ORIGINAL ARTICLE

Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making

Mahmoud Y. Shams[1] · Samah A. Gamel[2] · Fatma M. Talaat[1,3,4]

Abstract

Crop Recommendation Systems are invaluable tools for farmers, assisting them in making informed decisions about crop selection to optimize yields. These systems leverage a wealth of data, including soil characteristics, historical crop performance, and prevailing weather patterns, to provide personalized recommendations. In response to the growing demand for transparency and interpretability in agricultural decision-making, this study introduces XAI-CROP an innovative algorithm that harnesses eXplainable artificial intelligence (XAI) principles. The fundamental objective of XAI-CROP is to empower farmers with comprehensible insights into the recommendation process, surpassing the opaque nature of conventional machine learning models. The study rigorously compares XAI-CROP with prominent machine learning models, including Gradient Boosting (GB), Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), and Multimodal Naïve Bayes (MNB). Performance evaluation employs three essential metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R2). The empirical results unequivocally establish the superior performance of XAI-CROP. It achieves an impressively low MSE of 0.9412, indicating highly accurate crop yield predictions. Moreover, with an MAE of 0.9874, XAI-CROP consistently maintains errors below the critical threshold of 1, reinforcing its reliability. The robust R^2 value of 0.94152 underscores XAI-CROP's ability to explain 94.15% of the data's variability, highlighting its interpretability and explanatory power.

✉ Fatma M. Talaat
  fatma.nada@ai.kfs.edu.eg

  Mahmoud Y. Shams
  mahmoud.yasin@ai.kfs.edu.eg

  Samah A. Gamel
  sgamel@horus.edu.eg

1   Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh, Egypt

2   Faculty of Engineering, Horus University, Damietta, Egypt

3   Faculty of Computer Science and Engineering, New Mansoura University, Gamasa 35712, Egypt

4   Nile Higher Institute for Engineering and Technology, Mansoura, Egypt

1 Introduction

According to the Food and Agriculture Organization of the United Nations (FAO), agriculture is a significant sector in the economics of African nations, with approximately two-thirds of workers employed in this sector. Reforming agriculture in Africa is believed to be essential to eradicate poverty, hunger, and malnutrition [1]. However, the changing climate poses a challenge to the agricultural sector, with changing rainfall patterns, droughts, floods, and the spread of pests and diseases affecting crop production [2]. Therefore, predicting crop yield has become a difficult issue in precision agriculture. This is due to different variables especially the crop yield with the effect of climate change. Climate change is one of the major factors that can affect crop yields, making it essential to develop accurate predictive models to anticipate its effect. The changing environmental conditions, particularly global

warming, and climate variability, have a negative impact on the future of agriculture. These factors, combined with other variables such as climate, weather, soil, fertilizer use, and seed variety, necessitate the use of multiple datasets to address this complex problem [3]. Traditionally, predicting crop production has relied on statistical models, which can be time-consuming and arduous. However, the introduction of big data in recent years has opened up new possibilities for more advanced analysis methods, such as machine learning [4]. Machine learning models can be categorized as either descriptive or predictive, depending on the research questions and challenges at hand. Predictive models are employed to forecast the future, while descriptive models are used to learn from the data and explain past events [5–8].

Anthropogenic climate change will have a more severe impact on the agricultural industry due to its dependence on weather [9]. In estimating yield for the purpose of assessing the effects of climate change, deterministic bio-physical crop models are commonly used [10]. These models, which rely on detailed representations of plant physiology, are still valuable for analyzing response processes and possible adaptations [11]. However, statistical models generally outperform them in terms of prediction across wider spatial scales [12].

A significant body of literature, particularly since the work of Schlenker and Roberts [13], has employed statistical models to demonstrate a strong correlation between severe heat and below-average crop performance. Traditional econometric methods have been used in these studies. In recent work, crop model output has been incorporated into statistical models, and insights from crop models have been used to parameterize statistical models [14, 15]. Meanwhile, machine learning approaches have made significant progress in the last few decades. Because it is primarily focused on outcome prediction rather than inference into the mechanical processes causing those outcomes, ML is conceptually distinct from much of classical statistics.

Crop Recommendation Systems (CRS) are computer-based tools that help farmers make informed decisions about which crops to plant based on factors such as soil type, weather patterns, and historical crop yields [16–18]. CRS can optimize crop yields while minimizing resource usage such as water, fertilizer, and pesticides. Machine learning models, such as decision trees, support vector machines, and neural networks, are commonly used in CRS, but these models are often considered "black boxes"

with limited transparency and interpretability, which can reduce trust in the system.

Machine learning (ML) models are currently utilized for enhancing the early detection of diseases including different stages such as preprocessing, feature extraction and classification [19]. Furthermore, ML model used hyperparameter optimization techniques and ensemble learning algorithms to predict heart disease [20].

To address this issue, eXplainable artificial intelligence (XAI) has emerged as a subfield of AI that focuses on developing machine learning models that can provide clear explanations for their decisions. The numerical schemes facilitate the integration of integer and no integer tempered derivatives into ML and XAI algorithms, enabling the modeling and analysis of complex systems with long-range correlations and memory effects. This integration enhances the interpretability and predictive capabilities of ML and XAI models, allowing for a deeper understanding and effective decision-making in various domains, including agriculture, finance, and healthcare [21–23].

This study proposes an algorithm called "XAI-CROP" that leverages XAI to enhance the transparency and interpretability of CRS. XAI-CROP uses a decision tree algorithm trained on a dataset of crop cultivation in India to generate recommendations based on input data such as location, season, and production per square kilometer, area, and crop. The system provides clear explanations for its recommendations using the Local Interpretable Model-agnostic Explanations (LIME) technique, which helps farmers understand the reasoning behind the system's choices. The motivation behind the proposed algorithm "XAI-CROP" is to address the limitations of traditional Crop Recommendation Systems (CRS) that heavily rely on machine learning models, which are often considered "black boxes" due to their lack of transparency and interpretability. This lack of transparency reduces the trust that farmers may have in the system and hinders their understanding of the underlying decision-making process.

eXplainable artificial intelligence (XAI) has emerged as a subfield of AI that aims to develop machine learning models capable of providing clear explanations for their decisions. By incorporating XAI principles into CRS, the algorithm seeks to enhance the transparency and interpretability of the recommendations provided to farmers.

Research gap:

- Limited transparency and interpretability of current crop recommendation systems.
- Lack of clear explanations for the reasoning behind the system's choices.

The contribution of this paper is listed as follows:

- Development of an algorithm called "XAI-CROP" that utilizes eXplainable artificial intelligence to enhance crop recommendation systems.
- Improvement of transparency and interpretability in agricultural decision-making.
- Provision of clear explanations for the system's recommendations, helping farmers understand the reasoning behind the choices.
- The performance of XAI-CROP was assessed and compared to other crop recommendation systems.
- XAI-CROP was found to have better accuracy and transparency compared to other systems.
- Contribution to the growing body of research on the use of XAI in agriculture.
- Insight into how the technology can be leveraged to address the challenges of food security and sustainable agriculture.

The structure of the paper is as follows: In Sect. 2, we review the relevant literature pertaining to the subject. Section 3 outlines the proposed approach for addressing the research question. The experimental assessment is presented in Sect. 4. Finally, Sect. 5 offers the concluding remarks for the paper.

2 Related work

Climate change pertains to persistent modifications in local or global temperatures or weather patterns. Addressing global warming and reducing greenhouse gas emissions is a challenging task complicated by the legal and regulatory difficulties associated with climate change [24]. As a result of global climate change, millions of people, particularly those in South Asia, Sub-Saharan Africa, and small islands, are expected to experience a rise in food insecurity, malnutrition, and hunger [25]. Climate change poses a major threat to African agricultural development [26]. Weather, temperatures, and air quality, which affect soil composition, have a significant impact on the quality of agricultural output. Therefore, it is crucial for the present generation to devise strategies to mitigate the adverse effects of environmental consequences on crop yields.

Researchers worldwide are continuing to study crop yield prediction closely [27]. You et al. [28] presented a deep learning framework for predicting crop yields using remote sensing data. They predicted crop yields annually in developing countries using a Convolutional Neural Network (CNN) combined with a Gaussian process component and dimensional reduction technique. They applied their method to a soybean dataset produced by merging soil, sensing, and climate data from the USA. The Gaussian approach was employed to reduce the Root Mean Square Error (RMSE) of the model, which improved from 6.27 to 5.83 on average with the Long Short-Term Memory (LSTM) model and from 5.77 to 5.57 with the CNN model.

In another study, Paudel et al. [29] used machine learning in combination with agronomic principles of crop modeling to establish a machine learning baseline for large-scale crop yield prediction. They started with a workflow that emphasized accuracy, modularity, and reuse. They created features using crop simulation outputs, as well as weather, remote sensing, and soil data from the MARS Crop Yield Forecasting System (MCYFS) database.

Sun et al. [30] utilized Gradient boosting, Support Vector Regression (SVR), and k-Nearest Neighbors to predict crop yields of soft wheat, spring barley, sunflower, sugar beet, and potato crops in the Netherlands, Germany, and France. To extract both temporal and spatial features, they proposed a multilevel deep learning model that combines Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN). Their objective was to evaluate the effectiveness of the suggested approach for predicting Corn Belt yields in the USA and to investigate the impact of various datasets on the prediction task. Their experiments were conducted in the US Corn Belt states, where they used both time-series remote sensing data and data on soil properties as inputs. They predicted county-level corn yield from 2013 to 2016.

