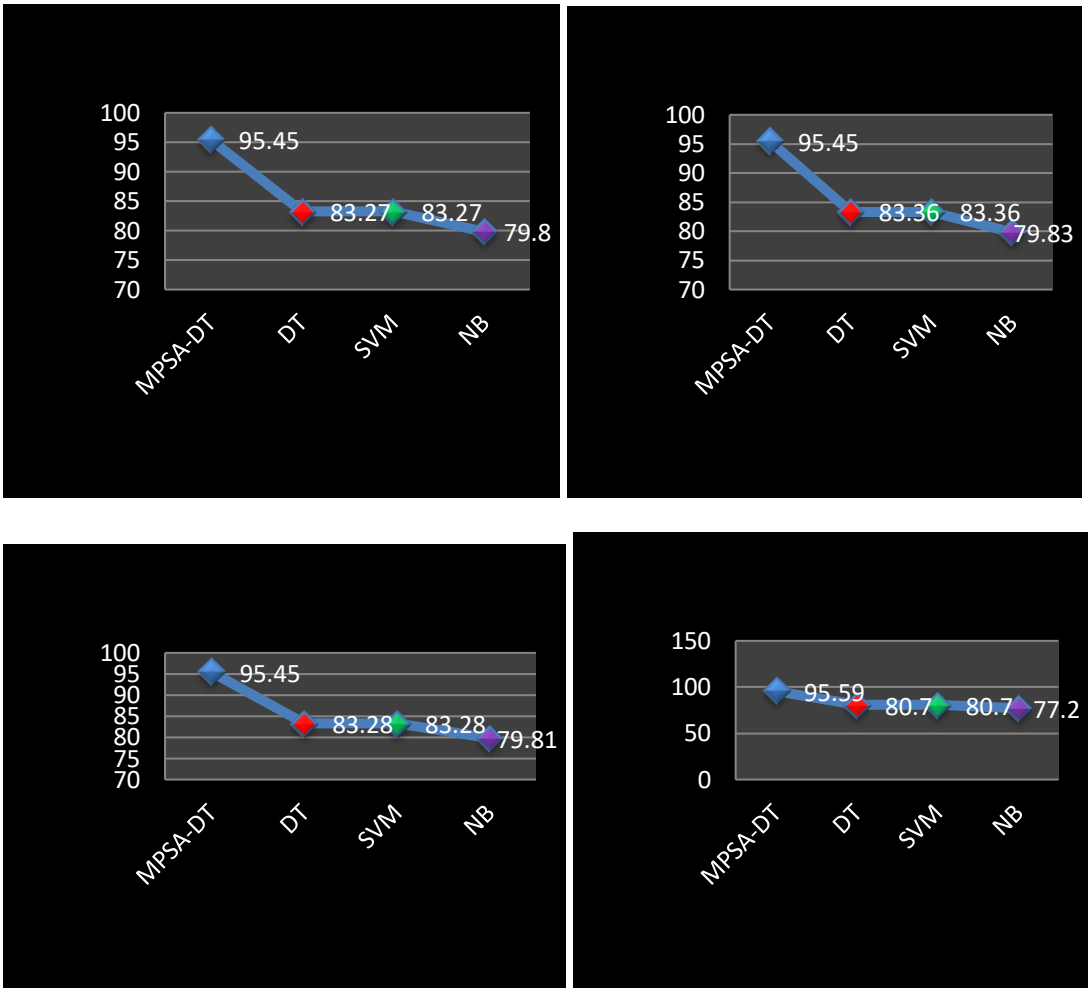
4.2.2. Yield prediction for Soybean

In this section, yield prediction of soybean is analysed. The evaluation of different prediction models for soybean yield prediction is illustrated in table 2 and figure 5. As depicted in the figure, the proposed MPSA-DT based prediction model attained F-measure of 95.45% while the conventional prediction models DT, SVM and NB obtained F-measure of 83.28%, 83.28% and 79.81% respectively. Likewise, compared to the conventional prediction models, the proposed MPSA-DT attained better accuracy of 95.59%.

