Journal of Information Systems Engineering and Management
2025, 10(15s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Multivariate Piecewise Rand Divergencive Cockroach Swarm Optimization for Agriculture Crops Recommendation

Kanuprasad. M. K[1*], Dr.N.V. Balaji[2]
[1]Research scholar, Karpagam Academy of Higher Education,Coimbatore - 641 021, Tamil Email:Prasadkanu32@gmail.com*
[2]Professor and Dean,Faculty of Arts, Science, Commerce and Managagement,Karpagam Academy of Higher Education, Coimbatore Email: balajiresearch261@gmail.com

ARTICLE INFO	ABSTRACT

Received: 02 Dec 2024
Revised: 25 Jan 2025
Accepted: 02 Feb 2025

Precision agriculture focuses on monitoring, management information scheme, and variable rate technologies in cropping systems. The primary advantages of precision agriculture results in enhancing crop recommendations, with optimal soil and environmental impact. Crop recommendation, predicting the most suitable crops for a specific region or farm based on factors such as soil type, weather conditions (including temperature, soil pH, and rainfall), is a challenging task in agriculture. Various approaches have been developed for predicting crop recommendations; however, achieving essential factors for timely and accurate detection remains difficult. In this paper, we introduce a novel technique called Multivariate Piecewise Rand Divergencive Cockroach Swarm Optimization (MPRDCSO) for accurate crop recommendation with minimum time consumption. The proposed MPRDCSO technique begins by gathering information from the crop recommendation dataset and consists of two significant steps: data preprocessing and feature selection. In the first step, data preprocessing is carried out to clean and transform input data using Multivariate Piecewise Constant Weighted Interpolation and Camargo's index-based preprocessing for accurate crop recommendation. Following this, the feature extraction process is executed by applying Rand Indexive Jensen–Shannon Divergenced Cockroach Swarm Optimization. In the feature extraction step, the number of features is collected from the dataset, and the rand similarity between the features is measured in the fitness measure to identify the most relevant features. With the extracted optimal attributes, crop recommendation is performed with higher accuracy. An experimental assessment of proposed technique is conducted with accuracy, precision, recall, F1-score, and crop recommendation time across different instances. The quantitatively discussed results indicate that the performance of the proposed MPRDCSO technique achieves higher accuracy with a reduced processing time compared to conventional methods.

Keywords: Crops Recommendation, preprocessing, Multivariate piecewise constant weighted interpolation, Camargo's index, feature extraction, Rand indexive Jensen–Shannon divergenced cockroach swarm optimization

1. INTRODUCTION

Agriculture plays a pivotal task in addressing worldwide challenges associated to food safety, environmental sustainability, and economic development. Research in agriculture is deals with multifaceted challenges that extend beyond traditional farming practices. Climate change, resource constraints, and soil nutrients are crucial factors for improving plant growth and crop production. To achieve precision agriculture, a crop recommendation system is a technology-driven solution that helps farmers make informed decisions about suitable crops to cultivate, considering various factors such as climate, soil conditions, and historical data. ML-based fertilizer recommendation methodology has been developed for enhancing crop yields.

An Improved Distribution-based Chicken Swarm Optimization with Weight-based Long Short-Term Memory (IDCSO-WLSTM) was developed in [1] for crop predictions and recommendations based on significant features. However, it failed to accurately detect crop predictions when dealing through huge number of attributes. A deep reinforcement learning (DRL) method was developed in [2] for crop prediction to enhance the performance of the crop recommendation system. However, the optimal feature selection in the crop classification system posed a major challenge.

ML-based fertilizer recommendation scheme was introduced in [3], focusing on real-time soil fertility context. However, complexity of fertilizer recommendation system was not reduced. An Agriculture Cultivation Recommender and Smart Irrigation approach were developed in [4] to assist farmers for increasing crop output. But, achieving accurate cultivation recommendations proved challenging issue.

A nutrient recommendation method was developed in [5] using an improved genetic algorithm (IGA) to optimize parameters for achieving maximum yield. However, it failed to reduce computational resources and was challenging to enhance recommendations for crop maintenance. The Modified Recursive Feature Elimination (MRFE) method was introduced in [6], aims to enhance crop prediction accuracy by selecting and ranking salient features. However, it faces challenges in handling large feature sets.

An IoT-enabled soil nutrient classification method was introduced in [7] for accurate crop recommendations, aiming to minimize the utilization of fertilizers in the soil and consequently improve productivity. However, the collection of datasets did not include a diverse range of crops. Novel machine learning techniques were developed in [8] for crop recommendation based on precise analysis. However, achieving crop recommendation with more optimal features were major concern. In [10], a Geographic Information System (GIS)-based multi-criteria technique was introduced to assess the suitability of land for soybean production.

1.1 Contributions

The significant contributions of the MPRDCSO technique are summarized as follows,

- Design an MPRDCSO technique to solve the crop recommendation problem with minimal time in the agriculture domain.
- To address null data during the preprocessing step, the Multivariate Piecewise Constant Weighted Interpolation method is employed. The Camargo's index method has been developed to effectively remove outlier data from the dataset, thereby minimizing the time required for crop recommendation.
- The MPRDCSO technique employs Rand Indexive Jensen–Shannon Divergence Cockroach Swarm Optimization to extract significant features for crop recommendations. The Rand Index is utilized for fitness estimation between features, and Jensen–Shannon Divergence is employed to discover global optimal solution. This enhances optimization algorithm as well as provides the extraction of more appropriate features.
- Finally, extensive experiments were conducted to evaluate the performance of our MPRDCSO technique and other related works. The results discussed indicate that the performance of our MPRDCSO technique is highly efficient, compared to other approaches.

1.2 Structure of the paper

The manuscript is structured into five sections as below: Literature review works are reviewed in Section 2. Section 3 explains MPRDCSO technique with neat diagram. In Section 4, experimental settings and implementation procedures are presented, followed by a comparative analysis of different methods, and various performance metrics are presented in Section 5. Finally, Section 6 gives conclusion.

2. LITERATURE REVIEW

A novel Artificial Neural Network (ANN) was developed in [11] to recommend suitable crops based on soil properties and weather parameters. However, the designed recommendation system was difficult to consider the crucial factors such as geographical, environmental, and economic aspects necessary for a successful farming system. A convolutional neural network (CNN) model, introduced in [12], incorporated a similarity tree for agriculture recommendations, achieving higher precision and accuracy in evaluation. However, the time complexity of the recommendation was not minimized.

An electronic agricultural record (EAR) was introduced in [13] to combine numerous separate datasets to recommend the optimum crops fertilization. The Nutrient Expert (NE) based decision support system was introduced in [14] to generate field or area-specific fertilizer recommendations. However, effective machine learning models were not employed for precise fertilizer recommendations. A new cloud-based, machine learning-powered crop recommendation approach was introduced in [15] to predict the crops that require to cultivated depend on assortment of known parameters. ML-based method for crop as well as fertilizer recommendation was developed [16]. This method was designed to detect crops for various seasons using multiple criteria.