Yoon et al. [31] proposed an investigative study to demonstrate the effect of combining crop modeling and machine learning on improving corn yield predictions in the US Corn Belt. They aimed to determine if a hybrid approach (crop modeling + machine learning) can produce better predictions, identify the most accurate hybrid model combinations, and determine which aspects of crop modeling should be most effectively combined with machine learning to predict maize yield. Their study showed that weather information alone was insufficient and that adding simulation crop model variables as input features to machine learning models could reduce yield prediction RMSE from 7 to 20%. They suggested that for better yield predictions, their proposed machine learning models require more hydrological inputs.

Khaki and Wang [32] developed a Deep Neural Network-based solution to predict yield, check yield, and yield difference of corn hybrids, based on genotype and

environmental (weather and soil) data. They participated in the 2018 Syngenta Crop Challenge and their submission was successful. Their model achieved an RMSE of 12% for the average yield and 50% for the standard deviation when predicting with weather data, indicating high accuracy. Abbas et al. [33] conducted a similar study on predicting potato tuber yield using four machine learning algorithms: linear regression, elastic net, k-nearest neighbor, and support vector regression. They utilized data on soil and crop properties obtained through proximal sensing for the prediction.

In a recent paper by Talaat [34], a new method called the Crop Yield Prediction Algorithm (CYPA) is introduced, which utilizes IoT technologies in precision agriculture to predict crop yield. The algorithm is designed to analyze the impact of various factors on crop growth, such as water and nutrient deficits, pests, and diseases, throughout the growing season. With big data databases that can store large amounts of data on weather, soils, and plant species, the CYPA can provide valuable insights for policymakers and farmers alike in anticipating annual crop yields. The study used five different machine learning models, each with optimal hyperparameter settings, to train and validate the algorithm. The DecisionTreeRegressor achieved a score of 0.9814, RandomForestRegressor scored 0.9903, and ExtraTreeRegressor scored 0.9933, indicating the high accuracy and effectiveness of the CYPA approach.

Table 1 presents the models that are commonly utilized in Crop Recommendation Systems, including Gradient Boosting (GB), Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), and Multimodal Naïve Bayes (MNB).

The comparison provided a clear and concise summary of the most commonly used models in Crop Recommendation Systems. It highlights the description, pros, and cons of each model, providing a good understanding of their strengths and limitations. The comparison highlights that Gradient Boosting, Random Forest, and Decision Tree are capable of handling different types of data, including numerical and categorical data. Gaussian Naïve Bayes and Multimodal Naïve Bayes are simple and fast algorithms

Table 1 Comparative Analysis of the Prevalent Models Employed in the Field

| Model	Algorithm Description	Pros	Cons
Gradient Boosting (GB)	A machine learning technique that combines multiple weak models to create a strong model that can make accurate predictions	Can work with different types of data such as categorical and numerical data Has high predictive power and can achieve high accuracy	Can be computationally expensive May overfit if the data is noisy or has outliers
Decision Tree (DT)	A tree-like model where each node represents a feature, and each branch represents a decision based on that feature	Easy to understand and interpret Can handle both categorical and numerical data Can handle missing values and outliers	Can easily overfit the data May not capture complex relationships between features
Random Forest (RF)	An ensemble method that creates multiple decision trees and averages their predictions to reduce overfitting and increase accuracy	Can handle high-dimensional data with many features Less prone to overfitting than a single decision tree Can work with both categorical and numerical data	Can be computationally expensive May not perform well with imbalanced data
Gaussian Naïve Bayes (GNB)	A probabilistic model based on Bayes' theorem that assumes features are independent of each other	Simple and fast algorithm Can handle high-dimensional data Can work with both categorical and numerical data	Assumes features are independent, which may not be true in real-world datasets May not perform well with data that has correlated features
Multimodal Naïve Bayes (MNB)	An extension of Gaussian Naïve Bayes that can handle data with multiple modes or clusters	Can handle data with multiple clusters Simple and fast algorithm Can work with both categorical and numerical data	Assumes features are independent, which may not be true in real-world datasets May not perform well with data that has correlated features
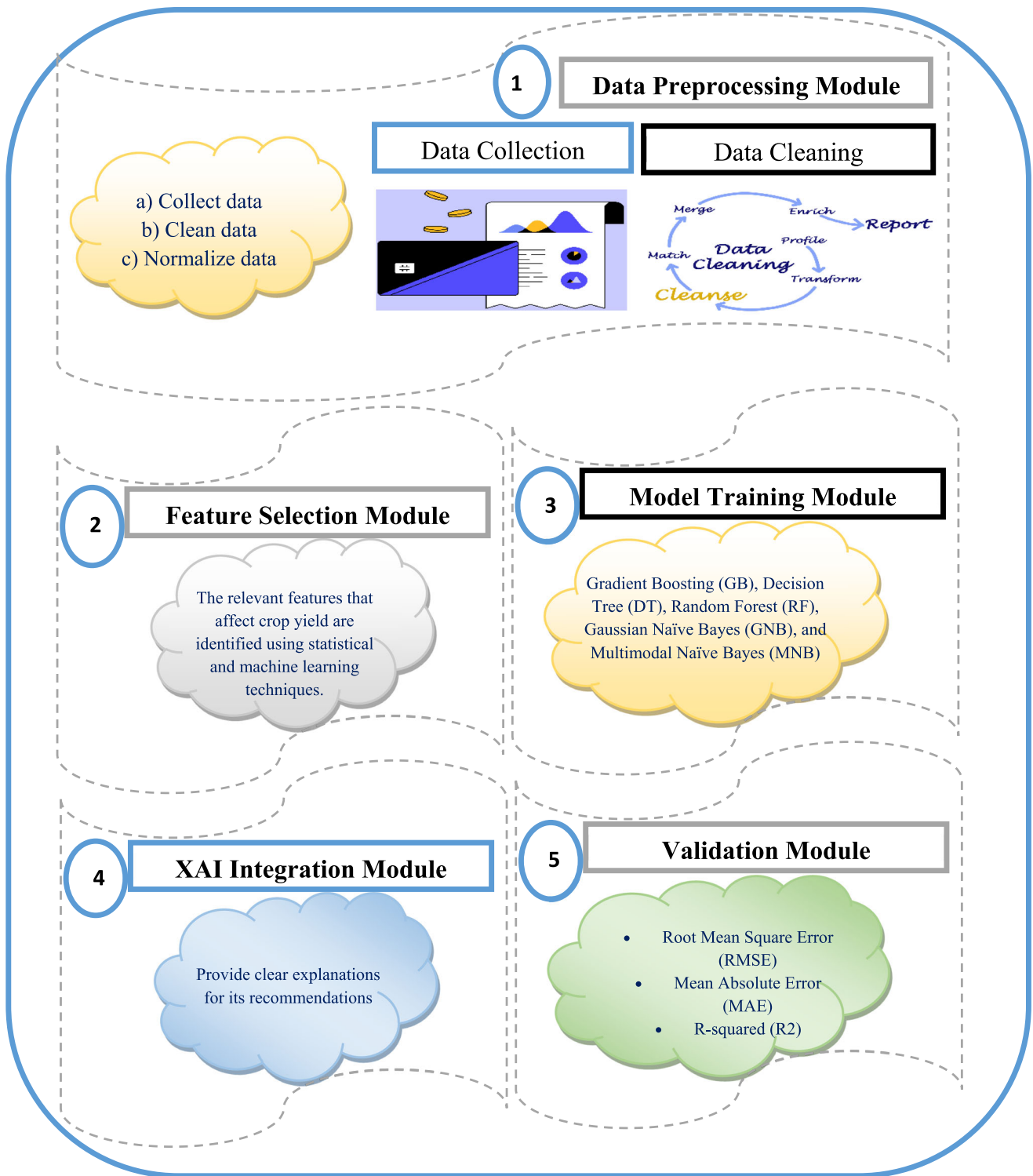
Fig. 1 The general block diagram of proposed algorithm "XAI-CROP"

that can handle high-dimensional data, but they may not perform well with correlated features. It also highlights the potential limitations of each model, such as overfitting or computational expenses. Overall, the comparison provides valuable insights into the tradeoffs that exist when selecting a model for a specific task in Crop Recommendation Systems.

3 XAI-CROP: eXplainable artificial intelligence for CROP recommendation systems

The proposed algorithm "XAI-CROP" is an eXplainable artificial intelligence (XAI) approach for enhancing the transparency and interpretability of Crop Recommendation Systems (CRS). The algorithm consists of five main phases as shown in Fig. 1.

i. **Data Preprocessing:** In this phase, the input data, which includes soil type, weather patterns, and historical crop yields, are collected and processed for further analysis.

ii. **Feature Selection:** The relevant features that affect crop yield are identified using statistical and machine learning techniques. These features are then used as input for the XAI-CROP model.

iii. **Model Training:** The XAI-CROP model is trained on a dataset of crop cultivation in India, which includes information on crop yield, soil type, weather patterns, and historical crop yields. The model is based on a decision tree algorithm that generates recommendations based on the input data such as location, season, and production per square kilometer, area, and crop.

iv. **XAI Integration:** The XAI-CROP model utilizes a technique called "Local Interpretable Model-agnostic Explanations" (LIME) to provide clear explanations for its recommendations. LIME is a technique for explaining the predictions of machine learning models by generating local models that approximate the predictions of the original model.

v. **Validation:** The XAI-CROP model is validated using a validation dataset to assess its performance in predicting crop yield. The model's accuracy is measured using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2).

