IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 7 September 2024, accepted 29 October 2024, date of publication 6 November 2024, date of current version 22 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3492973

RESEARCH ARTICLE

XAI-Powered Smart Agriculture Framework for Enhancing Food Productivity and Sustainability

R. JOHN MARTIN[1], (Member, IEEE), RUCHI MITTAL[2], VARUN MALIK[2], (Member, IEEE), FATHE JERIBI[1], SHAMS TABREZ SIDDIQUI[1], MOHAMMAD ALAMGIR HOSSAIN[1], AND S. L. SWAPNA[3]

[1]College of Engineering and Computer Science, Jazan University, Jazan 45142, Saudi Arabia
[2]Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura 140401, India
[3]Applexus Technologies, TechnoPark, Thiruvananthapuram 695583, India

Corresponding author: Fathe Jeribi (fjeribi@jazanu.edu.sa)

This work was supported by the Deanship of Graduate Studies and Scientific Research, Jazan University, Saudi Arabia, under Project RG24-S0140.

ABSTRACT A vital component of maintaining the world's expanding populace is farming. In agriculture field, factors such as soil quality, weather patterns, and crop yields are essential components of usual possessions that affect farming manufacture. Despite advancements, prevailing smart systems quiet struggle with handling big amounts in prediction claims, often facing difficulties in balancing prediction accuracy and learning efficiency. To ensure sustainable food production, integrating advanced machineries such as machine learning and artificial intelligence in agriculture is essential. This study proposes an explainable AI (XAI)-based smart agriculture system to provide holistic recommendation for precision farming aimed at improving productivity while reducing environmental impact. We compiled a comprehensive weather, soil, and crop dataset from official and verified sources in India. From this dataset, we extracted and optimized features using pre-trained architectures and enhanced barnacles mating optimization (EBMO) algorithm, addressing the high-measure mentality and computational complexity issues often encountered in agricultural data analysis. The selected features were analyzed to provide holistic recommendations for precision farming using baseline ML models such as support vector machine, random forest, neural network, and decision tree. Additionally, we integrate the XAI framework with an interpretable recommendation/s for optimizing the agricultural practices. The study developed an XAI-based smart agriculture system that provides holistic recommendations for precision farming to boost productivity. By using a comprehensive dataset and optimizing features with the EBMO algorithm, the research achieved high accuracy in crop yield predictions, particularly with the XLNet+SVM model, which outperformed existing models across various crops. The integration of SHapley Additive explanation (SHAP) and Local Interpretable Model-agnostic Explanation (LIME) further enabled interpretable AI-driven insights, enhancing transparency in decision-making. The results demonstrated significant improvements in prediction accuracy, resource management, and sustainability, offering valuable contributions to global food security through smart agriculture. The ability to explain these AI decisions further supports the adoption of AI technologies in agriculture, fostering resilient agricultural systems that are essential for feeding the world's growing population.

INDEX TERMS Smart agriculture, sustainable food production, precision farming, explainable AI, machine learning.

I. INTRODUCTION

Agriculture encompasses the cultivation of land, the raising of animals, and the production of diet, grit, and added essential crops for social use. It is vital for manufacturing, particularly in terms of the food and raw supplies necessary

The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li.

to support human life and economic expansion [1]. However, traditional agricultural practices face numerous challenges, including climate change, resource limitations, and the need for increased efficiency [2], [3], [4], [5], [6], [7]. By utilizing gears such as devices, hums, ML (Machine Learning) and AI (Artificial Intelligence) processes, smart agriculture enables the real-time monitoring and management of crops, soil, and weather conditions [8]. Smart agriculture aims to improve farming processes and boost efficiency while minimizing conservational influence. It transforms traditional farming by enabling efficient resource use, saving money, and reducing waste through IoT sensors and data analytics [9]. Customary predicting approaches rely on time series data analysis using possibility and amounts, but they are partial by past information and often fail to accurately predict and generalize [10]. In contrast, data-driven machine learning (ML) algorithms have rapidly developed, offering significant potential in agriculture [11]. However, challenges such as insufficient data groundwork, high sensor costs, and the need for specialized expertise hinder ML deployment in agriculture. Deep neural networks (DNNs) are increasingly used for forecasting applications [12]. Standard DNN models like Transformer models, encoder-decoder models, gated recurrent units (GRU), long short-term memory (LSTM), and convolutional neural networks (CNNs) are widely applied in fields such as natural language processing, time series prediction, computer vision, and image classification [13]. Compared to traditional methods, deep learning (DL) offers effective modeling capabilities with large datasets on modern computer system. However, neural networks can learn pseudo-relationships due to a lack of physical constraints. The decision-making process in traditional ML/DL models can be compound and stimulating, even for specialists [14]. Explainable AI (XAI) techniques [15] discourse this issue by in case clear and interpretable clarifications of AI-related choices [16], [17]. XAI assists farmers in making better decisions and maximizing crop output by elucidating the identification and resolution of these problems [15]. By offering clear and intelligible AI-based advice that farmers can rely on and comprehend, XAI dramatically increases the uptake and efficacy of innovative agricultural technologies [18]. Based on these references, we introduce an XAI-based recommendation framework for a smart agriculture system, designed to enhance precision farming through optimized irrigation scheduling and strategic crop rotation. The major contributions of this proposed framework are detailed as follows:

1. Data synthesis: We meticulously assembled an extensive dataset encompassing weather, soil, and crop information from authoritative and verified Indian sources, including, geoportal.natmo.gov.in (soil data), data.icrisat.org (crop production) and power.larc.nasa.gov (meteorological data).
2. Advanced feature extraction: Utilizing XLNet, we extracted profound features from the curated dataset, capturing intricate patterns and insights essential for accurate predictions.
3. Enhanced optimization: An enhanced barnacles mating optimization (EBMO) algorithm is used to tackle the challenges of high-measure mentality and computational complexity inherent in agricultural data analysis.
4. Holistic analysis: The selected optimal features were analyzed to deliver comprehensive recommendations for precision farming, leveraging established machine learning models such as Support vector machines, random forests, decision trees, and neural networks
5. Interpretable insights: By integrating XAI frameworks SHapley Additive explanation (SHAP) and Local Interpretable Model-agnostic Explanation (LIME), we offer transparent and actionable insights ensure recommendations for optimizing agricultural practices.

This paper is structured as follows: Section II summarizes related work on recommendation framework for a smart agriculture system. Section III expounds on the conceptual architecture of the proposed framework and describes the process details. Section IV presents the experimental results and analysis. Finally, the conclusions of this study are summarized in Section V.

II. RELATED WORKS

This section provides a literature review on smart agriculture systems aimed at sustainable food production. Table 1 summarizes the research gaps identified from the existing state-of-the-art studies.

A. STATE OF ART WORKS

By combining the Internet of Things (IoT) with a machine learning (ML) model, the three-layer architecture known as FARMIT was able to continuously evaluate the quality of the crop using data collected from various sources [19]. The crop quality was evaluated using a random forest model, which yielded results that were highly comparable to those found by a trained model. The percentage error it attained was 6.59%, which is lower than the typical solution's 6.71%. A potential solution to the problems of food insecurity, climate change adaptation, and mitigation is climate-smart agriculture (CSA) [20]. For smart agriculture, a novel hybrid model is presented, utilizing the Internet of Things (IoT) and artificial intelligence (AI) algorithms to develop a reliable and cost-effective decision-making system [21]. A remarkable accuracy of 0.91 is displayed by the hybrid model that incorporates both the LSTM layer and inception V3. In order to ensure the precision and dependability of the data acquired, they employed ML algorithm-based techniques such as k-NN, gradient boosting, XGBoost, and multilayer perceptron for crop yield projection [22]. A smart irrigation system that runs on the cloud was suggested as a way to link several modest smart farms and consolidate relevant data [23]. The data can be used to make informed decisions about water management, which helps with conservation efforts, especially in dry areas [24]. The findings provide important data for studies aiming to improve agricultural quality and

TABLE 1. Research gap summary from existing works on smart agriculture system.

Ref.	Smart agriculture	ML/DL	Constraints	Findings	Research gap
[19]	IoT-based smart farming	Random Forest (RF)	Temperature, ambient humidity, soil humidity	Error 6.71%	Weak forecasting models and not provide clear explanations
[20]	Climate-smart agriculture	Linear regression	Rainfall distribution	MAE 12.526	Leading to discriminatory outcomes due to data imbalance
[21]	IoT-based monitoring for smart	MLP, Naïve Bayes, and SVM	Soil color, moisture content, composition	Accuracy 91%	The decision-making processes of is extremely complex, making it hard to validate.
[22]	Predicting potato crop yield	LSTMs and GRUs	Spatial and temporal patterns	MSE 0.03177 and 0.03150	Concentrate only on crop yield prediction and not suitable for smart agriculture
[23]	Water management in smart agriculture	Compact RIO controller	Groundwater, exploitable, mobilized potential, over-exploited volume	MAE 0.058	Non-justified threshold formation makes no sense in the outcomes
[24]	Crop yield prediction	LSTM and RNN	Soil color, moisture content, composition	Accuracy 85%	Difficult to handle real-time data because of model complexity
[25]	Smart irrigation system	XGBoost	Moisture, humidity, temperature, water levels	70% reduction of water consumption	Not provide clear explanations for their water management outputs
[26]	Safeguard crops from wildlife threats	AI-CAM and EvoNet	Crop damage	Accuracy96.7%	Not suitable for resource-constrained agricultural environments
[27]	Image analysis for cotton crop classification	AlexNet, GoogLeNet, InceptionV3 & VGG-19	Area, production, yield	Accuracy 93.4%	Limited understanding of the factors that influence farmers' decisions
[28]	Camera informed agricultural land monitoring	Near Surface Camera (NSCam)	Long-term or abrupt agricultural land changes	F-measure 84.56%	Data sources are limited by weather conditions, leads to gaps in agricultural monitoring