A random forest regression (RFR) was introduced in [17] to predict the season of corn N nutrition index (NNI) by integrating various parameters, including soil, weather, and management data. A distinct machine learning (ML) model was developed in [18] for crop prediction, incorporating efficient feature selection methods and raw data preprocessing. However, achieving higher accuracy in crop prediction posed a significant challenge. A crop recommendation system was implemented in [19] based on regional basis, weather data and various machine learning algorithms. But, developing a fully automated system to capture reliable meteorological data was a major challenging issue. A Random Forest-based system was introduced in [20] to recommend crops for farmers based on input from temperature, soil, moisture, and nutrients. However, achieving higher accuracy was a major challenge for drawing improved conclusions about which crops to cultivate.

3. METHODOLOGY

Agriculture is the most important source of income for a country. Precision agriculture is explained as set of decision support methods in agriculture aimed at handling spatial as well as temporal inconsistencies to increase crop yield, quality and so on. Due to rapid changes in environmental conditions, soil conditions, and water availability, selecting the most appropriate crop to cultivate for a given piece of land is a major challenging task in agriculture. This is because every crop has its specific climate conditions. To address these issues, a crop recommendation system plays a vital role in predicting sustainable crops and management systems. This system provides valuable insights for farmers to optimize crop production and enhance crop yield. In this paper, the MPRDCSO technique is introduced for accurate crop recommendations with minimal time consumption in the agriculture domain.

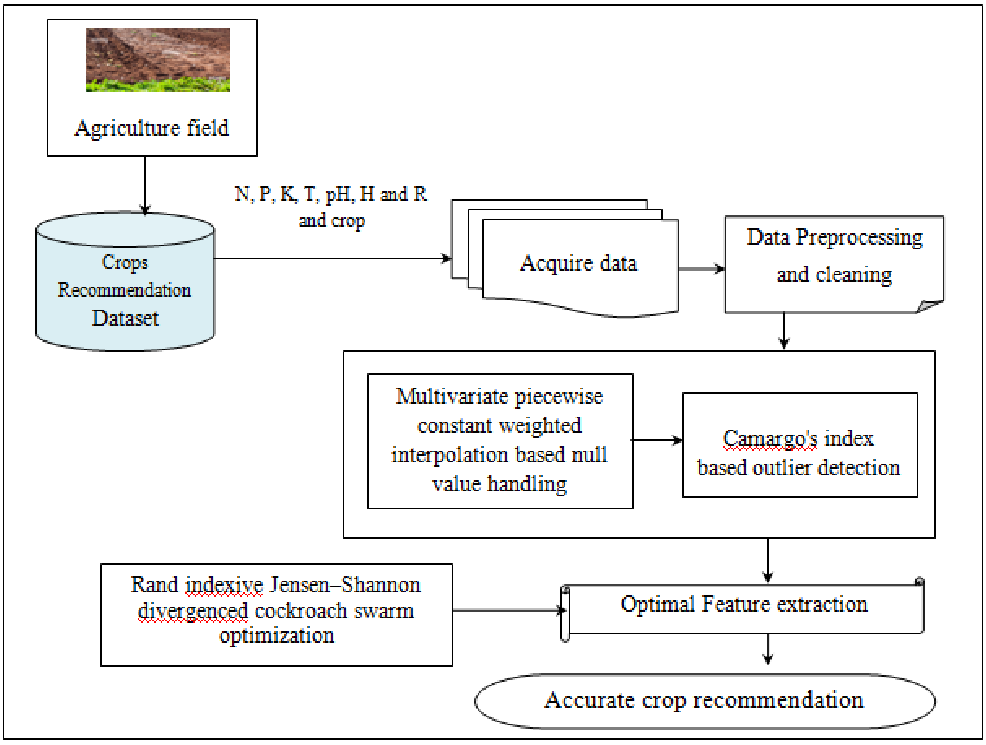

Figure 1 architecture of proposed MPRDCSO technique for crop recommendation

Figure 1 illustrates architectural design of proposed MPRDCSO technique for crop recommendation. In this design, farmers select suitable crops for cultivation based on soil and climate conditions. The initial dataset comprises various features or attributes denoted as $X_1, X_2, X_3, \dots X_m$, and a set of instances or data records denoted as

$D_1, D_2, D_3, \dots D_n$ representing training instances. The crop recommendation dataset encompasses several variables such as soil characteristics and weather conditions, including phosphorus (P), nitrogen (N), potassium (K), temperature (T), humidity (H), pH, as well as rainfall (R). These data are obtained from the dataset for crop recommendation analysis.

Following data acquisition, the proposed technique involves two stages namely preprocessing and feature extraction, aimed at enhancing the accuracy of crop recommendations. Initially, data preprocessing stage is performed to clean and transform the raw dataset into a suitable format, thus improving the crop recommendation process. Subsequently, relevant features are extracted from the processed dataset to enable efficient crop recommendation predictions and enhance accuracy with minimal time complexity.

These two dissimilar processes of the MPRDCSO technique are explained briefly in the following subsections.

3.1 Preprocessing stage

Data preprocessing plays a vital step in the data mining particularly in the context of crop recommendation systems. It involves a sequence of operations to clean, transform, and organize the raw dataset, and making it suitable format for further analysis. The preprocessing stage comprises of Null data handling and Remove outliers.

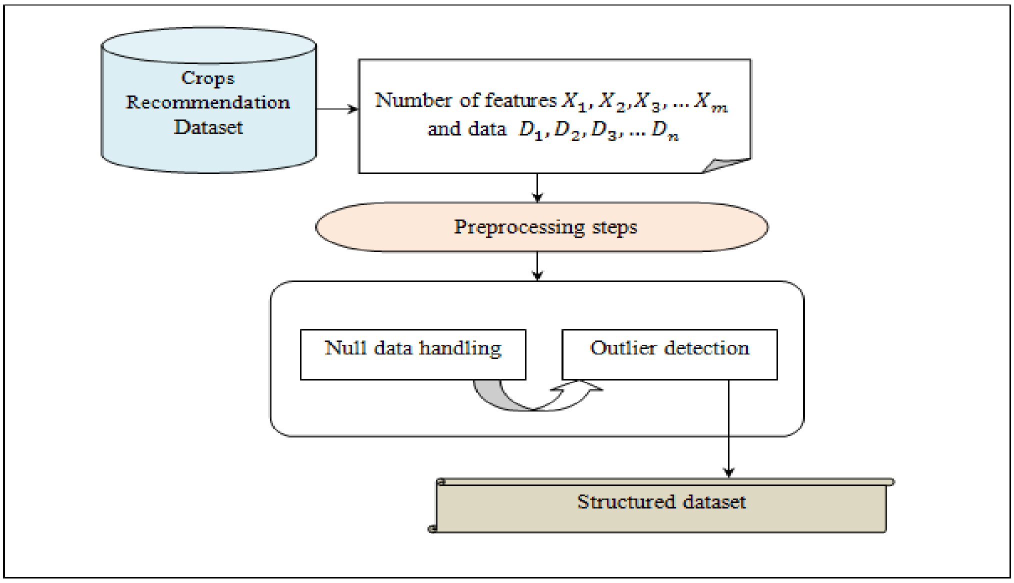

Figure 2 flow process of data preprocessing

As shown in the above figure 2, flow processes of data preprocessing is carried out with the objective of attaining the structured dataset by includes different steps namely Null data handling and outlier detection.