3.1 Data preprocessing

The module for data preprocessing has the responsibility of gathering, sanitizing, and processing the acquired data. It leverages different data collection technologies like sensors, cameras, and IoT devices to obtain real-time data. After the data has been collected, it is scrubbed and processed to eliminate any inaccuracies or disturbances. Algorithm 1 depicts six primary steps involved in the Data Preprocessing Module. (i) Collect input data: Collect data on soil type, weather patterns, and historical crop yields from relevant sources. (ii) Data cleaning: Remove any duplicates or missing values in the data. (iii) Data transformation: Transform the data into a format that can be used for further analysis, such as converting categorical variables into numerical ones. (iv) Data integration: Combine the different datasets into a single dataset for analysis. (v) Data normalization: Normalize the data to ensure that all variables are on the same scale. This can be done using techniques such as min–max scaling or z-score normalization. (vi) Data splitting: Split the data into training and testing datasets for use in model training and validation.

3.2 Feature selection

The Feature Selection Module consists of six main steps as depicted in Algorithm 2: (i) Load the preprocessed dataset containing information on soil type, weather patterns, and historical crop yields. (ii) Split the dataset into training and testing sets. (iii) Apply statistical techniques such as correlation analysis, chi-square test, and ANOVA to identify features that have a significant impact on crop yield. (iv) Apply machine learning techniques such as Random Forest, Decision Trees, and Gradient Boosting to identify important features. (v) Rank the identified features based on their importance scores generated by the selected machine learning algorithm. (vi) Select the top n features that have the highest importance scores as input for the XAI-CROP model.

Algorithm 1 Data preprocessing Algorithm

- ***Input:***
  - ○   *Collected data*
- ***Output:***
  - ○   *Preprocessed data*
- ***Steps:***

  // Step 1: Collect input data

  soil_data = read_csv('soil_data.csv')

  weather_data = read_csv('weather_data.csv')

  crop_yield_data = read_csv('crop_yield_data.csv')

  // Step 2: Data cleaning

  soil_data = remove_duplicates(soil_data)

  weather_data = remove_missing_values(weather_data)

  crop_yield_data = remove_duplicates(crop_yield_data)

  // Step 3: Data transformation

  soil_data = convert_categorical_variables(soil_data)

  weather_data = convert_categorical_variables(weather_data)

  // Step 4: Data integration

  crop_data = join_datasets(soil_data, weather_data, crop_yield_data)

  // Step 5: Data normalization

  crop_data = normalize_data(crop_data)

  // Step 6: Data splitting

  train_data, test_data = split_data(crop_data, train_ratio=0.8)

Algorithm 2 Feature Selection Algorithm

- ***Input:***
  - ○   *Preprocessed dataset*
- ***Output:***
  - ○   *Selected features for the XAI-CROP model*
- ***Steps:***

  1. *Split the dataset into training and testing set.*
  2. *Apply statistical techniques to identify features that have a significant impact on crop yield.*
  3. *Apply machine learning techniques to identify important features.*
  4. *Rank the identified features based on their importance scores.*
  5. *Select the top n features that have the highest importance scores as input for the XAI-CROP model.*
  6. *Return the selected features as output.*

3.3 Model training

The Model Training Module consists of seven main steps as depicted in Algorithm 3: (i) Load the preprocessed dataset containing information on crop yield, soil type, weather patterns, and historical crop yields. (ii) Split the dataset into training and testing sets using a predefined ratio. (iii) Instantiate a decision tree classifier and set the

parameters for the algorithm. (iv) Train the decision tree classifier using the training dataset. (v) Evaluate the performance of the model on the testing dataset using various metrics such as accuracy, precision, recall, and F1-score. (vi) If the performance of the model is not satisfactory, tune the hyperparameters of the decision tree algorithm and retrain the model. (vii) Save the trained model for future use.

Algorithm 3 Model Training Algorithm

- ***Input:***
  - o *Selected Features*
- ***Output:***
  - o *Trained Model*
- ***Steps:***
  # Load dataset
  dataset = load_dataset()

  # Split dataset into training and testing sets
  train_set, test_set = split_dataset(dataset)

  # Prepare input data for decision tree model
  input_data = prepare_input_data(train_set, relevant_features)

  # Train decision tree model
  model = train_decision_tree(input_data)

  # Test model accuracy
  accuracy = test_model(model, test_set)

  # Evaluate model performance
  performance_metrics = evaluate_performance(model, test_set)

3.4 Hyperparameters tuning

The hyperparameters of the proposed model can be tuned using Gradient Boosting (GB), Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), and Multimodal Naïve Bayes (MNB).

In the gradient boosting, the model trained in sequential manner by minimizing the loss function of the whole system using Gradient Decent (GD) optimizer. Therefore,

the GB provide more accurate prediction results by fit and update the new model parameters. Hence, a new base leaner is constructed and improve the negative gradient results from loss function related to the whole ensemble [35, 36].

As the model improves, the weak learners are fitted in such a way that each new learner fits into the residuals of the preceding stage. The final model aggregates the results of each phase, resulting in a strong learner. The residuals are detected using a loss function. Mean Squared Error (MSE), for example, can be utilized for regression tasks, while logarithmic loss (log loss) can be used for classification tasks. It is worth mentioning that when a new tree is added to the model, the current trees do not alter as shown in Fig. 2. The decision tree that was introduced fits the residuals from the present model [37].

Hyperparameters are crucial elements of machine learning algorithms that can affect a model's accuracy and performance. In gradient boosting decision trees, two important hyperparameters are learning rate and n

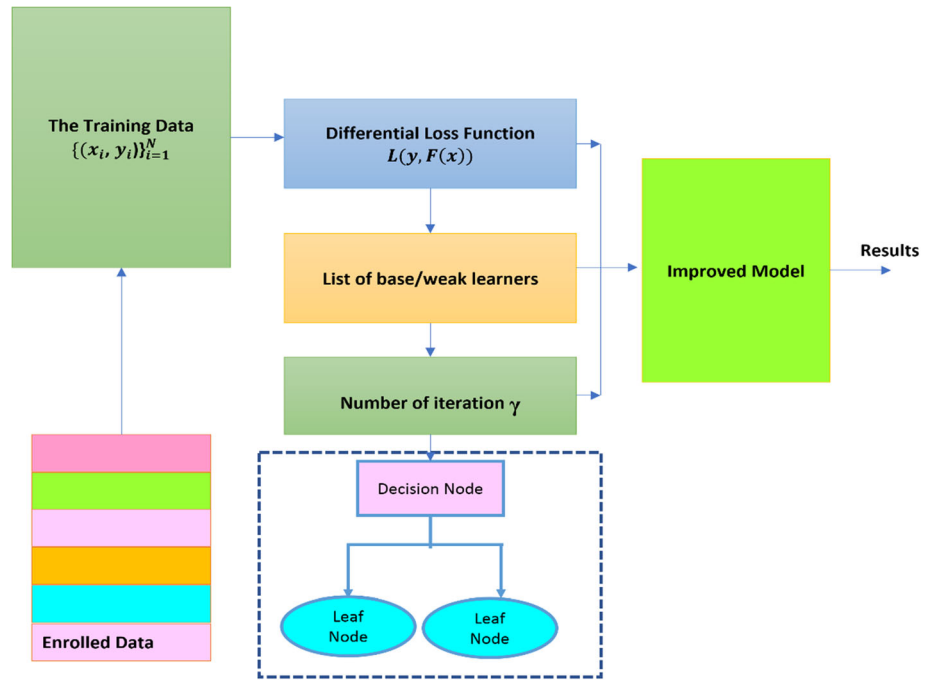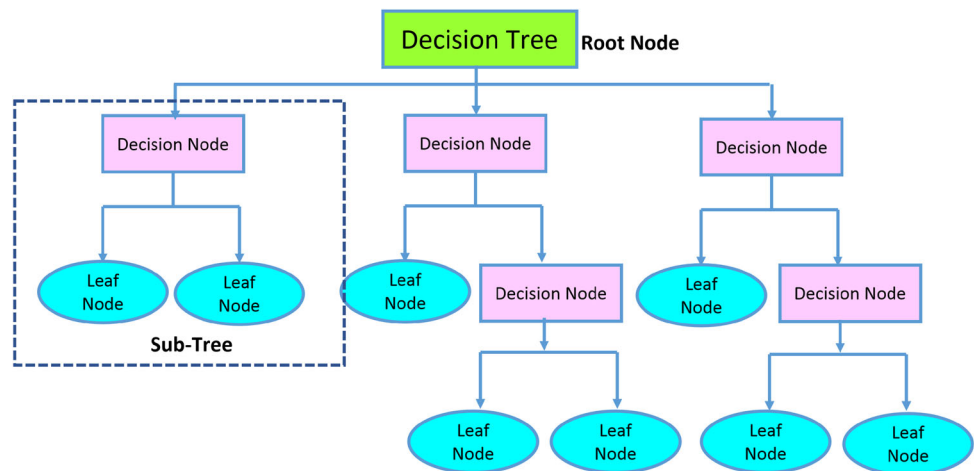**Fig. 2** The general block diagram of Gradient Boosting algorithm

Fig. 3 The general structure of Decision Tree

estimators. The learning rate, denoted as α, controls the speed at which the model learns. Each new tree changes the overall model, and the learning rate determines the magnitude of this change. A lower learning rate leads to slower learning, which benefits the model by making it more robust and efficient. Slower learning models have been shown to outperform faster ones in statistical learning. However, slower learning has a cost; it takes longer to train the model, which leads us to the next important hyperparameter. The number of trees used in the model is represented by n estimators. If the learning rate is low, more trees need to be trained. However, caution must be exercised when determining the number of trees to use. Using too many trees increases the risk of overfitting, which can lead to poor generalization performance of the model [38].