profitability through the use of artificial intelligence and the internet of things (IoT). Through the use of cutting-edge agricultural technology, an intelligent irrigation system that incorporates embedded systems, the internet of things (IoT), and cloud computing can increase food security [25]. One novel approach uses the Internet of Things (IoT) and edge AI to create a system that can identify and prevent animal intrusions using small DL algorithms based on machine learning [26]. DL with feature extraction methods such as continuous wavelet transform (CWT) and fast Fourier transform (FFT). The detection process involved employing pre-trained models such as AlexNet, GoogLeNet, InceptionV3, and VGG-19. The model reaches the maximum accuracy of 96.7%, according to the results [27]. When educated on features obtained through CWT instead of FFT, the GoogLeNet learning model achieved an impressive precision of 93.4% and an F1-score of 0.953 [28].

B. PROBLEM DESCRIPTION

Agriculture is vital for sustaining the growing global population, faces numerous challenges due to climate change, soil degradation, and need for increased productivity. Despite technological advancements, traditional agricultural practices and existing intelligent models encounter several significant problems. Many studies focus separately on climate forecasting, soil prediction, or crop yield prediction, resulting in fragmented insights. Few studies consider the interdependence of soil quality, weather patterns, and crop yields together, leading to incomplete decision-making frameworks. Agricultural data is vast and multi-measure mental, presenting challenges in processing and analysis. Traditional models struggle with the computational demands of large datasets, leading to inefficiencies and inaccuracies [29]. Many intelligent models do not achieve the desired accuracy levels due to the complexity and variability of agricultural data. Models that require extensive training data and computational power are impractical for many agricultural applications. Traditional ML/DL models [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29] are often "black boxes," providing little insight into how decisions are made. To address these issues, we proposed an XAI-based smart agriculture system that integrates climate, soil, and crop yield data to provide interpretable recommendations for sustainable food production. By combining weather, soil, and crop data from verified sources, our model provides a complete picture of the factors affecting agricultural productivity. We compiled a comprehensive dataset from official sources in India, including data.icrisat.org for crop production, power.larc.nasa.gov for meteorological data, and geoportal.natmo.gov.in for soil data. Pre-trained architectures are used to extract meaningful features from the dataset. The enhanced barnacles mating optimization (EBMO) algorithm addresses high-measure mentality and computational complexity, optimizing features for better model performance. We use support vector machines, decision trees, random forests, and neural networks to analyze the optimized features and provide actionable insights. The integration of XAI techniques ensures that the decision-making process is transparent and interpretable, allow farmers to understand the rationale behind recommendations.

III. METHODS

Fig. 1 depicts a framework for holistic recommendations in precision farming using an XAI-based smart agriculture system. The process starts with gathering extensive datasets on weather, soil, and crop yields from verified Indian sources. This data undergoes preprocessing to ensure quality and reliability. Data preprocessing in the proposed XAI-based smart agriculture system is a crucial step to ensure that the collected datasets on weather, soil, and crop yields are of high quality, reliable, and suitable for feature extraction and subsequent analysis. The process begins with data cleaning, which involves handling missing values, correcting errors, and removing irrelevant or redundant data points. For example, any incomplete weather or soil records are either imputed using statistical methods or excluded if deemed unreliable. Next, the data undergoes normalization and scaling to standardize the values across different features, which is essential when working with machine learning models. This step ensures that no single variable disproportionately influences the model's performance, especially given the diverse nature of the data sources (e.g., temperature, humidity, soil pH levels, and crop yield metrics). Techniques such as Min-Max scaling or Z-score normalization are employed, depending on the distribution of the data.