The raw input crop recommendation dataset 'CRD' and the input features $X_1, X_2, X_3, \dots X_m$ and each input data are represented by $D_1, D_2, D_3, \dots D_n$ are acquired from the database. In the dataset, the data are organized into 'n' number of rows and 'm' columns, where each row represents an instance or record, and each column represents features or attributes. The sample crop recommendation dataset details of five samples, along with their features and respective data values that are stated in table 1.

Table 1 crop recommendation dataset details

S.No	N	P	K	temperature	humidity	ph	rainfall	label
1	90	42	43	20.87974	82.00274	6.502985	202.9355	rice
2	85	58	41	21.77046	80.31964	7.038096	226.6555	rice
3	60	55	44	23.00446	82.32076	7.840207	263.9642	rice
4	74	35	40	26.4911	80.15836	6.980401	242.864	rice
5	78	42	42	20.13017	81.60487	7.628473	262.7173	rice

With the above sample obtained from crop recommendation dataset, data processing is performed as follows,

3.1.1 Handling null data

Handling null (or missing) data is a fundamental step in data preprocessing stage to ensure the accuracy and reliability of analyses. The proposed MPRDCSO technique uses the multivariate piecewise constant weighted interpolation method for dealing with null data (or missing). This method handling the missing data, based on the values of their nearest neighbors. Interpolation is a method used for determining a new data points based on the set of known nearest neighbor's data points in the dataset.

The missing value is determined as a weighted average of its nearest neighbors. Let's consider the feature 'X' with missing data values 'Dp_M', and to estimate the weighted average of its nearest neighbors. This is mathematically expressed as given below.

$$Dp_M = w_{avg}\ (ND) \quad (1)$$

$$w_{avg}\ (ND) = \frac{\sum_{i=1}^{n} \beta_i\, Dp_i}{\sum_{i=1}^{n} \beta_i} \quad (2)$$

Where, Dp_M indicatees the missing value to be estimated, w_{avg} denotes a weighted average, ND proximity data (i.e. nearest data), β_i denotes a weight assigned to the nearest neighbor based on a distance metric. The distance between the two data points are estimated using Euclidean distance measure as give below,

$$dst = \sqrt{\left(Dp_j - Dp_i\right)^2} \quad (3)$$

Where, 'dst' denotes a distance between the two data points Dp_i and Dp_j. In this way, null or missing values are handled.

3.1.2 Outlier detection

Detecting and handling outliers is a crucial step in data preprocessing to ensure the quality of machine learning models. The process involves identifying noisy or outlier data points in a raw dataset that significantly deviate from the majority of other data points in a specific feature. The Camargo's index is utilized as qualitative method to assess the dependency between data points within a particular feature. It serves as a measure to determine whether two data points are independent or dependent.

Let's consider the feature 'X' with number of data points 'Dp_M', The Camargo's index function is formulated as follows,

$$\varphi_{CI} = 1 - \sum_{j=1}^{m} \left(\frac{|Dp_i - Dp_j|}{m}\right) \quad (4)$$

Where, φ_{CI} indicates a Camargo's index function, Dp_i denotes a current data point, Dp_j indicates other data points in the particular feature, m indicates a number of data points. From the analysis, the index function provides the output ranges from the 0 to 1. Set thresholds to determine which data points are considered outliers.

$$Z = \begin{cases} ODp, & \varphi_{CI} < T \\ NDp, & \varphi_{CI} > T \end{cases} \quad (5)$$

Where, Z denotes a outcome, ODp denotes an outliers data points, NDp indicates a normal data points, φ_{CI} denotes a Camargo's index, T indicates a threshold. From the analysis, outlier's data points are removed. The pseudo code representation of data preprocessing is given below.

Algorithm 1: data preprocessing
Input: Dataset 'CRD', Features '$X = \{X_1, X_2, \dots, X_m\}$', records or instances '$D = \{D_1, D_2, \dots, D_n\}$'
Output: preprocessed dataset
Begin **Step 1:** Collect number of Features '$X = \{X_1, X_2, \dots, X_m\}$', records or instances '$D = \{D_1, D_2, \dots, D_n\}$ **Step 2**: **For** each features 'X' and instances 'D' **Step 3**: Handle null or missing value 'Dp_M' by measuring the weighted average w_{avg} (ND) **Step 4:** Assign weight based on distance using (3) **Step 5**: **For** each data point Dp_i and Dp_j **Step 6:** Measure Camargo's index function using (4) **Step 7:** **if** $(\varphi_{CI} < T)$ **then** **Step 8:** Outlier's data points **Step 9:** **else** **Step 10**: Normal data points **Step 11: End if** **Step 12: Remove** outlier's data points **Step 13: Return** preprocessed data **Step 14: End for** **Step 15: End for** **End**

Algorithm 1 outlines the data preprocessing procedure aimed at minimizing the error rate associated with crop recommendation detection. Initially, input features and data records or instances are gathered from the dataset. Subsequently, null or missing data are handled using a multivariate piecewise constant weighted interpolation method. Missing data are filled through weighted average of other data points in specific features. Following this, outlier data points are identified using the Camargo's index function. A threshold value is then assigned to distinguish between outlier data points and normal data points. As a result, the preprocessed results are obtained.

3.2 Rand indexive Jensen–Shannon divergenced cockroach swarm optimization based feature extraction

After the data preprocessing, feature extraction is performed to reduce dimensionality of the dataset. Feature extraction is the process of extracting important features from the original dataset time-consuming and required expertise. These redundant and unnecessary features reduce the performance model. Therefore, the proposed MPRDCSO technique uses Rand indexive Jensen–Shannon divergenced cockroach swarm optimization to extract important features from the original dataset to minimize time-consuming for crop recommendation.