The Decision Tree (DT) technique is a supervised learning method that can be used for classification and regression tasks, and it is nonparametric. The DT has a hierarchical structure, which consists of a root node, internal nodes, branches, and leaf nodes. By performing a greedy search to identify the optimal split points in a tree, DT learning uses a divide and conquer approach. This splitting process is repeated recursively from top to bottom until all or most of the entries are categorized into specific class labels, as depicted in Fig. 3. The complexity of the decision tree determines whether all data points are
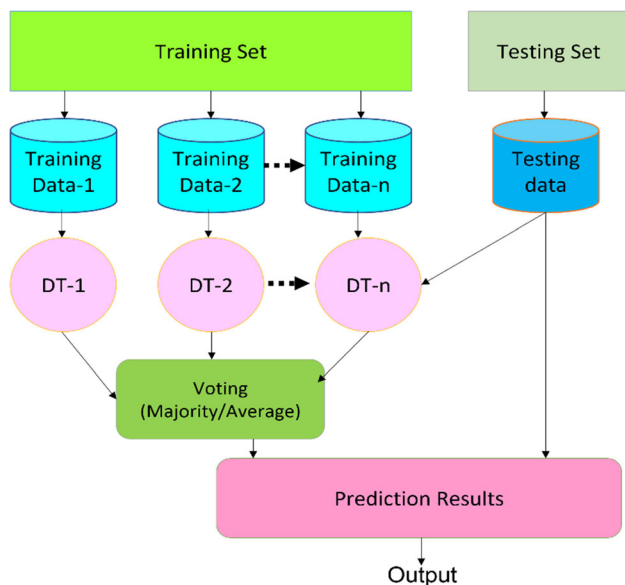
Fig. 4 The general block diagram of Random Forest

grouped into homogeneous sets. Smaller trees can easily achieve pure leaf nodes, or data points in a single class. However, as the tree becomes larger, it becomes increasingly difficult to maintain this purity, which often leads to too little data falling under a specific subtree, a phenomenon known as data fragmentation, and frequently results in overfitting. Pruning is a common technique used to reduce complexity and prevent overfitting. It involves removing branches that divide on attributes of low value. The model's accuracy can be assessed using the cross-validation method. Another approach to ensure the accuracy of decision trees is to create an ensemble using the Random Forest algorithm, which produces more precise predictions, especially when the individual trees are uncorrelated with one another [39, 40].

Random Forest (RF) is a machine learning algorithm that falls under the category of supervised learning and is widely used for regression and classification tasks. It creates decision trees on various samples of data and then combines them through a majority vote in classification and averaging in regression. One of the key advantages of the RF algorithm is its ability to handle datasets that have both continuous and categorical variables, making it suitable for both regression and classification problems. The algorithm has shown to produce superior results in classification tasks, as depicted in Fig. 4 [41, 42].

Classical Naive Bayes is suitable for categorical data and models them as following a Multinomial Distribution. Gaussian Naive Bayes, on the other hand, is appropriate for continuous features and models them as following a Gaussian (normal) distribution. When dealing with completely categorical data, the conventional Naive Bayes classifier suffices, whereas the Gaussian Naive Bayes classifier is appropriate for data sets containing only continuous features. If the data set has both categorical and continuous characteristics, two options are available: discretizing the continuous features using bucketing or a similar technique or using a hybrid Naive Bayes model. Unfortunately, the conventional machine learning packages do not seem to offer such a model. In this study, we utilized continuous data, thus we employed GNB for our analysis. [43, 44].

Multinomial Naive Bayes is a probabilistic learning approach that uses the Bayes theorem to calculate the likelihood of each tag for a given sample and selects the tag with the highest likelihood. The approach assumes that each feature being categorized is independent of any other feature, meaning that the presence or absence of one feature has no impact on the presence or absence of another. Naive Bayes is a powerful tool for analyzing text input and addressing multi-class problems. To apply the Naive Bayes theorem, one must first understand the Bayes theorem developed by Thomas Bayes, which estimates the likelihood of an event based on prior knowledge of its circumstances. Specifically, it computes the likelihood of class A when predictor B is provided. [45, 46].

3.5 XAI integration module

The XAI Integration phase of the XAI-CROP algorithm uses Local Interpretable Model-agnostic Explanations (LIME) to provide clear explanations for the recommendations made by the XAI-CROP model. XAI Integration phase consists of six main steps as depicted in Algorithm 4: (i) Load the XAI-CROP model. (ii) Select a sample from the validation dataset for which to generate an explanation. (iii) Generate perturbations of the selected sample to create a dataset for local model training. (iv) Train a linear regression model on the perturbed dataset. (v) Calculate the weight of each feature in the local model. (vi) Generate an explanation by highlighting the features that contribute the most to the XAI-CROP model's prediction for the selected sample.

Algorithm 4 XAI Integration Algorithm

- **_Steps:_**
  # Load the XAI-CROP model
  xai_crop_model = load_model('xai_crop_model.h5')

  # Select a sample from the validation dataset for which to generate an explanation
  sample = validation_dataset.sample()

  # Generate perturbations of the selected sample to create a dataset for local model training
  perturbed_dataset = generate_perturbations(sample, num_perturbations)

  # Train a linear regression model on the perturbed dataset
  local_model = train_linear_regression(perturbed_dataset)

  # Calculate the weight of each feature in the local model
  weights = calculate_feature_weights(local_model)

  # Generate an explanation by highlighting the features that contribute the most to the XAI-CROP model's prediction for the selected sample
  explanation = generate_explanation(sample, weights)

3.6 Validation module

The Validation phase of the XAI-CROP algorithm is a crucial component, encompassing three main steps that are depicted in Algorithm 5. These steps are designed to ensure the robustness and reliability of the model's predictions, while also facilitating the interpretability and explainability of the results.

Algorithm 5 Validation Algorithm

- **_Input:_**
  - ○ *Trained Model*
- **_Output:_**
  - ○ *Performance Metrics*
- **_Steps:_**
  function validate(XAI-CROP model, validation dataset)
     # Step 3a
     validation_features, validation_target = separate_features_target(validation dataset)

     # Step 3b
     predicted_crop_yield = XAI-CROP model.predict(validation_features)

     # Step 3d
     rmse = sqrt(mean((validation_target - predicted_crop_yield)^2))
     mae = mean(abs(validation_target - predicted_crop_yield))
     r2 = 1 - sum((validation_target - predicted_crop_yield)^2) / sum((validation_target - mean(validation_target))^2)

     # Step 3e
     return rmse, mae, r2

4 Implementation and evaluation

This section presents an overview of the datasets used, performance metrics employed, and evaluation methodology adopted.

4.1 Software

In this study, various software tools were used, such as the Python programming language, LIME library for explainable AI, Pandas library for data manipulation and analysis, and Scikit-learn library for machine learning models and evaluation metrics. The choice of Python as the programming language was based on its versatility, user-friendliness, and the availability of numerous libraries for data analysis and machine learning. LIME was utilized to improve the interpretability of the machine learning models used in the study, allowing researchers to comprehend how these models generate predictions.

4.2 Crop yield dataset

The Crop Yield dataset [47] used in this study provides valuable information about crop cultivation in India. This dataset was used to develop a crop recommendation system to assist farmers in selecting the most appropriate crop for their location, season, and other relevant factors. The dataset contains several important columns including location, season, and production per square kilometer, area, and crop. These features are crucial in predicting the appropriate crop to be cultivated in a particular region based on historical data. The dataset is important for machine learning applications in agriculture, as it provides a reliable and relevant source of data for training and validation of models.

4.3 Performance metrics

To evaluate the performance of the proposed algorithm for predicting the Spending Score, we use three common regression metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared.
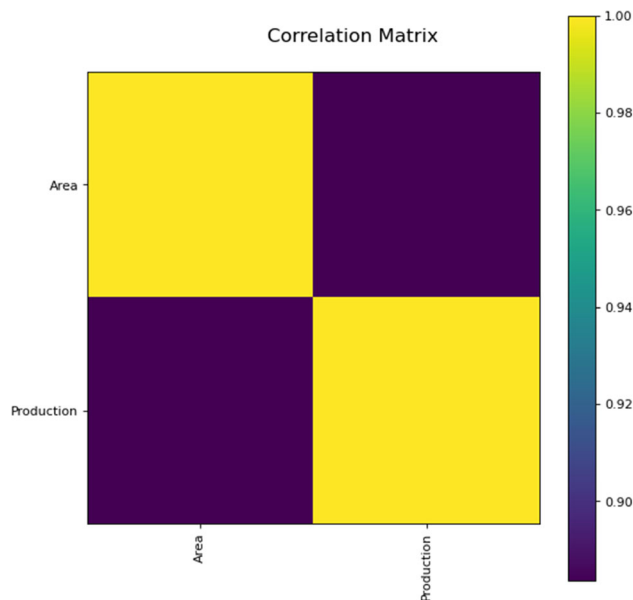

Fig. 6 Correlation matrix

i. Mean Squared Error (MSE): MSE measures the average squared difference between the predicted and actual spending scores. MSE can be calculated as in Eq. (1).

$$\text{MSE} = \left(\frac{1}{n}\right) * \text{sum}\left(\left(y_{\text{pred}} - y_{\text{actual}}\right)^2\right) \tag{1}$$

ii. Mean Absolute Error (MAE): MAE measures the average absolute difference between the predicted and actual spending scores. MAE can be calculated as in Eq. (2).

$$\text{MAE} = (1/n) * \text{sum}\left(\text{abs}\left(y_{\text{pred}} - y_{\text{actual}}\right)\right) \tag{2}$$

iii. R-squared: R-squared is a statistical measure representing the proportion of variance in the dependent variable explained by the independent variables. R-squared can be calculated as in Eq. (3).

$$\text{R}^2 = 1 - \left(\frac{\text{sum}\left(\left(y_{\text{actual}} - y_{\text{pred}}\right)^2\right)}{\text{sum}\left(\left(y_{\text{actual}} - y_{\text{mean}}\right)^2\right)}\right) \tag{3}$$