Table 2: The evaluation of different prediction models for soybean yield prediction

Metrics / Models	Precision	Recall	F-measure	Accuracy
MPSA-DT	95.45	95.45	95.45	95.59
DT	83.27	83.36	83.28	80.7
SVM	83.27	83.36	83.28	80.7
NB	79.8	79.83	79.81	77.2

Figure 5: Graphical representation of the evaluation of different prediction models for soybean yield prediction



4.2.3. Yield prediction for Mustard

In this section, yield prediction of mustard is analysed. Table 3 and figure 6 illustrate the evaluation of different prediction models for mustard yield prediction. Compared to DT, SVM and NB, F-measure of MPSA-DT is increased to 12%, 12% and 18% respectively. Likewise, accuracy of MPSA-DT is increased to 17% , 17% and 22% than that of DT, SVM and NB respectively.

Table 3: The evaluation of different prediction models for mustard yield prediction

Metrics / Models	Precision	Recall	F-measure	Accuracy
MPSA-DT	94.54	94.54	94.54	94.59
DT	84.48	84.73	84.49	81.2
SVM	84.48	84.73	84.49	81.2
NB	79.72	79.83	79.87	77.3

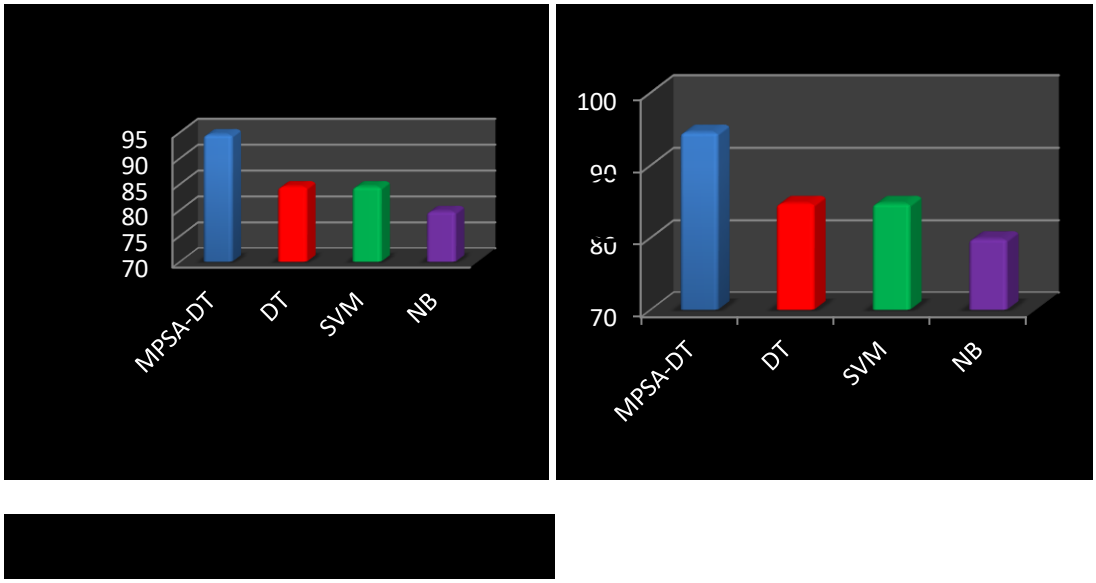


Figure 6: Graphical representation of the evaluation of different prediction models for mustard yield prediction

4.2.4. Yield prediction for Maize

In this section, yield prediction of maize is analysed. Table 4 and figure 7 illustrate the evaluation of different prediction models for maize yield prediction. 7As depicted in the figure, the proposed MPSA-DT based prediction model attained F-measure of 95.7% while the conventional prediction models DT, SVM and NB obtained F-measure of 83.39%, 83.39% and 79.92% respectively. Likewise, compared to the conventional prediction models, the proposed MPSA-DT is increased to 19%, 19% and 24% than that of DT, SVM and NB respectively. .