Data transformation is another key aspect of preprocessing, where categorical variables (like soil types or weather conditions) are encoded into numerical values using one-hot encoding or label encoding. This step enables the machine learning models to process and understand these variables effectively. Additionally, outlier detection is performed to identify and manage extreme values that could skew the results. Methods like the interquartile range (IQR) or Z-score analysis help in pinpointing these anomalies. Once cleaned and transformed, the dataset is subjected to feature engineering to create new, more informative variables that capture essential relationships among the data points. For example, combining temperature and humidity data might generate a new feature indicating the likelihood of drought or optimal growth conditions for specific crops. Lastly, the dataset is split into training and testing subsets to validate the models effectively, ensuring that the system's performance is generalizable and robust. This comprehensive preprocessing pipeline is essential for preparing high-dimensional and complex agricultural data, thereby laying a solid foundation for accurate and efficient model training and predictions in the smart agriculture framework. Features are extracted with XLNet, and data dimensionality is reduced using the EBMO algorithm to streamline the data. These predictions are then interpreted using XAI frameworks like LIME and SHAP, providing farmers with clear, actionable recommendations to enhance productivity and sustainability.

A. FEATURE EXTRACTION

Feature extraction is a critical step in the data processing pipeline that ensures machine learning models are provided with meaningful and significant attributes, which greatly influence the predictions and outcomes. In the realm of smart agriculture, feature extraction transforms raw and often unstructured agricultural data into a set of relevant features that can be efficiently utilized by machine learning algorithms for precision farming recommendations. The process of feature extraction begins with an in-depth analysis of the raw data to identify the elements that hold the most predictive power. For example, weather data is analyzed to extract features such as:

- Temperature: Daily maximum and minimum values, weekly and monthly averages, and temperature variation trends, which are crucial for understanding crop growth and health.
- Precipitation: Rainfall amounts and distribution patterns, which impact soil moisture and water availability for crops.
- Humidity levels: These affect evapotranspiration rates and plant stress conditions.
- Wind speed: Used to assess the risk of crop damage and potential evaporation losses.
- Solar radiation: Captured in terms of daily light intensity and duration, affecting photosynthesis and plant growth.
- Historical weather patterns: Trends and anomalies that could influence future crop cycles or predict extreme weather events.

From soil data, important features are extracted to provide insights into soil health and fertility, such as:

- Soil moisture content: Indicates the water-holding capacity and irrigation needs of the soil.
- pH levels: Determines soil acidity or alkalinity, which affects nutrient availability and crop suitability.
- Nutrient concentrations: Levels of essential nutrients like nitrogen, phosphorus, and potassium, which are critical for plant growth.
- Soil texture: Proportions of sand, silt, and clay, which influence water retention, aeration, and nutrient supply.
- Organic matter content: Reflects the soil's fertility and its ability to retain moisture and nutrients.
- Soil temperature: Impacts seed germination, root development, and microbial activity in the soil.

Crop yield data provides additional features that help predict and optimize agricultural outcomes, including:

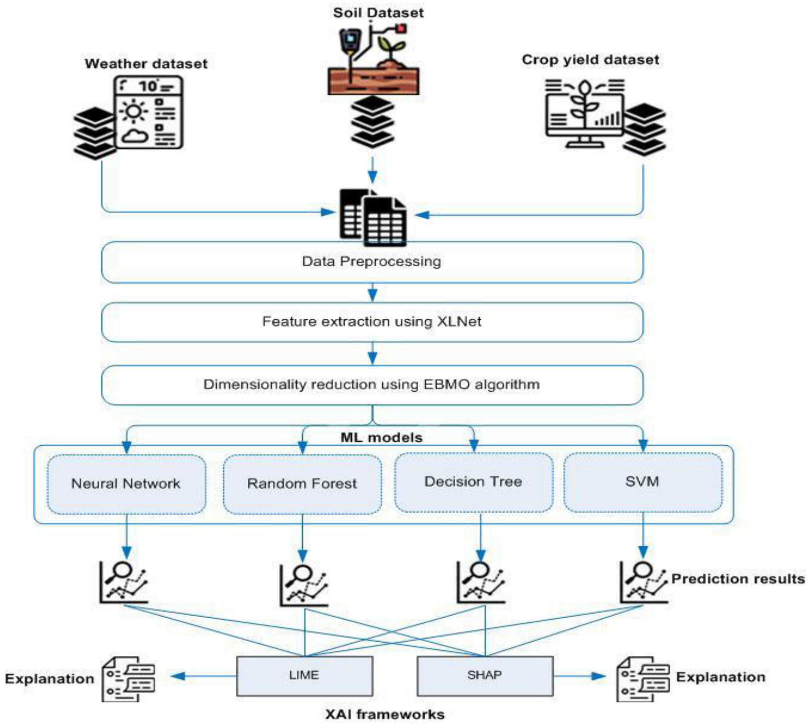

FIGURE 1. Conceptual framework of holistic recommendation for precision farming using XAI based smart agriculture.