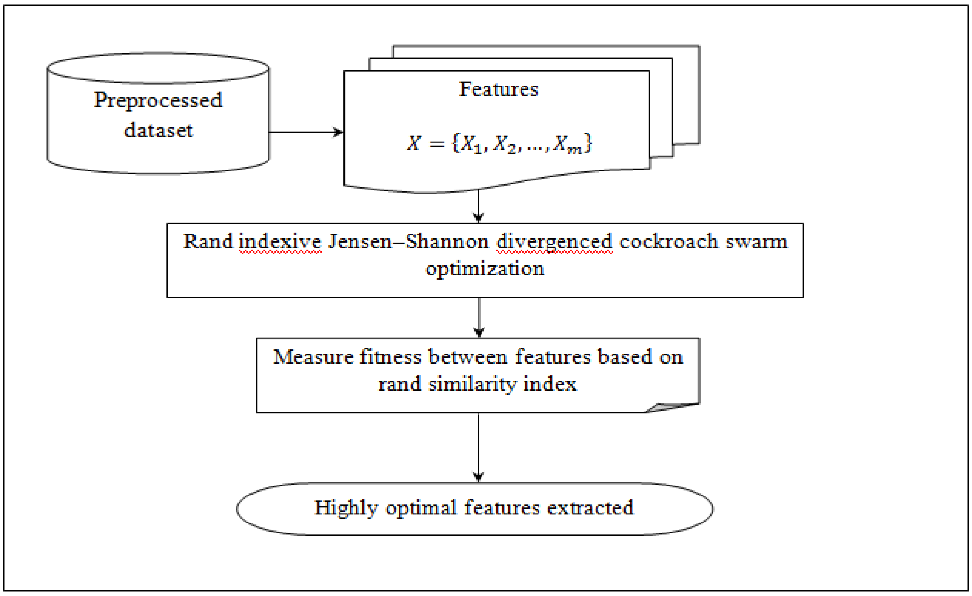

Figure 3 flow process of Rand indexive Jensen–Shannon divergenced cockroach swarm optimization based Feature extraction

As illustrated in the above figure 3 with the preprocessed data acquired as input, first, the number of features are subjected to Rand indexive Jensen–Shannon divergenced cockroach swarm optimization method is used for extracting significant features in high-dimensional space into low-dimensional space.

Cockroach swarm optimization is a metaheuristic optimization enthused through basic natural activities of cockroach searching for their food source. Initialize populations of cockroach i.e. features' in search space.

$$X = \{X_1, X_2, \ldots, X_m\} \quad (6)$$

After the initialization process, the fitness is computed based on the Rand similarity. It is a statistical technique used to measure the relationship between the features. The rand similarity is mathematically computed as follows,

$$f \Rightarrow RS = 1 - \frac{|X_i \,\Delta\, X_j|}{m} \quad (7)$$

Where, RS indicates a rand similarity measures, f denotes a fitness, X_i, X_j denotes features in the dataset, m denotes a data sample size i.e. number of features. The similarity 'RC returns a value from 0 to 1.

Based on the fitness evaluation, the algorithm performs three fundamental behaviors for such as chase-swarming, dispersing, and ruthless.

- **Chase-swarming behaviors**

Algorithm selects current best features from the population based on its fitness value, and this cockroach moves towards the global optimum. If the fitness of one cockroach (i.e. X_i) is greater than that of another (i.e. X_j), after that location of local best solutions in the search space is updated towards the global best solution. The updated results are obtained as follows,

$$Q\ (t+1) = Q(t) + M * R * 0.5\left|X_g - Q(t)\right| \quad (8)$$

Where, $Q\ (t+1)$ denotes an updated solution,$Q(t)$ denotes a current local best features, 'M' indicates step that is a fixed value,R denotes random number [0, 1],X_g indicates the global best solution, $0.5\left|X_g - Q(t)\right|$ represents the Jensen–Shannon divergence for measuring the similarity between the global best solution and current local best features. Otherwise, the cockroach goes local best solution.

$$Q(t+1) = Q(t) + M * R * 0.5|X_l - Q(t)| \quad (9)$$

Where, $Q\ (t+1)$ indicates an updated solution, $Q(t)$ indicates a current local best solutions, 'M' designates step that is fixed value, R denotes random number [0, 1], X_l indicates local best solutions.

- **Dispersion**

Another activity is dispersion and it is implemented to maintain the diversity among the cockroaches. Some cockroaches move away from the current solutions, discovering new areas of the search space. This helps to prevent premature convergence to suboptimal solutions. The dispersion behavior of the cockroach in the search space is expressed as follows.

$$Q\ (t+1)\ = Q(t) + B\ (1,d) \quad (10)$$

Where, $Q(t+1)$ indicates an updated solution, $Q(t)$ represents a current local best solutions $B\ (1,d)$ designates d-dimensional random vector which value is set in a assured range.

- **Ruthless behavior**

The term ruthless in this context, is likely metaphorical and refers to the selection process where weaker solutions are eliminated by better ones. The ruthless replacement involves replacing a randomly chosen individual in the swarm with this best solution.

$$X_r = X_g \quad (11)$$

Where, X_r denotes a random individual in the swarm and X_g indicates the global best position. The above-said process is continued until the maximum iteration gets achieved. In this way, the global optimum feature is chosen for crop recommendation. As a result, optimal features are extracted from the dataset. With the extracted features, crop recommendation is performed with higher accuracy with minimum time consumption. The pseudo code for Rand indexive Jensen–Shannon divergenced cockroach swarm optimization is given below,

// **Algorithm 2: Rand indexive Jensen–Shannon divergenced cockroach swarm optimization**
Input : Dataset 'DS', Features '$X = \{X_1, X_2, \dots, X_m\}$'
Output: Extracted optimal features
Begin **Step 1:** Initialize the populations of X_i in search space **Step 2:** **For each** X_i **Step 3:** Measure the fitness '$f(x)$' using (7) **Step 4:** **While** (t < Max_iter) **do** **Step 5:** **If** $(f\ (X_j) > f\ (X_i))$ **then** **Step 6:** Update the position $Q\ (t+1) = Q(t) + M * R * 0.5\|X_g - Q(t)\|$ **Step 7:** **else** **Step 8:** Update the position $Q(t+1)\ = Q(t) + M * R * 0.5\|X_l - Q(t)\|$ **Step 9:** **End if** **Step 10:** **for** j = 1 to N **do** **Step 11:** Execute the Dispersion $Q\ (t+1)\ = Q(t) + B\ (1,d)$ **Step 12:** Execute the Ruthless behavior $X_r\ = X_g$ **Step 13:** **End for** **Step 14:** **t = t+1** **Step 15:** **Go to step 4** **Step 16:** **End while** **Step 17:** Obtain the global best features **Step 18:** **End for** **End**

The Rand Indexive Jensen–Shannon Divergence Cockroach Swarm Optimization process for crop recommendation accuracy is described as shown in algorithm 2. Initially, the number of features in the search space is initialized. After initialization, fitness is measured based on the Rand similarity measure between features. If the fitness of one feature is higher than other, the position of the current best solution is updated, establishing it as global best. Otherwise, local best solution is updated. Subsequently, Dispersion and Ruthless behavior are executed to replace a randomly chosen individual with the global best. This iterative process continues until a highest number of iterations are reached. Finally, this optimization method extracts optimized features as output for accurate crop recommendation and effectively minimizing the dimensionality of the dataset.

4. EXPERIMENTAL SCENARIO

The experimental setup is designed to evaluate the performance of the proposed Multivariate Piecewise Rand Divergencive Cockroach Swarm Optimization (MPRDCSO) and existing IDCSO-WLSTM [1] and DRL [2] using Python coding. A comprehensive crop recommendation dataset is gathered from https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset. This dataset includes relevant information such as soil characteristics, weather conditions, and region-specific information, historical crop performance. The dataset includes 8 features and 2200 instances for crop recommendation. Features in the dataset might include information such as: Soil-related factors such as pH, potassium, Phosphorous, Nitrogen. Climate-related factors are temperature, rainfall and humidity.