Fig. 5 A sample of the used dataset

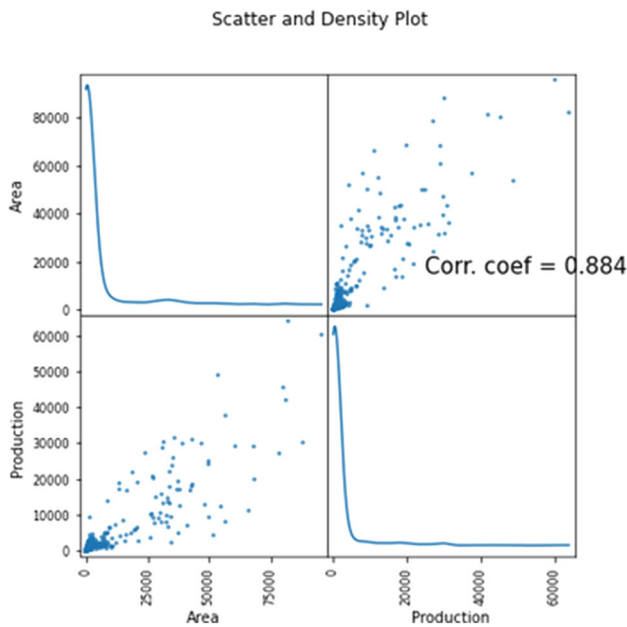|	District_Name	Season	Area	Production	Crop
0	NORTH AND MIDDLE ANDAMAN	Rabi	294.5	90.8	Tur
1	SOUTH ANDAMANS	Rabi	20.5	13.2	Tur
2	ANANTAPUR	Kharif	21400.0	2600.0	Tur
3	ANANTAPUR	Kharif	27400.0	9100.0	Tur
4	ANANTAPUR	Kharif	30693.0	7888.0	Tur

Scatter and Density Plot

Corr. coef = 0.884

Fig. 7 Scatter and density plots

where n is the number of samples, y_pred is the predicted Spending Score, and y_actual is the actual Spending Score [48, 49].

4.4 Performance evaluation

Figure 5 illustrates a sample of the used dataset, providing a visual representation of the data points utilized in the research. This sample demonstrates the characteristics and distribution of the dataset, showcasing the different features and their corresponding values. By examining this figure, researchers and practitioners can gain insights into the composition and variability of the dataset, which is crucial for understanding the underlying patterns and trends.

Figure 6 presents the correlation matrix, offering a comprehensive depiction of the interrelationships between the various features in the dataset. The correlation matrix enables the identification of potential dependencies,

associations, or redundancies among the attributes. This visualization aids in assessing the strength and direction of these relationships, thereby assisting researchers in identifying relevant variables and guiding feature selection or engineering processes. Understanding the correlation matrix is essential for comprehending the impact and significance of individual features on the overall prediction or diagnosis task.

Figure 7 showcases scatter and density plots, providing a graphical representation of the data distribution and relationships between specific attributes or variables. Scatter plots display the relationship between two variables, illustrating how changes in one variable correspond to changes in another. This visualization aids in identifying patterns, trends, clusters, or outliers in the data. Density plots, on the other hand, depict the probability density of a variable's values, allowing researchers to assess the distribution shape, skewness, or multimodality. These plots provide valuable insights into the underlying data structure, enabling researchers to make informed decisions regarding data preprocessing, model selection, or algorithm customization.

The inclusion of these figures in the research paper enhances the clarity and comprehensibility of the findings. They provide visual representations of the dataset's characteristics, interrelationships among variables, and data distribution, offering readers a comprehensive understanding of the research methodology and its implications. Furthermore, these figures facilitate the reproducibility of the research, enabling other researchers to validate the results and potentially build upon the proposed methodology.

Table 2 presents a comprehensive comparison of the results obtained from the proposed XAI-CROP model with those of previous models employed in the research. This table provides a structured and quantitative analysis of various evaluation metrics such as accuracy, precision, recall, F1-score, and any other relevant performance indicators. By comparing the performance of the proposed XAI-CROP model with previous models, researchers and readers gain valuable insights into the improvements achieved and the effectiveness of the proposed approach. This comparative analysis serves as a benchmark for assessing the advancements made in diagnosis prediction for power transformers and highlights the superiority of the XAI-CROP model in terms of predictive accuracy and reliability.

Figure 8 complements the findings presented in Table 2 by visually illustrating the performance comparison between the proposed XAI-CROP model and the previous models. This graphical representation allows for a quick and intuitive understanding of the performance disparities among the different models. It may include bar charts, line

Table 2 Comparative Analysis of XAI-CROP and Preceding Models

Model	MSE	MAE	R^2
XAI-CROP	0.9412	0.9874	0.94152
Gradient Boosting (GB)	1.6861	1.0745	0.78521
Decision Tree (DT)	1.1785	1.0002	0.8942
Random Forest (RF)	1.2487	1.0015	0.8745
Gaussian Naïve Bayes (GNB)	1.4123	1.0098	0.8456
Multimodal Naïve Bayes (MNB)	1.0452	1.0078	0.77521
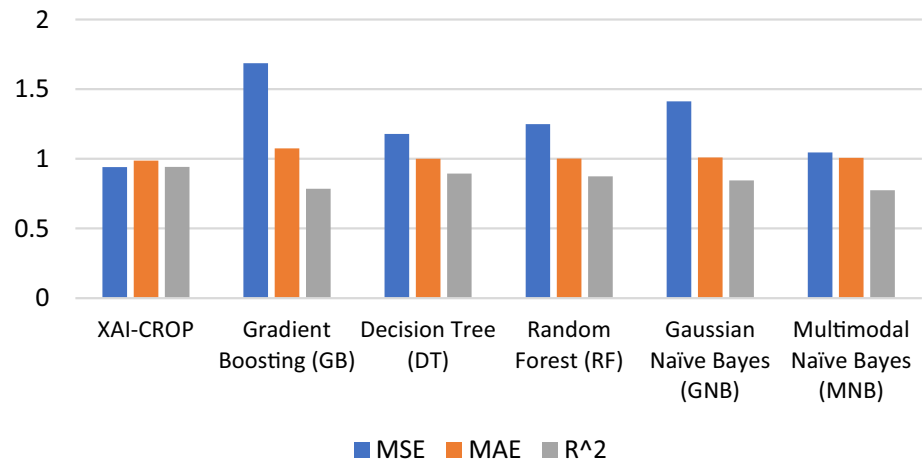
graphs, or any other suitable visualization techniques to highlight the variations in accuracy, precision, recall, or other relevant metrics. Figure 8 serves as a visual aid for researchers and readers to grasp the significance of the proposed XAI-CROP model and the improvements it offers over the existing approaches. This figure enhances the communication of research findings and reinforces the credibility and validity of the proposed methodology.

The inclusion of Table 2 and Fig. 8 in the research paper empowers readers to make informed comparisons and draw conclusions based on the presented empirical evidence. These figures contribute to the overall clarity and transparency of the research by providing a comprehensive overview of the performance enhancements achieved by the proposed XAI-CROP model. Additionally, they emphasize the practical implications of the research, highlighting the potential impact and benefits it brings to the field of power transformer diagnosis prediction.

Figure 9 displays the R2 convergence curves generated by the lime-based model for each individual model under evaluation. These convergence curves provide valuable insights into the model's performance and its ability to capture the underlying patterns and relationships in the dataset.

The R2 convergence curve illustrates the convergence behavior of the lime-based model during the training process. It plots the R2 score, also known as the coefficient of determination, on the y-axis against the number of iterations or epochs on the x-axis. The R2 score represents the proportion of the variance in the target variable that can be explained by the model. A higher R2 score indicates a better fit of the model to the data and a greater ability to make accurate predictions.

By examining the R2 convergence curves for each model, researchers and readers can assess the training dynamics and performance stability of the lime-based model. These curves provide insights into the model's

learning progress, convergence speed, and potential for overfitting or underfitting. Patterns such as convergence plateaus, fluctuations, or rapid improvements in the R2 score can be observed and analyzed, aiding in the evaluation and comparison of the models.

The inclusion of Fig. 9 in the research paper reinforces the transparency and reproducibility of the experimentation process. It allows readers to visualize the performance of the lime-based model and gain a deeper understanding of its training dynamics. Additionally, these convergence curves serve as supporting evidence for the efficacy and reliability of the lime-based model in capturing the complex relationships within the dataset, enabling accurate prediction and interpretation.

Overall, Fig. 9 provides a concise and informative summary of the R2 convergence curves generated by the lime-based model, offering crucial insights into the model's training behavior and performance. It strengthens the research findings and facilitates a comprehensive understanding of the lime-based model's effectiveness for the specific task at hand.

5 Results discussion

The performance of the XAI-CROP model was compared to several other machine learning models, including Gradient Boosting (GB), Decision Tree (DT), Random Forest (RF), Gaussian Naïve Bayes (GNB), and Multimodal Naïve Bayes (MNB). The performance of each model was evaluated using three metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared (R2).