- Crop type: Classifying different crops allows the model to tailor recommendations to specific needs.
- Growth stages: Tracking the development phase of the crop (e.g., germination, flowering, harvest) to optimize interventions like fertilization and irrigation.
- Yield per hectare: Historical yield data helps forecast future productivity and assess the impact of various farming practices.
- Plant health indices: Metrics like the Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), derived from remote sensing data, provide insights into crop vigor and early detection of stress or diseases.
- Pest and disease incidence: Tracking occurrences of pests and diseases to preemptively address crop threats.
- Harvest times: Analyzing the optimal timing for harvesting to maximize yield and quality.

In this study, XLNet [30], [31] is utilized for feature extraction due to its effectiveness in capturing intricate patterns and relationships within large, diverse datasets. XLNet, an autoregressive pre-training model, is particularly powerful in modeling sequential dependencies, making it well-suited for the complex temporal and spatial relationships present in agricultural data. XLNet uses a generalized permutation-based training objective that captures bidirectional context, allowing it to understand the full sequence of features, which is crucial for interpreting how weather patterns influence crop growth over time. The model is adept at recognizing subtle interactions between variables, such as how humidity, soil pH, and nutrient levels collectively affect crop health. XLNet can process different types of data, whether sequential (e.g., time-series weather data) or categorical (e.g., soil type classifications), and effectively combine them into a comprehensive feature set. By learning from permutations of input sequences, XLNet is better equipped to uncover non-linear relationships that simpler models might overlook, such as the impact of sporadic rainfall on crop yield over multiple growing seasons.

B. ENHANCED OPTIMIZATION

Enhanced optimization handling the vast and complex datasets encountered in agricultural data analysis. The enhanced barnacles mating optimization (EBMO) algorithm is used to address the challenges of high-measurement and computational complexity inherent in this field. It is an advanced variant of the barnacles mating optimization (BMO) algorithm [32], which is inspired by the natural mating behavior of barnacles. In the feature selection, EBMO works by evaluating a vast amount of potential feature subsets to determine the most informative and relevant features for predictive modeling. In the case of agricultural data, this involves assessing how well the selected features contribute to accurate predictions of crop yields, soil conditions, or weather patterns. Inspired by barnacle mating behaviors, the algorithm utilizes pair's solution to produce offspring. The mating strategy ensures that the algorithm explores a diverse set of feature combinations.

In the breeding cluster, all individuals are currently inactive in mating. The mating behavior of crabs served as the inspiration for developing the barnacle mating optimization algorithm. First, the candidate solution is considered as a collection that can represent the population using a matrix as follows.

$$P = \begin{bmatrix} p_1^1 & \cdots & p_1^b \\ \vdots & \ddots & \vdots \\ p_B^1 & \cdots & p_B^b \end{bmatrix} \tag{1}$$

Population estimation and cataloging procedure is achieved to find the best explanation at the top of P. After that they choose to marry their parents

$$Barnacale\_c = Randperm(B) \tag{2}$$

$$Barnacale\_m = Randperm(B) \tag{3}$$

where B is the amount of barnacle populaces, b is the amount of switch factors, and Barnacle_c and Barnacle_m signify the mating parentages. Barnacle creates spring variables from the parent.

$$p_h^{B\_New} = xp^B_{Barnacle\_c} + yp^B_{Barnacle\_a} \tag{4}$$

where x is a casual amount normally dispersed amid [0, 1] and y = (1 - p) $xp^B_{Barnacle\_c}$ and $xp^B_{Barnacle\_m}$ represents the chosen father and mother barnacle variables. x and y signify the genetic occurrences of the father and mother crabs in the new progenies. During this time, offspring are produced through the course of fertilization. BMO is measured process and is expressed as follows.

$$p_h^{B\_New} = Rand() \times p^b_{Barnacle\_a} \tag{5}$$

where the significance of rand() is a significance amid zero and one. The use of arbitrary variables and early generalization are two hallmarks of a well-designed optimizer. This approach allows for workforce diversification and facilitates the exploration of potential solutions across all areas of the office. The mutation is likely to provide a new solution that is quite similar to the initial state since the posterior tail of the Gaussian range is short. For this reason, the search procedure iteratively searches every possible solution space position using minimal increments. What follows is a definition of the Gaussian density function.