Table 2 description of features

S. No	Features or attributes	Description
1	N	ratio of Nitrogen content in soil
2	P	ratio of Phosphorous content in soil
3	K	ratio of Potassium content in soil
4	Temperature	temperature in degree Celsius
5	ph	ph value of the soil
6	rainfall	rainfall in mm
7	Label	**22 unique labels** 0: 'apple', 1: banana, 2: blackgram,3: chickpea, 4:coconut, 5: coffee, 6: cotton, 7: grapes, 8: jute, 9: kidneybeans, 10: lentil, 11: maize, 12: mango, 13: mothbeans, 14: mungbean, 15:muskmelon, 16: orange, 17: papaya, 18: pigeonpeas, 19: pomegranate, 20: rice, 21:watermelon

5. PERFORMANCE RESULTS ANALYSES

Experimental outcomes of MPRDCSO and existing IDCSO-WLSTM [1] and DRL [2] are discussed with accuracy, precision, recall, F-score, and crop recommendation time across different instances.

Accuracy: Achieving accuracy in crop recommendation involves developing models that effectively predict the most suitable crops for specific weather and soil conditions. The accuracy is mathematically stated as given below.

$$Accuracy = \left[\frac{TR_p+FL_p}{TR_p+FL_p+TR_N+FL_N}\right] * 100 \qquad (12)$$

Where TR_p indicates a true positive, FL_p denotes a false positive, TR_N indicates the true negative, FL_N represents the false negative. The accuracy is measured in percentage (%).

Precision: It is the performance metrics that assesses the relevancy of the recommended crops. It is mathematically formulated as given below,

$$Pc = \left(\frac{TR_p}{TR_p+ FL_p}\right)(13)$$

Where, Pc represents a Precision, TR_p symbolizes the true positive, FL_p indicates false positive.

Recall: Recall also known as Sensitivity that is calculated based on number of true positives as well as false negatives during the crop recommendation. It is calculated as follows,

$$Rl = \left(\frac{TR_p}{TR_p+FL_n}\right) \quad (14)$$

Where 'Rl 'indicates a recall, TR_p denotes a true positive, FL_n denotes the false negative.

F-score: It is average value of precisions as well as recall. It is calculated as given below,

$$F - score = \left[2 * \frac{Pc * Rl}{Pc + Rl}\right] \quad \textbf{(15)}$$

Where F-score is computed based on precision Pc and recall 'Rl'.

Crop recommendation time: It is measured as the amount of time taken to predict the most appropriate crops with the extracted optimal features. This is mathematically estimated as follows:

$$CRT = \sum_{i=1}^{n} I_i * \ time\ [CP] \quad (16)$$

Where, CRT denotes a crop recommendation time, I_i indicates a number of instances $time\ [CP]$ denotes a time for crop prediction. It is computed in the unit of milliseconds (ms).

Table 3 comparison of accuracy

Number of instances	Accuracy (%)		
	MPRDCSO	**IDCSO-WLSTM**	**DRL**
200	90	85	82.5
400	91.25	87.5	83.75
600	90.83	85.83	81.66
800	90	85	82.5
1000	90.5	86	84
1200	90.83	87.5	85
1400	89.28	87.14	85.71
1600	89.37	86.87	85
1800	91.11	87.22	85.55
2000	89.5	86	84.5

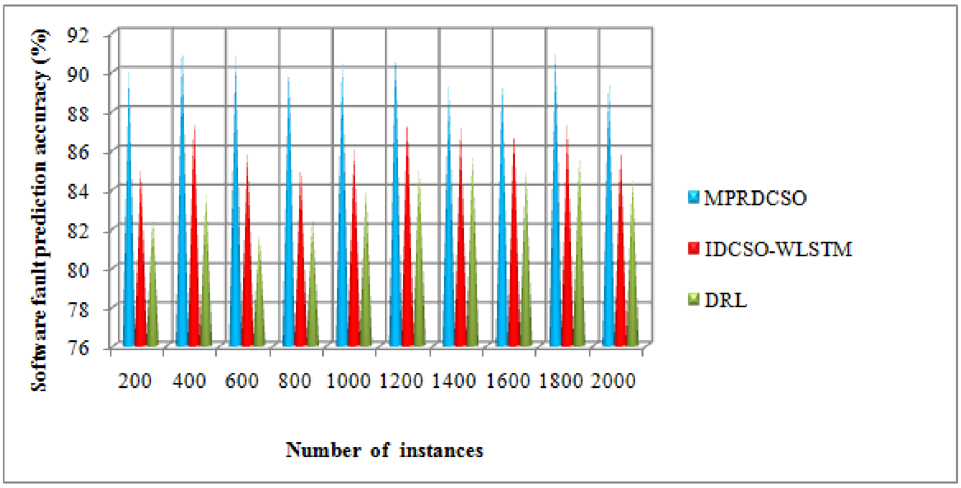

Figure 4 graphical illustration of accuracy

Figure 4 illustrates the graphical representation of crop prediction accuracy through the recommendation process using three methods namely MPRDCSO, existing IDCSO-WLSTM [1] and DRL [2]. The observed results demonstrate that the proposed MPRDCSO technique outperforms the other methods, [1] and [2]. In the experiment conducted with 200 instances, the accuracy was observed to be 90% using the proposed MPRDCSO technique, and 85% and 82.5% using existing [1] and [2], respectively. The overall analysis of ten results indicates that the prediction accuracy, using the proposed MPRDCSO technique is significantly increased by 4% compared to [1] and 7% compared to [2]. This superior performance is achieved in the MPRDCSO technique's ability to perform optimal feature extraction from the dataset using the Rand Indexive Jensen–Shannon Divergence Cockroach Swarm Optimization algorithm in crop recommendation. The Rand index measures the similarity between features in fitness estimation. Consequently, the MPRDCSO technique extracts the most optimal features for predicting the crop, thereby enhancing accuracy.