Table 4: The evaluation of different prediction models for maize yield prediction

Metrics / Models	Precision	Recall	F-measure	Accuracy
MPSA-DT	95.7	95.7	95.7	95.79
DT	83.38	83.47	83.39	80.8
SVM	83.38	83.47	83.39	80.8
NB	79.9	79.94	79.92	77.3

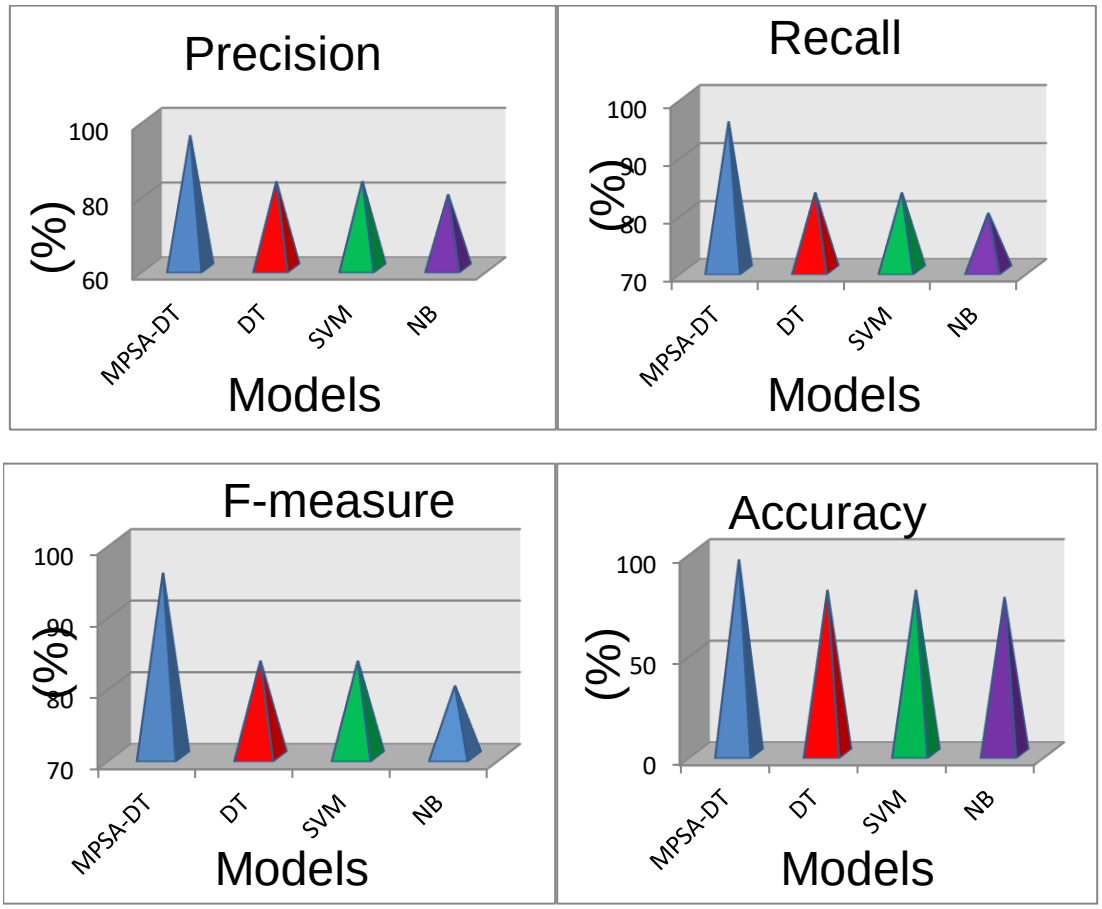


Figure 7: Graphical representation of the evaluation of different prediction models for maize yield prediction

5. Conclusion

To enhance the prediction accuracy of crop yield in IoT based agricultural environment, an optimized DT classifier with FRA has been presented. Feature redundancy in the dataset has been reduced by selecting the optimal features using FRA. These chosen features are fed into the prediction model. The optimized prediction model where decision nodes are selected optimally using MPSA algorithm predicted the yield of various crops. From the simulation results, the proposed MPSA-DT based crop yield prediction model attained accuracy of 93.2% for sunflower, 95.59% for soybean, 94.59 for mustard and 95.79 for maize.

Declarations

Funding

The authors declare that they have no competing interests and funding

Conflict of Interests

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Availability of data and material

Data sharing is not applicable to this article because of proprietary nature.

Data availability statement

Not Applicable

Code Availability

Code sharing is not applicable to this article because of proprietary nature.