$$F(p) = \frac{1}{\sqrt{2\pi\sigma^2}} E^{-\frac{(p-\mu)^2}{2\sigma^2}} \tag{6}$$

where μ denotes the expected significance and σ^2 denotes the variance. Assuming $\mu = 0, \sigma^2 = 1$,, this calculation reduces to the produced casual fickle.

$$p_h^* = p_h + J(\partial) \cdot p_h \tag{7}$$

where $J(\partial)$ resembles to the generated Gaussian step trajectory. ∂ is the Gaussian casual significance amid [0,1]. An efficient optimizers must to accomplish impressive levels of search at the commencement of the pursuit and high mistreatment at the end. In BMO, the significance of p_h plays an imperative role in defining the mistreatment and investigation developments. We developed a scientific system to modify the significance of p_h so that it can be dynamically attuned as the iterations decrease. The logistic model and its mathematical expression are given as follows.

$$\begin{cases} \dfrac{dxl(s)}{dt} = \lambda \cdot \left(1 - \dfrac{xl(s)}{xl_{Max}}\right) \cdot xl(s) \\ xl(0) = xl_{Min} \end{cases} \tag{8}$$

where xl_{Max} and xl_{Min} represent the extreme and least significances of xl, individually, where s embodies the amount of iterations and λ signifies the first decline proportion. The technique of parting of variables is used to solve the objective function.

$$xl(s) = \frac{xl_{Max}}{1 + \left(\frac{xl_{Max}}{xl_{Min}} - 1\right) \cdot E^{-\lambda s}} \tag{9}$$

The transition parameter $xl(s) = xl_{Min}$ when s = 0; s $\to \infty, xl(s) = xl_{Max}$. The impact of the alternative stricture on the optimization course is investigated as trails. As cited overhead, a lot of research is obligatory at the initial stage, and a minor significance of p_h will aid the search course. Hence, when s=0, the manipulation stage is usually achieved later the examination stage. As the amount of repetitions surges, the significance of p_h also grows. A large significance of xl is useful for the extraction process. In a dynamic transition system, a better and better balance the refraction index η is calculated as follows.

$$\eta = \frac{Sin\theta_1}{\sin\theta_2} = \frac{((m+n)/2 - p)/i}{(p' - (m+n)/2)/i'} \tag{10}$$

Let the rate $K = \frac{i}{i'}$. can be changed into the subsequent form:

$$p' = (m+n)/2 + (m+n)/(2K\eta) - p/K\eta \tag{11}$$

where a characterizes the higher limit and b characterizes the inferior limit. The p' inverse solution of P is called refractive learning. Here extended to n-measure optimal space.

$$p'_g = (m_g + n_g)/2 + (m_g + n_g)/(2K\eta) - p_g/K\eta \tag{12}$$

where m_g denotes the g-th measurement of higher destined, n_g denotes the g-th measurement of the lower bound. p_g and p'_g are the g-th measurement of p and p', individually. Algorithm 1 depicts the steps involved in the enhanced optimization using EBMO.

C. HOLISTIC ANALYSIS

Holistic analysis in precision farming involves an integrated approach to analyzing agricultural data to provide comprehensive and actionable recommendations. This method looks at the entire ecosystem, including weather patterns, soil condition, and crop health, to make informed decision that enhance productivity, sustainability, and efficiency. Holistic analysis is a thorough and comprehensive examination of data that considers multiple factors and their interactions within a system. In precision farming, this means evaluating a wide

Algorithm 1 Enhanced Optimization Using EBMO

Input: Initial population, fitness function, maximum iteration, threshold condition
Output: Best optimal features

1. Initialize the parameters and the population P
2. Define marry their parents for each iteration

$$Barnacale\_c = Randperm(B) \text{ and}$$
$$Barnacale\_m = Randperm(B)$$

3. **While** s<Maximum iterations **do**
4. Compute Barnacle creates new Spring variables from the parent.

$$p_h^{B\_New} = xp^B_{Barnacle\_c} + yp^B_{Barnacle\_a}$$

5. Compute sperm casting in BMO is process

$$p_h^{B\_New} = Rand() \times p^b_{Barnacle\_a}$$

6 For each variable
7. Update offspring generation
8. Compute Gaussian density function

$$F(p) = \frac{1}{\sqrt{2\pi\sigma^2}} E^{-\frac{(p-\mu)^2}{2\sigma^2}}$$

7. Define separation of variables

$$xl(s) = \frac{xl_{Max}}{1 + \left(\frac{xl_{Max}}{xl_{Min}} - 1\right) \cdot E^{-\lambda s}}$$

8. Compute refraction index

$$\eta = \frac{Sin\theta_1}{\sin\theta_2} = \frac{((m+n)/2 - p)/i}{(p' - (m+n)/2)/i'}$$

9. Formulate n-measurement space

$$p'_g = (m_g + n_g)/2 + (m_g + n_g)/(2K\eta) - p_g/K\eta$$

10. End while
11. Return the best solution

range of variables such as climate data, soil properties, and crop characteristics to understand their combined impact on agricultural outcomes. The goal is to provide full picture of the farming environment and deliver tailored recommendations that optimize farming practices. The recommendation system for precision farming is used to guide farmers in making data-driven decisions that improve crop yield and farm management. Four different ML models are trained on the optimized features.