Table 4 comparison of Precision

Number of instances	Precision		
	MPRDCSO	IDCSO-WLSTM	DRL
200	0.903	0.866	0.857
400	0.940	0.906	0.868
600	0.931	0.906	0.869
800	0.927	0.893	0.875
1000	0.936	0.904	0.888
1200	0.942	0.911	0.897
1400	0.933	0.915	0.904
1600	0.935	0.911	0.895
1800	0.944	0.915	0.907
2000	0.931	0.911	0.898

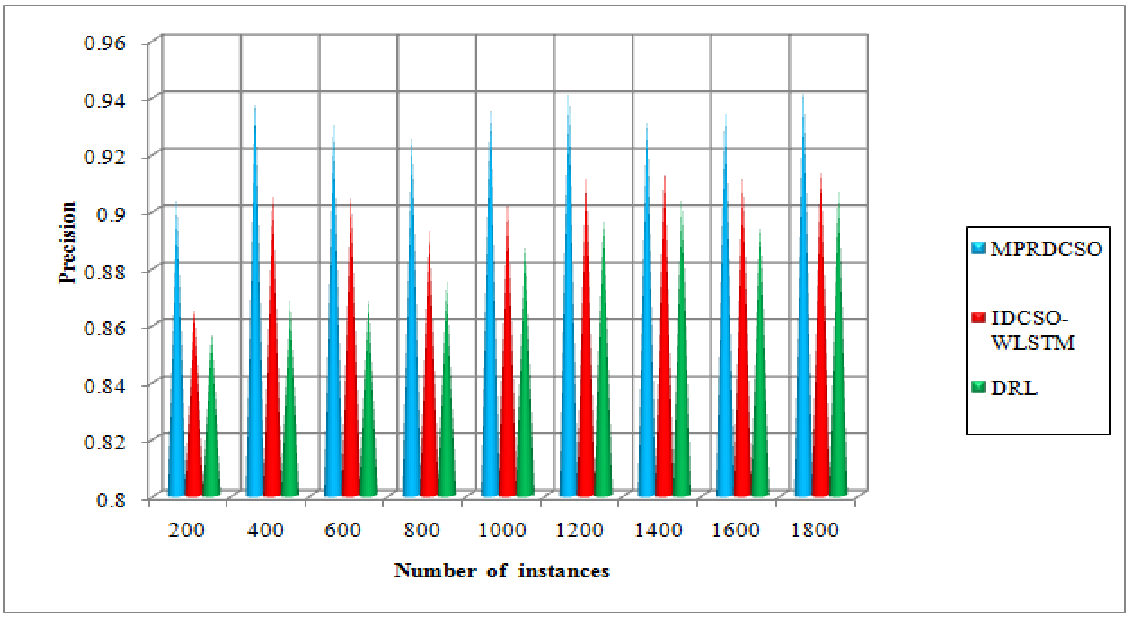

Figure 5 graphical illustration of precision

Figure 5 presents a graphical illustration of precision using three different methods namely MPRDCSO technique, existing methods IDCSO-WLSTM [1] and DRL [2]. The precision performance using the MPRDCSO technique was higher than that of the existing methods. This is because of the accurate classification of all instances into their respective classes (i.e., different crops), with only a minimal number of instances being incorrectly categorized. In simulations conducted with 200 inductances, the precision was observed to be 0.903 using the MPRDCSO technique, and it was 0.866 and 0.857 using methods [1] and [2], respectively. The analysis indicates that the true positive rate is enhanced by implementing an outlier data removal process in the preprocessing phase of the MPRDCSO technique. Additionally, the most optimal feature selection process contributes to the improved precision performance. Upon comparing the overall performance of the MPRDCSO technique with existing methods, it is evident that precision has significantly increased by 3% and 5% compared to [1] and [2], respectively.

Table 5 comparison of recall

Number of instances	Recall		
	MPRDCSO	IDCSO-WLSTM	DRL
200	0.965	0.928	0.888
400	0.954	0.935	0.913
600	0.959	0.915	0.888
800	0.955	0.921	0.903
1000	0.953	0.926	0.911
1200	0.951	0.939	0.916
1400	0.941	0.931	0.920
1600	0.942	0.932	0.923
1800	0.955	0.933	0.919
2000	0.947	0.922	0.914

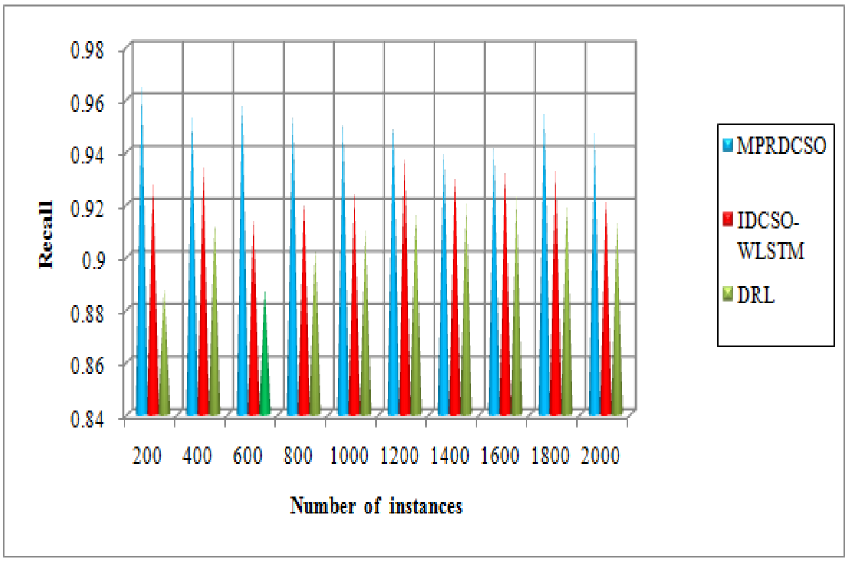

Figure 6 graphical illustration of recall

Figure 6 reveals the graphical illustration of recall versus the number of instances, showing the outcomes of three different methods namely the MPRDCSO technique, and existing methods IDCSO-WLSTM [1] and DRL [2].

Analysis of Figure 6 reveals that the MPRDCSO technique consistently outperforms conventional methods in terms of recall. In the first iteration with 200 instances, the observed recall performance using the MPRDCSO technique was 0.965. In contrast, the recall values for existing methods [1] and [2] were 0.928 and 0.888, respectively. The statistical evaluation demonstrates a notable improvement in recall using the MPRDCSO technique. To quantify this improvement, a comparative analysis was conducted. The overall recall comparison results indicate that the MPRDCSO technique outperforms existing methods by 3% and 5% when compared to [1] and [2], respectively. This improvement is achieved by the MPRDCSO technique's ability to extract optimal features for crop prediction, providing accurate labels that enhance true positives and minimize true negatives.