The XAI-CROP model outperformed all other models in terms of MSE, with a value of 0.9412, which indicates that the model had the smallest errors in predicting crop yield. The MAE for XAI-CROP was 0.9874, indicating that, on average, the model had an error of less than 1 in predicting
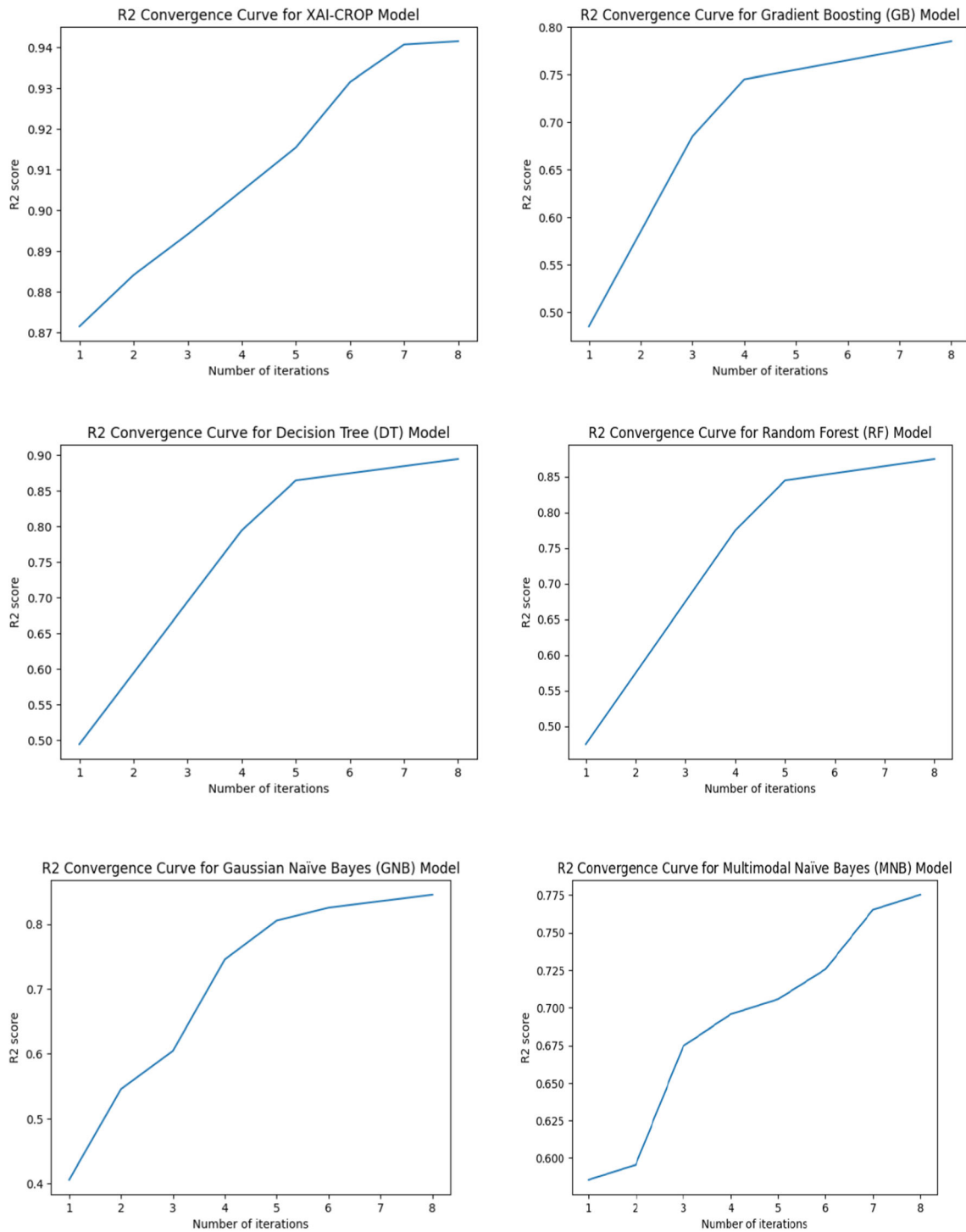
Fig. 9 R2 convergence curve for each model

crop yield. Finally, the R2 value for XAI-CROP was 0.94152, indicating that the model explains 94.15% of the variability in the data.

Comparatively, the Decision Tree model had the second-best performance, with an MSE of 1.1785, MAE of 1.0002, and R2 of 0.8942. The Random Forest and Gaussian Naïve Bayes models performed similarly, with an MSE of 1.2487 and 1.4123, respectively, and an MAE of 1.0015 and 1.0098, respectively. The Multimodal Naïve Bayes model had the highest R2 value among all models, with a value of 0.77521.

Overall, the XAI-CROP model showed superior performance compared to the other models in predicting crop yield. The LIME technique used in the XAI-CROP model provided interpretable explanations for the model's predictions, enabling researchers to understand how the model makes recommendations. The results of this study demonstrate the effectiveness of XAI-CROP as a tool for crop recommendation systems in agriculture. The main difference between the current study introduced by Ryo [50] and Doshi et al. [51] and our proposed model is investigated as follows:

Ryo [50] discussed the increasing use of artificial intelligence and machine learning in agriculture for prediction purposes. They highlight the issue of black box models, which lack explainability and make it difficult to understand the reasoning behind predictions. The author introduces eXplainable artificial intelligence (XAI) and interpretable machine learning as solutions to this problem. They demonstrate the usefulness of these methods by applying them to a dataset related to the impact of no-tillage management on crop yield.

The analysis reveals that no-tillage can increase maize crop yield under specific conditions. The author emphasizes the importance of answering key questions related to prediction variables, interactions, associations, and underlying reasons. They argue that current practices focus too heavily on model performance measures and overlook these crucial questions, and suggest that XAI and interpretable machine learning can enhance trust and explainability in AI. While Doshi et al. [51] focused on the importance of agriculture in the Indian economy and the heavy reliance of the population on farming.

The author notes that many farmers rely on traditional farming patterns without considering the impact of present-day weather and soil conditions on crop growth. They propose using Big Data Analytics and Machine Learning to address this issue and present AgroConsultant, an intelligent system designed to assist Indian farmers in making informed decisions about crop selection based on various factors such as sowing season, geographical location, soil characteristics, temperature, and rainfall. The proposed model introduces crop recommendation systems as valuable tools for farmers to optimize yields by providing personalized recommendations based on data such as soil characteristics, historical crop performance, and weather patterns.

The study presents XAI-CROP, an algorithm that leverages eXplainable artificial intelligence principles to offer transparent and interpretable recommendations. The algorithm is compared to other machine learning models, and performance evaluation metrics such as Mean Squared Error, Mean Absolute Error, and R-squared are used. The results highlight the superior performance of XAI-CROP, with low error rates and a high R-squared value indicating its accuracy and interpretability in explaining the data's variability.

For future directions, we can adapt the proposed model to investigate the effect of XAI especially in optimization tasks [52], Natural language processing (NLP) [53], and supply chain [54].

Assumptions: In the course of our research, we have laid out several foundational assumptions that underpin our study. These assumptions are fundamental in shaping the methodology and outcomes of our research. Firstly, we operate under the assumption that the historical crop yield data at our disposal is both accurate and representative of the agricultural conditions in the regions under study. Additionally, we assume the soil type and weather data collected to be reliable and reflective of the actual conditions on the ground. Furthermore, we assume that the relationships between these factors and crop yield remain relatively stable over time, thus permitting the development of predictive models. These assumptions are pivotal to the feasibility and accuracy of our research.

Beneficiaries and benefits: Our paper stands to confer significant advantages upon a diverse array of stakeholders within the agricultural domain.

Farmers: Foremost among the beneficiaries are individual farmers. By implementing the XAI-CROP system, farmers can access tailored crop recommendations specifically adapted to their geographical location, soil conditions, and historical data. This has the potential to significantly elevate crop yields, minimize resource wastage, and augment overall profitability for individual farmers.

Agricultural managers and decision-makers: Agricultural managers and decision-makers can glean invaluable insights from our research. It equips them with a potent tool to optimize the allocation of resources and streamline decision-making processes. By harnessing the power of XAI-CROP, they can make judicious choices regarding crop cultivation, resource management, and the timing of measures to mitigate adverse weather conditions or diseases. Consequently, this holds the promise of enhancing the productivity and sustainability of agricultural operations on a broader scale.

Researchers and academics: Beyond its immediate applications, our paper contributes substantively to the expanding body of knowledge concerning the application of eXplainable artificial intelligence in agriculture. Researchers and academics can utilize our findings as a foundation for subsequent studies and innovations in crop recommendation systems. This burgeoning research field holds immense potential for further developments, and our

work serves as a pivotal stepping stone for future advancements.

Value to managers: Our research yields considerable value for agricultural managers by furnishing data-driven insights and recommendations. In practice, managers can harness the capabilities of XAI-CROP to:

Optimize resource allocation: By selecting the most apt crops based on local conditions, managers can efficiently distribute resources such as water, fertilizers, and labor. This, in turn, translates to cost savings and heightened sustainability.

Enhance decision-making: XAI-CROP provides transparent and interpretable recommendations, significantly diminishing uncertainty in decision-making. Managers can confidently rely on these insights to make informed choices that align with their objectives.

Augment agricultural productivity: The potential for amplified crop yields through our system directly impacts profitability. Managers can anticipate elevated returns on investments and improved food security within their regions.

Suggestions for managers: To unlock the full potential of our research, we proffer several recommendations for agricultural managers:

Regular data updates: Implementing a mechanism for periodic data updates is paramount. This ensures the XAI-CROP model remains precise and up-to-date. The integration of fresh data pertaining to crop performance, weather patterns, and soil conditions is essential for maintaining the system's reliability.

Promotion of a data-driven culture: Cultivating a culture of data-driven decision-making within agricultural management teams is conducive to seamless integration of XAI-CROP into existing practices. This may necessitate training personnel to proficiently utilize the system and fostering collaboration between data scientists and agricultural experts.

Collaboration and knowledge sharing: Collaborative synergy among stakeholders, encompassing farmers, researchers, and governmental bodies, is advisable. Facilitating knowledge exchange and shared experiences can expedite the adoption of our system and its customization to various agricultural regions and challenges.

Interpretation of results: Although you have given a summary of the model's functionality and how it compares to other models, you might want to go into further detail when interpreting the findings. Give an explanation of why XAI-CROP performed better than the other models and an explanation of the factors that led to its higher R2 values, lower MSE, and lower MAE. This can make it easier for the reader to comprehend your model's advantages.