- A neural network (NN) [33] is used for precision farming recommendations and the input layer receives data like weather conditions and soil properties. The hidden layers, equipped with ReLU activation functions, process this data by learning complex patterns and relationships. The output layer generates predictions or classifications based on the task at hand—continuous significances for regression (e.g., crop yield predictions) or categorical outcomes for classification (e.g., pest presence). ReLU used non-linearity, enhancing the network's ability to capture intricate patterns. Algorithms like Adam adjust network weights to minimize loss and improve performance. Techniques like dropout prevent overfitting, ensuring the model generalizes well to new data. This structure enables the neural network to provide actionable recommendations for precision farming, optimizing agricultural practices for better outcomes.
- A popular ensemble learning method for classification and regression problems is random forest (RF) [34]. It works by training a large number of decision trees and then predicting new data points using the mean or mode of the classes. As a first step, random forest uses bootstrap sampling—which is sampling with replacement—to create numerous subgroups of the training data. A distinct decision tree is trained using each subset. A decision tree takes a selected set of features into account at each split. Each tree contributes a class label for a classification job, and the final forecast is decided by a majority vote. The final prediction for regression tasks is the average of the predicted significances from each tree.
- Decision tree (DT) [35] is an ML model used for both classification and regression tasks. The decision tree splits the dataset into subsets based on the significance of a feature that results in the greatest reduction in impurity. At each node, the tree selects the feature and threshold that best separates the data into distinct groups. This process is repeated recursively to build a tree structure with branches representing different decision paths. For a new data point, the Decision Tree follows the path determined by the feature significances of the data point, traversing from the root to a leaf node. The leaf node provides the final recommendation or prediction. In classification, this is the majority class of the data points in the leaf. In regression, it is the average significance of the target variable in that leaf.
- Support vector machine (SVM) [36] is classifier used for pattern recognition and classification tasks. SVM used to provide actionable recommendations based on various features of the agricultural environment. Collect and preprocess data including features such as soil moisture, temperature, crop type, and historical yield data. The data is normalized and cleaned to ensure accuracy in predictions. Depending on the nature of the data, choose an appropriate kernel function. For non-linear relationships amid features, kernels like RBF are commonly used to transform the input space into a higher-measurement space where a linear separation is possible. During training, the SVM algorithm solves an optimization problem to find this hyperplane. The data points that are closest to the hyperplane and are crucial for determining its position are called support vectors that used to define the decision boundary.

Each model generates predictions based on the input features. These predictions provide insights into various aspects of farming, such as expected crop yield, optimal irrigation schedules, and nutrient management strategies.

D. INTERPRETABLE INSIGHTS

Interpretable insights are crucial for transforming complex machine learning model outputs into understandable and actionable recommendations. In the precision farming, leveraging XAI frameworks such as SHAP (Shapely Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) [37], [38] ensures that the recommendations provided by ML models are transparent and meaningful to users, such as farmers and agricultural specialists. LIME works by approximating the behavior of ML model with a simpler, interpretable model in the vicinity of a particular prediction. LIME generates a local interpretable model around a specific prediction. For example, if the model predicts a particular crop's yield for a given set of weather and soil conditions, LIME will create a simpler model that approximates the decision of the complex model for those specific conditions. The interpretable model identifies which features (e.g., soil moisture, temperature) had the most significant impact on the prediction. For instance, if the recommendation is to increase irrigation, LIME can show how soil moisture levels and historical yield data influenced this decision. LIME provides visualizations that make it easier to grasp the contributions of different features to a prediction. Farmers can see which factors are driving the model's recommendation and adjust their practices accordingly. In order to provide an explanation for a locally faithful prediction, LIME reduces the function of loss L, that involves the initial structure f, the simplest linear model h, and the closeness measure among the instances, which measures the level of detail of the explanation model. Simplifying the model makes the explanation more comprehensible.

$$\xi\left(z\right)=\underset{h\in H}{\arg\min}\;K\left(d,h,\pi_z\right)+\Omega\left(z\right) \tag{13}$$