Table 6 comparison of F-score

Number of instances	F-score		
	MPRDCSO	IDCSO-WLSTM	DRL
200	0.932	0.895	0.872
400	0.946	0.920	0.889
600	0.944	0.910	0.878
800	0.940	0.906	0.888
1000	0.944	0.914	0.899
1200	0.946	0.924	0.906
1400	0.936	0.922	0.911
1600	0.938	0.921	0.908
1800	0.949	0.923	0.912
2000	0.938	0.916	0.905

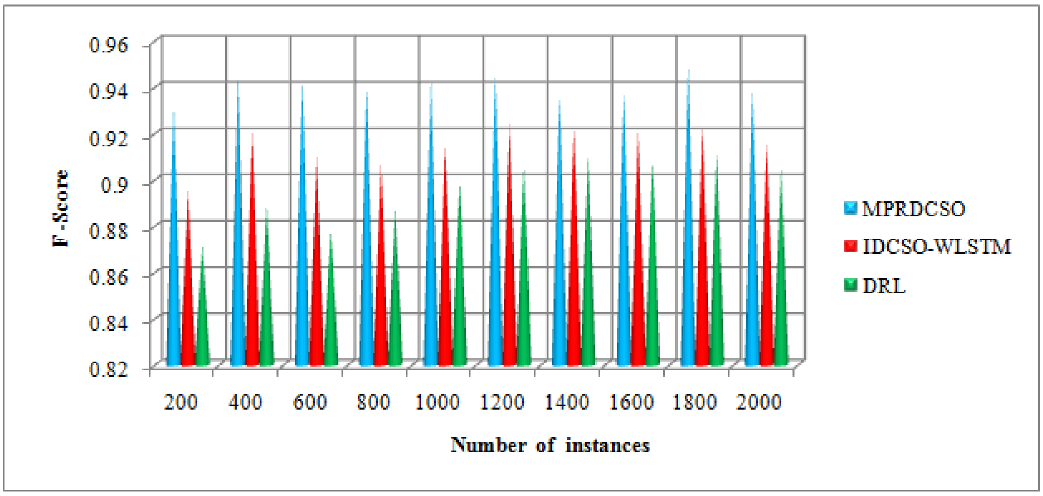

Figure 7 graphical illustrations of F-score

Figure 7, depicted above, illustrates the graphical representation of F-score across a range of instances, ranging from 200 to 2000. The observed result reveals that the F-score using the MPRDCSO technique was found to be 0.932, and recall was found to be 0.895 and 0.872 by applying IDCSO-WLSTM [1] and DRL [2] respectively. These findings indicate an important improvement in F-measure performance with the proposed MPRDCSO technique in comparison to existing methods. The application of the MPRDCSO technique resulted in enhanced precision as well as recall during crop prediction. Subsequently, the actual predictions were made, leading to an overall improvement in F-score using the MPRDCSO technique. The comprehensive performance results indicate a considerable enhancement in the F-score of the MPRDCSO technique by 3% compared to [1] and 5% compared to [2], respectively.

Table 7 comparison of crop recommendation time

Number of instances	Crop recommendation time (ms)		
	MPRDCSO	IDCSO-WLSTM	DRL
200	30	33	35.25
400	33	36	39
600	36.45	40.5	42.75
800	43.2	45	47.4
1000	46.5	48.75	51
1200	49.5	51.3	53.1
1400	55.65	57.75	59.85
1600	58.8	60	62.4
1800	60.75	62.1	64.8
2000	63	65.25	67.5

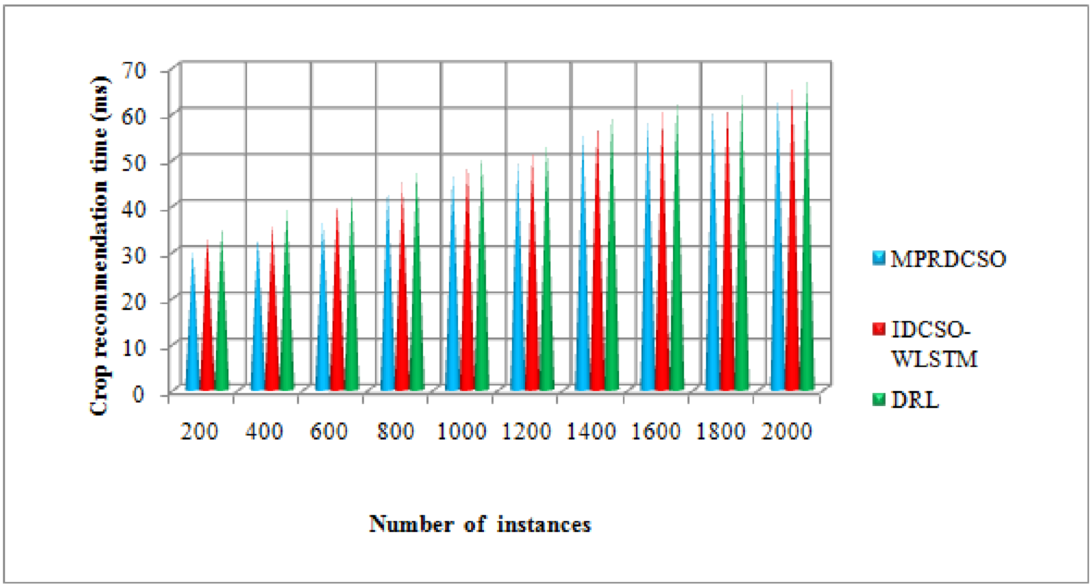

Figure 8 graphical illustrations of crop recommendation time

Figure 8 reveals the graphical results of crop recommendation time using the three methods, MPRDCSO technique, and existing methods IDCSO-WLSTM [1] and DRL [2] respectively. From the figure it noticed that crop recommendation time enhances through increase in number of instances from 200 to 2000. This is due to the reason that with larger number of instances involved during experimentation, large amount of time is said to be consumed during analyzing of different modules this in turn increases the time also. However, with experiments conducted with 200 instances, the time consumed in crop recommendation time being '30*ms*', the overall time using [1] method was 33ms and '35.25ms using [2]. From this result it is inferred that the crop recommendation time using MPRDCSO technique was considerably reduced by 5% and 10% when compared to [1] and [2]. The improvement is due to the data preprocessing. The null or missing data are handled by applying a multivariate piecewise constant weighted interpolation method. The duplicate data are identified and removed from the dataset using Camargo's index. This helps to minimize the time consumption of crop recommendation.

6. CONCLUSION

Crop prediction in agriculture for cultivation in agriculture is a complex process and multiple models have been proposed and tested. In this paper, an efficient crop recommendation technique MPRDCSO designed to predict suitable crops for cultivation based on environmental and soil qualities. The primary goal of MPRDCSO is to enhance the accuracy of crop prediction through efficient preprocessing and optimal feature extraction. The algorithm employs various steps to achieve this objective. Firstly, the raw dataset preprocessed to minimize the time complexity of crop recommendation. Following data processing, optimal features are extracted from the dataset using Rand Indexive Jensen–Shannon Divergence and Cockroach Swarm Optimization. These techniques are chosen for their effectiveness in identifying key features that significantly contribute to accurate crop predictions. The experimental assessment of the MPRDCSO algorithm is conducted with conventional techniques, using a crop recommendation dataset. The results obtained highlight the superior performance of the MPRDCSO algorithm. Specifically, it demonstrates increased accuracy in crop prediction, precision, recall, and F-score. Moreover, the algorithm proves effective in minimizing the time required for the recommendation process.