Useful use cases: Describe particular situations in which the XAI-CROP model can be put to use. You could,

for example, explain how a farmer's decision-making process could incorporate the model. This could entail a comprehensive explanation of how the model's suggestions are implemented in the real world or a step-by-step manual.

6 Several real-world implications can be done in the real world such as:

1. Precision Agriculture: The proposed model can be utilized to provide personalized crop recommendations to farmers based on factors such as soil quality, weather conditions, historical data, and specific crop requirements. This can help optimize resource allocation, improve crop yield, and minimize environmental impact by reducing the use of fertilizers and pesticides.

2. Sustainable Farming Practices: By incorporating explainable AI into crop recommendation systems, the proposed model can assist farmers in adopting sustainable farming practices. It can provide insights into the ecological impact of different crop choices and recommend environmentally friendly strategies, such as crop rotation or intercropping, to enhance soil health and biodiversity preservation.

3. Climate Change Adaptation: With climate change affecting agricultural productivity and patterns, the proposed model can aid farmers in adapting to changing conditions. By analyzing historical climate data and incorporating predictive models, it can generate recommendations for resilient crop choices that are better suited to withstand extreme weather events or shifting climate patterns.

4. Small-Scale Farming Support: Small-scale farmers often face unique challenges in terms of limited resources and access to information. The proposed model can offer tailored crop recommendations and provide valuable insights to support decision-making for small-scale farmers, helping them maximize their productivity and profitability.

5. Decision Support for Agricultural Advisors: Agricultural advisors and consultants can utilize the proposed model to provide expert recommendations to farmers. By incorporating explainable AI, the model can transparently present the underlying reasoning and justifications for specific crop recommendations, enabling advisors to effectively communicate and gain trust from farmers.

7 Conclusion

In this study, we developed XAI-CROP, a crop recommendation system that leverages machine learning and eXplainable artificial intelligence techniques. Through extensive training and evaluation on a dataset encompassing crop cultivation in India, we demonstrated the superior performance of XAI-CROP compared to other machine learning models. Our results showcased its higher accuracy, as indicated by lower Mean Squared Error (MSE), Mean Absolute Error (MAE), and higher R-squared (R2) value.

The key innovation of XAI-CROP lies in its ability to provide interpretable explanations for its recommendations. By incorporating eXplainable artificial intelligence (XAI) techniques such as LIME, XAI-CROP demystifies the decision-making process, making it transparent and understandable for farmers and agricultural stakeholders. This transparency not only enhances the system's usability but also builds trust and confidence among end-users by enabling them to comprehend the rationale behind the recommended crop choices.

Moreover, XAI-CROP holds profound implications for real-world applications, particularly in the context of precision agriculture and data-driven farming. By factoring in geographical location, seasonal variations, and historical data, XAI-CROP empowers users to make informed decisions regarding crop selection. This capability contributes to improved crop yield predictions, optimized resource allocation, and ultimately, enhanced food security in agricultural regions. The physical relevance of XAI-CROP is further highlighted by its compatibility with different geographical locations, scalability to various crops, and potential integration with state-of-the-art techniques. By considering these aspects, our model demonstrates its potential to address diverse agricultural challenges and play a pivotal role in sustainable and efficient crop cultivation worldwide.

In conclusion, XAI-CROP has the potential to revolutionize the agricultural sector by enabling farmers to make data-driven decisions, leading to increased crop yields and improved food security. Our study underscores the importance of utilizing machine learning and explainable AI techniques in the development of practical and effective crop recommendation systems. Future research can focus on enhancing the accuracy and scalability of XAI-CROP, extending its application to other regions and crops, and exploring integration with state-of-the-art techniques to further advance agricultural decision support. By bridging the gap between advanced technology and the physical world of agriculture, XAI-CROP represents a significant step forward in enabling farmers to navigate the complexities of modern farming and make informed choices for sustainable and efficient crop cultivation. It is our hope that this research contributes to the advancement of agricultural practices worldwide, promoting food security and environmental sustainability.

In the future, the proposed algorithm can be used with OCNN [55–58] and make use of Resnet [59]. Attention mechanism can be used as in [60] and correlation algorithms as in [61]. YOLO v8 can be used as in [62]. Additionally, future work can explore the integration of XAI-CROP with state-of-the-art techniques such as those demonstrated in references [63–71], offering even more sophisticated capabilities for financial decision support.

Author contributions It is a collaborative effort where Mahmoud, Fatma and Samah worked together. Fatma came up with the idea and wrote the abstract and the proposal, while Samah contributed by making comparisons and made the experiments. Mahmoud wrote the introduction and the related work.

Data availability https://www.kaggle.com/datasets/ananysharma/crop-yield.

Declarations

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

Ethical approval There is no any ethical conflicts.

References

1. Bhadouria R, et al. (2019) Agriculture in the era of climate change: Consequences and effects. In Climate Change and Agricultural Ecosystems, Elsevier, 1–23.
2. Xu X et al (2019) Design of an integrated climatic assessment indicator (ICAI) for wheat production: a case study in Jiangsu Province, China. Ecol Ind 101:943–953

3. Bali N, Singla A (2021) Deep learning based wheat crop yield prediction model in punjab region of north india. Appl Artif Intell 35(15):1304–1328

4. Van Klompenburg T, Kassahun A, Catal C (2020) Crop yield prediction using machine learning: A systematic literature review. Comput Electron Agric 177:105709

5. Alpaydin E (2020) Introduction to machine learning. MIT press.

6. Tarek Z et al (2023) Soil erosion status prediction using a novel random forest model optimized by random search method. Sustainability 15(9):9. https://doi.org/10.3390/su15097114

7. Shams MY, Tarek Z, Elshewey AM, Hany M, Darwish A, Hassanien AE (2023) A machine learning-based model for predicting temperature under the effects of climate change. In: The Power of Data: Driving Climate Change with Data Science and Artificial Intelligence Innovations, A. E. Hassanien and A. Darwish, Eds., in Studies in Big Data. Cham: Springer Nature Switzerland, 2023: 61–81. https://doi.org/10.1007/978-3-031-22456-0_4.

8. Elshewey AM et al (2023) A novel WD-SARIMAX model for temperature forecasting using daily Delhi climate dataset. Sustainability 15(1):1. https://doi.org/10.3390/su15010757

9. Porter JR, Xie L, Challinor AJ, Cochrane K, Howden SM, Iqbal MM, Lobell DB, Travasso MI (2014) Food security and food production systems. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 1(1): 485–533 (2014).

10. Rosenzweig C et al (2013) The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. Agric For Meteorol 170:166–182

11. Khater HA, Gamel SA (2023) Early diagnosis of respiratory system diseases (RSD) using deep convolutional neural networks. J Ambient Intell Human Comput 14:12273–12283

12. Lobell DB, Asseng S (2017) Comparing estimates of climate change impacts from process-based and statistical crop models. Environ Res Lett 12(1):015001

13. Schlenker W, Roberts MJ (2009) Nonlinear temperature effects indicate severe damages to US crop yields under climate change. Proc Natl Acad Sci 106(37):15594–15598

14. Roberts MJ, Braun NO, Sinclair TR, Lobell DB, Schlenker W (2017) Comparing and combining process-based crop models and statistical models with some implications for climate change. Environ Res Lett 12(9):095010

15. Roberts MJ, Schlenker W, Eyer J (2013) Agronomic weather measures in econometric models of crop yield with implications for climate change. Am J Agr Econ 95(2):236–243

16. Patel K, Patel HB (2023) Multi-criteria agriculture recommendation system using machine learning for crop and fertilizesrs prediction. Curr Agricult Res J 11(1), 2023.

17. Mittal N, Bhanja A (2023) Implementation and identification of crop based on soil texture using AI. In: 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE. 1467–1471.

18. Fenz S, Neubauer T, Heurix J, Friedel JK, Wohlmuth M-L (2023) AI- and data-driven pre-crop values and crop rotation matrices. Eur J Agron 150:126949. https://doi.org/10.1016/j.eja.2023.126949

19. Arif MS, Mukheimer A, Asif D (2023) Enhancing the early detection of chronic kidney disease: a robust machine learning model. Big Data Cognit Comput 7(3):3. https://doi.org/10.3390/bdcc7030144

20. Asif D, Bibi M, Arif MS, Mukheimer A (2023) Enhancing heart disease prediction through ensemble learning techniques with hyperparameter optimization. Algorithms 16(6):6. https://doi.org/10.3390/a16060308

21. Nawaz Y, Arif MS, Shatanawi W, Nazeer A (2021) An explicit fourth-order compact numerical scheme for heat transfer of boundary layer flow. Energies 14(12):12. https://doi.org/10.3390/en14123396

22. Nawaz Y, Arif MS, Abodayeh K (2022) A third-order two-stage numerical scheme for fractional stokes problems: a comparative computational study. J Comput Nonlinear Dyn 17:101004. https://doi.org/10.1115/1.4054800

23. Nawaz Y, Arif MS, Abodayeh K (2022) An explicit-implicit numerical scheme for time fractional boundary layer flows. Int J Numer Meth Fluids 94(7):920–940. https://doi.org/10.1002/fld.5078

24. McEldowney JF (2021) Climate change and the law. In: the impacts of climate change, Elsevier. 503–519.

25. de Oliveira AC, Marini N, Farias DR (2014) Climate change: New breeding pressures and goals. Encyclopedia Agricult Food Syst 2014:284–293

26. Williams TO, et al. (2015) Climate smart agriculture in the African context. Unlocking Africa's Agricultural Potentials for Transformation to Scale , FAO and UNEP , Abdou Diouf International Conference, Dakar, Senegal, pp. 1–26, 2015.