Authors' contributions

All authors read and approved the final manuscript

References

[1] Chlingaryan, Anna, Salah Sukkarieh, and Brett Whelan. "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review." Computers and electronics in agriculture 151 (2018): 61-69.
[2] M. TAMIL SELVI, B. JAIS ON: Lemuria: A Novel Future Crop Prediction Algorithm Using Data Mining, The Computer Journal, (2020).
[3] A. GONZÁLEZ SÁNCHEZ, J. FRAU STO SO LÍS, W. OJEDA BUSTAMANTE: Predictive ability of machine learning methods for massive crop yield prediction, (2014).
[4] Rajshekhar Borate., "Applying Data Mining Techniques to Predict Annual Yield of Major Crops and Recommend Planting Different Crops in Different Districts in India", International Journal of Novel Research in Computer Science and Software Engineering,April 2016.
[5] D Ramesh, B Vishnu Vardhan, "Analysis of Crop Yield Prediction using Data Mining Techniques", International Journal of Research in Engineering and Technology (IJRET),Vol.4, 2015.
[6] Ramesh A. Medar and Vijay. S. Rajpurohit "A Survey of data mining techniques for crop yield prediction", IJARCSMS, Volume 2, Issue 9, September 2014.
[7] P.Surya, Dr.I.Laurence Aroquiaraj, "Crop yield prediction in agriculture using data mining predictive analytic techniques", International Journal of Research and Analytical Reviews(IJRAR),2018.
[8] Mokaya and Victor, "Future of precision agriculture in india using machine learning and artificial intelligence." Int. J. Comput. Sci. Eng 7, no. 2 (2019): 1020-1023.
[9] Suresh, A., N. Manjunathan, P. Rajesh, and E. Thangadurai. "Crop Yield Prediction Using Linear Support Vector Machine." European Journal of Molecular & Clinical Medicine 7, no. 6 (2020): 2189-2195.
[10] Dahikar, Snehal S., Sandeep V. Rode, and PramodDeshmukh. "An artificial neural network approach for agricultural crop yield prediction based on various parameters." International Journal of Advanced Research in Electronics and Communication Engineering 4, no. 1 (2015): 94-98.
[11] Parida, BikashRanjan, Amritesh Kumar, and Avinash Kumar Ranjan. "Crop Types Discrimination and Yield Prediction Using Sentinel-2 Data and AquaCrop Model in Hazaribagh District, Jharkhand." KN-Journal of Cartography and Geographic Information (2021): 1-13.
[12] Rezk, Nermeen Gamal, Ezz El-Din Hemdan, Abdel-Fattah Attia, Ayman El-Sayed, and Mohamed A. El-Rashidy. "An efficient iot based smart farming system using machine learning algorithms." Multimedia Tools and Applications 80, no. 1 (2021): 773-797.

[13] Elavarasan, Dhivya, and PM Durai Raj Vincent. "A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters." Journal of Ambient Intelligence and Humanized Computing (2021): 1-14.
[14] Vani, P. Suvitha, and S. Rathi. "Improved data clustering methods and integrated A-FP algorithm for crop yield prediction." Distributed and Parallel Databases (2021): 1-15.
[15] Prasad, N. R., N. R. Patel, and Abhishek Danodia. "Crop yield prediction in cotton for regional level using random forest approach." Spatial Information Research 29, no. 2 (2021): 195-206.
[16] Shidnal, Sushila, Mrityunjaya V. Latte, and Ayush Kapoor. "Crop yield prediction: two-tiered machine learning model approach." International Journal of Information Technology 13, no. 5 (2021): 1983-1991.
[17] Khosla, Ekaansh, Ramesh Dharavath, and Rashmi Priya. "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression." Environment, Development and Sustainability 22, no. 6 (2020): 5687-5708.
[18] Tian, Li, Chun Wang, Hailiang Li, and Haitian Sun. "Yield prediction model of rice and wheat crops based on ecological distance algorithm." Environmental Technology & Innovation 20 (2020): 101132.
[19] Gopal, PS Maya, and R. Bhargavi. "A novel approach for efficient crop yield prediction." Computers and Electronics in Agriculture 165 (2019): 104968.
[20] Gheraibia, Youcef, and Abdelouahab Moussaoui, "Penguins search optimization algorithm (PeSOA)," In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, pp. 222-231, Springer, Berlin, Heidelberg, 2013.