REFERENCES

[1] S. Kiruthika and, D. Karthika, “IOT-BASED professional crop recommendation system using a weight-based long-term memory approach”, Measurement: Sensors, Elsevier, Volume 27, June 2023, Pages 1-10. https://doi.org/10.1016/j.measen.2023.100722

[2] Mohamed Bouni, Badr Hssina, Khadija Douzi, Samira Douzi, “Towards an Efficient Recommender Systems in Smart Agriculture: A deep reinforcement learning approach”, Procedia Computer Science, Elsevier, Volume 203, 2022, Pages 825-830. https://doi.org/10.1016/j.procs.2022.07.124

[3] Arfat Ahmad Khan, Muhammad Faheem, Rab Nawaz Bashir, Chitapong Wechtaisong, And Muhammad Zahid Abbas, “Internet of Things (IoT) Assisted Context Aware Fertilizer Recommendation”, IEEE Access , Volume 10, 2022, Pages 129505 – 129519. **DOI:** 10.1109/ACCESS.2022.3228160

[4]V. Elizabeth Jesi, Anil Kumar, Bappa Hosen and Stalin David D, “IoT Enabled Smart Irrigation and Cultivation Recommendation System for Precision Agriculture”, ECS Transactions, Volume 107, 2022, Pages 5953-5967. **DOI** 10.1149/10701.5953ecst

[5] Usman Ahmed, Jerry Chun-Wei Lin, Gautam Srivastava, Youcef Djenouri, “A nutrient recommendation system for soil fertilization based on evolutionary computation”, Computers and Electronics in Agriculture, Elsevier, Volume 189, October 2021, Pages 1-7. https://doi.org/10.1016/j.compag.2021.106407

[6] G. Mariammal, A. Suruliandi, S. P. Raja, E. Poongothai, “Prediction of Land Suitability for Crop Cultivation Based on Soil and Environmental Characteristics Using Modified Recursive Feature Elimination Technique With Various Classifiers”, IEEE Transactions on Computational Social Systems , Volume 8, Issue 5, 2021, Pages 1132 – 1142. **DOI:** 10.1109/TCSS.2021.3074534

[7] Murali Krishna Senapaty, Abhishek Ray and Neelamadhab Padhy , “IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture”, Computers, Volume 12, Issue 3, 2023, Pages 1-34. https://doi.org/10.3390/computers12030061

[8] Husam Lahza, K. R. Naveen Kumar, B. R. Sreenivasa, Tawfeeq Shawly, Ahmed A. Alsheikhy, Arun Kumar Hiremath and Hassan Fareed M. Lahza , “Optimization of Crop Recommendations Using Novel Machine Learning Techniques”, Sustainability, Volume 15, Issue 11, 2023, Pages 1-18. https://doi.org/10.3390/su15118836

[9] Zujiao Shi, Donghua Liu, Miao Liu, Muhammad Bilal Hafeez, Pengfei Wen, Xiaoli Wang, Rui Wang, Xudong Zhang, Jun Li, “Optimized fertilizer recommendation method for nitrate residue control in a wheat–maize double cropping system in dryland farming”, Field Crops Research, Elsevier, Volume 271, 2021, Pages 1-12. https://doi.org/10.1016/j.fcr.2021.108258

[10] Fatihu Kabir Sadiq, Suleiman Lawan Ya’u, Jamila Aliyu, Lemuel Musa Maniyunda, “Evaluation of land suitability for soybean production using GIS-based multi-criteria approach in Kudan Local Government area of Kaduna State Nigeria”, Environmental and Sustainability Indicators, Elsevier, Volume 20, December 2023, Pages1 -10. https://doi.org/10.1016/j.indic.2023.100297

[11] J Madhuri, M Indiramma, “Artificial Neural Networks Based Integrated Crop Recommendation System Using Soil and Climatic Parameters”, Indian Journal of Science and Technology, Volume 14, Issue 19, 2021, Pages 1587-1597. https://doi.org/ 10.17485/IJST/v14i19.64

[12] Nguyen Ha Huy Cuong, Trung Hai Trinh , Duc-Hien Nguyen, Thanh Khiet Bui, Tran Anh Kiet, Phan Hieu Ho, Nguyen Thanh Thuy, “An approach based on deep learning that recommends fertilizers and pesticides for agriculture recommendation”, International Journal of Electrical and Computer Engineering (IJECE), Volume 12, Issue 5, 2022, Pages 5580-5588. http://doi.org/10.11591/ijece.v12i5.pp5580-5588

[13] Vuong M. Ngo, Thuy-Van T. Duong, Tat-Bao-Thien Nguyen, Cach N. Dang & Owen Conlan, “A big data smart agricultural system: recommending optimum fertilisers for crops”, International Journal of Information Technology, Springer, volume 15, 2023, pages 249–265. https://doi.org/10.1007/s41870-022-01150-1

[14] Jairos Rurindaa, Shamie Zingorea, Jibrin M. Jibrinc , Tesfaye Balemid , Kenneth Masukie , Jens A. Anderssonf , Mirasol F. Pampolinog , Ibrahim Mohammedh , James Mutegia , Alpha Y. Kamarah , Bernard Vanlauwei , Peter Q. Craufurd, “Science-based decision support for formulating crop fertilizer recommendations in sub-Saharan Africa”, Agricultural Systems, Elsevier, Volume 180, 2020, Pages 1-13. https://doi.org/10.1016/j.agsy.2020.102790

[15] Navod Neranjan Thilakarathne, Muhammad Saifullah Abu Bakar , Pg Emerolylariffion Abas and Hayati Yassin., “A Cloud Enabled Crop Recommendation Platform for Machine Learning-Driven Precision Farming”, Sensors, Volume 22, Issue 16, 2022, Pages 1-21. https://doi.org/10.3390/s22166299

[16] Krupa Patel and Hiren B. Patel, “Multi-criteria Agriculture Recommendation System using Machine Learning for Crop and Fertilizesrs Prediction”, Current Agriculture Research Journal, Volume 11, Issue 1, 2023, Pages 137-149. http://dx.doi.org/10.12944/CARJ.11.1.12

[17] Xinbing Wang, Yuxin Miao, Rui Dong, Hainie Zha, Tingting Xia, Zhichao Chen, Krzysztof Kusnierek, Guohua Mi, Hong Sun, Minzan Li, “Machine learning-based in-season nitrogen status diagnosis and side-dress nitrogen recommendation for corn”, European Journal of Agronomy, Elsevier, Volume 123, 2021, Pages 1-13. https://doi.org/10.1016/j.eja.2020.126193

[18] S. P. Raja; Barbara Sawicka; Zoran Stamenkovic; G. Mariammal , “Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers”, IEEE Access, Volume 10, 2022, Pages 23625 – 23641. **DOI:** 10.1109/ACCESS.2022.3154350

[19] Saikat Banerjee, Dr. Abhoy Chand Mondal, “A Region-Wise Weather Data-Based Crop Recommendation System Using Different Machine Learning Algorithms”, International Journal of Intelligent Systems and Applications in Engineering, Volume 11, Issue 3, Pages 283–297. https://orcid.org/0000-0002-7361-1553

[20] Pachade, R. S., and Sharma, A., “Machine learning for weather-specific crop recommendation”, International Journal of Health Sciences, Volume 6, 2022, Pages 4527–4537. https://doi.org/10.53730/ijhs.v6nS8.13222