27. Reddy PS, Amarnath B, Sankari M (2023) Study on machine learning and back propagation for crop recommendation system. In: 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), IEEE. 1533–1537.

28. You J, Li X, Low M, Lobell D, Ermon S (2017) Deep gaussian process for crop yield prediction based on remote sensing data. In: Thirty-First AAAI conference on artificial intelligence.

29. Paudel D et al (2021) Machine learning for large-scale crop yield forecasting. Agric Syst 187:103016

30. Sun J, Lai Z, Di L, Sun Z, Tao J, Shen Y (2020) Multilevel deep learning network for county-level corn yield estimation in the us corn belt. IEEE J Selected Top Appl Earth Obs

31. Yoon HS et al (2021) Akkermansia muciniphila secretes a glucagon-like peptide-1-inducing protein that improves glucose homeostasis and ameliorates metabolic disease in mice. Nat Microbiol 6(5):5. https://doi.org/10.1038/s41564-021-00880-5

32. Khaki S, Wang L (2022) Crop Yield Prediction Using Deep Neural Networks. Front Plant Sci 10, 2019, Accessed: Sep. 27, 2022. Available: https://doi.org/10.3389/fpls.2019.00621

33. Abbas F, Afzaal H, Farooque AA, Tang S (2020) Crop yield prediction through proximal sensing and machine learning algorithms. Agronomy 10(7):7. https://doi.org/10.3390/agronomy10071046

34. Talaat FM (2023) Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. Neural Comput Applic. https://doi.org/10.1007/s00521-023-08619-5

35. Friedman JH (2001) Greedy function approximation: a gradient boosting machine. Ann Stat 29(5):1189–1232

36. Natekin A, Knoll A (2022) Gradient boosting machines, a tutorial. Front Neurorobotics 7, 2013, Accessed: Sep. 27, 2022. Available: https://doi.org/10.3389/fnbot.2013.00021

37. Ke G, et al. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Advances in Neural Information Processing Systems, 2017, 30. Accessed: Sep. 27, 2022. Available: https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html

38. Rao H et al (2019) Feature selection based on artificial bee colony and gradient boosting decision tree. Appl Soft Comput 74:634–642. https://doi.org/10.1016/j.asoc.2018.10.036

39. Freund Y, Mason L (1999) The alternating decision tree learning algorithm. In: Icml, 1999, 99, pp. 124–133.

40. Feng J, Yu Y, Zhou Z-H (2018) Multi-Layered Gradient Boosting Decision Trees. In: Advances in Neural Information Processing Systems, 2018, 31. Accessed: Sep. 27, 2022. Available: https://proceedings.neurips.cc/paper/2018/hash/39027dfad5138c9ca0c474d71db915c3-Abstract.html

41. Pretorius A, Bierman S, Steel SJ (2016) A meta-analysis of research in random forests for classification. In: 2016 Pattern Recognition Association of South Africa and Robotics and Mechatronics International Conference (PRASA-RobMech), 2016, pp. 1–6.

42. Sun C, Li X, Guo R (2021) Research on electrical fire risk assessment technology of cultural building based on random forest algorithm. In: 2021 International Conference on Aviation Safety and Information Technology, New York, NY, USA, Dec. 2021, pp. 769–773. https://doi.org/10.1145/3510858.3511382.

43. Geenen PL, van der Gaag LC, Loeffen WLA, Elbers ARW (2011) Constructing naive Bayesian classifiers for veterinary medicine: A case study in the clinical diagnosis of classical swine fever. Res Vet Sci 91(1):64–70. https://doi.org/10.1016/j.rvsc.2010.08.006

44. Xu S (2018) Bayesian Naïve Bayes classifiers to text classification. J Inf Sci 44(1):48–59. https://doi.org/10.1177/0165551516677946

45. Kibriya AM, Frank E, Pfahringer B, Holmes G (2005) Multinomial naive bayes for text categorization revisited. In: AI 2004: Advances in Artificial Intelligence, Berlin, Heidelberg, 2005, pp. 488–499. https://doi.org/10.1007/978-3-540-30549-1_43.

46. Jiang L, Wang S, Li C, Zhang L (2016) Structure extended multinomial naive Bayes. Inf Sci 329:346–356. https://doi.org/10.1016/j.ins.2015.09.037

47. https://www.kaggle.com/datasets/ananysharma/crop-yield.

48. Elshewey A, Shams M, Tarek Z, Megahed M, El-kenawy E-S, El-dosuky M (2023) Weight prediction using the hybrid stacked-LSTM food selection model. CSSE, 46(1): 765–781, 2023, https://doi.org/10.32604/csse.2023.034324.

49. Shams MY, Elshewey AM, El-kenawy E-SM, Ibrahim A, Talaat FM, Tarek Z (2023) Water quality prediction using machine learning models based on grid search method. Multimed Tools Appl. https://doi.org/10.1007/s11042-023-16737-4

50. Ryo M (2022) Explainable artificial intelligence and interpretable machine learning for agricultural data analysis. Artif Intell Agricult 6:257–265. https://doi.org/10.1016/j.aiia.2022.11.003

51. Doshi Z, Nadkarni S, Agrawal R, Shah N (2018) AgroConsultant: Intelligent crop recommendation system using machine learning algorithms. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Aug. 2018, pp. 1–6. https://doi.org/10.1109/ICCUBEA.2018.8697349.

52. Taleizadeh AA, Amjadian A, Hashemi-Petroodi SE, Moon I (2023) Supply chain coordination based on mean-variance risk optimisation: pricing, warranty, and full-refund decisions. Int J Syst Sci: Oper Logist 10(1):2249808. https://doi.org/10.1080/23302674.2023.2249808

53. Gharaei A, Amjadian A, Shavandi A, Amjadian A (2023) An augmented Lagrangian approach with general constraints to solve nonlinear models of the large-scale reliable inventory systems. J Comb Optim 45(2):78. https://doi.org/10.1007/s10878-023-01002-z

54. Taleizadeh AA, Varzi AM, Amjadian A, Noori-daryan M, Konstantaras I (2023) How cash-back strategy affect sale rate under refund and customers' credit. Oper Res Int J 23(1):19. https://doi.org/10.1007/s12351-023-00752-2

55. Talaat FM (2022) Effective deep Q-networks (EDQN) strategy for resource allocation based on optimized reinforcement learning algorithm. Multimed Tools Appl 81(17). https://doi.org/10.1007/s11042-022-13000-0

56. Talaat FM (2022) Effective prediction and resource allocation method (EPRAM) in fog computing environment for smart healthcare system. Multimed Tools Appl

57. Talaat Fatma M, Alshathri Samah, Nasr Aida A (2022) A new reliable system for managing virtualcloud network. Comput Mater Continua 73(3):5863–5885. https://doi.org/10.32604/cmc.2022.026547

58. El-Rashidy N, ElSayed NE, El-Ghamry A, Talaat FM (2022) Prediction of gestational diabetes based on explainable deep learning and fog computing. Soft Comput 26(21):11435–11450

59. El-Rashidy N, Ebrahim N, el Ghamry A, Talaat FM (2022) Utilizing fog computing and explainable deep learning techniques for gestational diabetes prediction. Neural Comput Applic. https://doi.org/10.1007/s00521-022-08007-59.FaivdullahL,AzaharF,HtikeZZ,Naing

60. Hanaa S, Fatma BT (2022) Detection and classification using deep learning and sine-cosine fitnessgrey wolf optimization. Bioengineering 10(1):18. https://doi.org/10.3390/bioengineering10010018

61. Talaat FM (2023) Real-time facial emotion recognition system among children with autism based on deep learning and IoT. Neural Comput Appl 35(3), https://doi.org/10.1007/s00521-023-08372-9

62. Talaat FM (2023) Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes, April 2023, Neural Comput Appl 35(2), https://doi.org/10.1007/s00521-023-08619-5

63. Hassan E, El-Rashidy N, Talaat FM (2022) Review: Mask R-CNN Models. May 2022, https://doi.org/10.21608/njccs.2022.280047.

64. Siam AI, Gamel SA, Talaat FM (2023) Automatic stress detection in car drivers based on non-invasive physiological signals using machine learning techniques. Neural Comput Applic. https://doi.org/10.1007/s00521-023-08428-w

65. Talaat FM, Gamel SA (2023) A2M-LEUK: attention-augmented algorithm for blood cancer detection in children, June 2023, Neural Comput Appl. https://doi.org/10.1007/s00521-023-08678-8

66. Gamel SA, Hassan E, El-Rashidy N et al (2023) Exploring the effects of pandemics on transportation through correlations and deep learning techniques. Multimed Tools Appl. https://doi.org/10.1007/s11042-023-15803-1

67. Talaat FM, ZainEldin H (2023) An improved fire detection approach based on YOLO-v8 for smart cities. Neural Comput Applic. https://doi.org/10.1007/s00521-023-08809-1

68. Alnaggar M, Siam AI, Handosa M, Medhat T, Rashad MZ (2023) Video-based real-time monitoring for heart rate and respiration rate. Expert Syst Appl 1(225):120135

69. Alnaggar M, Handosa M, Medhat T, Z Rashad M (2023) Thyroid Disease multi-class classification based on optimized gradient boosting model. Egypt J Artif Intell. 2(1):1–4.

70. Alnaggar M, Handosa M, Medhat T, Rashad MZ (2023) An IoT-based framework for detecting heart conditions using machine learning. Int J Adv Comput Sci Appl. 14(4).

71. Alhussan AA, Talaat FM, El-kenawy ES, Abdelhamid AA, Ibrahim A, Khafaga DS, Alnaggar M (2023) Facial expression recognition model depending on optimized support vector machine. Comput Mater Continua. 76(1).