In order to gain confidence in the overall model's behavior, the LIME developers supported an additional calculation called the Sub-modular Pick (SP). In this case, n is the number of explanations picked by the SP, d is the number of interpretable characteristics, and I is the sum of the importance of all interpretable features in one or more instances. To avoid duplication of effort, a function called c can be used, which calculates the sum of all characteristics' importance in the collection of explanations for Q and U.

$$v\left(C,Q,U\right)=\sum_{h=1}^{f'}1_{[\,\mathrm{E}u\in C;Q_{uh}\geq 0]}U_h \tag{14}$$

The SP finds the situations in set C that have the greatest impact on insurance and adds them together to optimize the protection function.

$$Pick\left(Q,U\right)=\underset{C,|C|\leq B}{\arg\max}\;v\left(C,Q,U\right) \tag{15}$$

With its roots in cooperative game theory, SHAP offers a solid framework for dissecting the prediction process and identifying the role of individual features. As far as comprehensible insights in precision agriculture are concerned, it provides the following advantages: Both at the dataset level and for specific forecasts, SHAP significance levels shed light on how each attribute contributed to the final forecast. By equitably dispersing the significance of the prediction among the characteristics, SHAP significances provide a uniform indicator of feature relevance. For instance, SHAP can determine the relative importance of each feature (such as soil pH and temperature) in a model's forecast of a crop's success under a certain set of circumstances. Instead of using the original, simplified inputs z, explanation models use the following mapping function: These methods' capacity to clarify is equivalent to the direct capability of two components. SHAP takes advantage of the original model's prediction function by adding up the feature contributions u, where z is an integer between 0 and 1, and N is the number of reduced input features.

$$h\left(x'\right)=\phi_0+\sum_{u=1}^{N}\phi_u+x'_u \tag{16}$$

There are three distinct desirable characteristics of SHAP significances. Nearby precision is the first and endorses that the clarification model should have the option to rough the result of the first model either when

$$h\left(x'\right)=\phi_0+\sum_{u=1}^{N}\phi_u+x'_u$$
$$d\left(z\right)=h\left(x'\right)=\phi_0+\sum_{u=1}^{N}\phi_u+x'_u \tag{17}$$

In order for the feature absence to not affect what the model produces, the additional requirement of absence must be satisfied.

$$z'_u=0\Rightarrow\phi_u=0 \tag{18}$$

The initial input should decrease if the impact of a reduced input does not similarly decrease independent of the other properties. This is stated in the third belongings, constancy.

$$d'_z\left(x'\right)-d'_z\left(x'\backslash u\right)\geq d_z\left(x'\right)-d_z\left(x'\backslash u\right) \tag{19}$$

The model must be qualified for each probable subclass A of the total set of structures D in order to compute Shapley significances. A model $d_{A\cup[u]}$ is trained with and without the same feature, $d_{A\cup[u]}$ in order to compute this significance, which compares the predictions from the two models: $d_{A\cup[u]}$ where $d_{A\cup[u]}$ is the subset S's input features. By responding to this procedure for each feature, a feature attribution i can be obtained for each observation.

$$\phi_u=\sum_{A\subseteq M\backslash u}\frac{|A|\left(|D|-|A|-1\right)}{|N|}\left[d_{A\cup[u]}\left(z_{A\cup[u]}\right)-d_a\left(z_a\right)\right] \tag{20}$$

TABLE 2. Statistical characteristics of soil dataset.

Description	District				
	Anantapur	Chitoor	Guntur	Kadapa	Nellore
Nitrogen	Low	Medium	Low	Low	Medium
Phosphorus	Low	Low	Low	Low	Low
Potassium	Medium	Low	Low	Low	Low
Soil type	Red	Red	Black	Red + Black	Alluvial
Soil depth	100-300	0-25	300	100-300	300
pH	Neutral	Slightly acidic	Strongly alkaline	Strongly alkaline	Slightly alkaline

The model-specific tree interpreter determines the contributions based on a portion S of features, where S represents the non-zero z-indexes and N is the set of all of the input characteristics, in contrast to the model-agnostic SHAP kernel explainer, which determines the mean minimal impact of each feature. The expected significance is $[d_z (A \cup \{u\}) - d_z (A)]$

$$\phi_u = \sum_{A \subseteq M \setminus u} \frac{|A| (N - A - 1)}{|N|} [d_z (A \cup \{u\}) - d_z (A)] \quad (21)$$

DeepLIFT adapts the innovative significances into binary implications using the mapping function $[d_z(A \cup \{u\}) - d_z(A)]$ where 0 embodies the input xi enchanting the situation significance and 1 represents the original implication. We chains individual neural network gears into the model as a whole, which are defined as $n_{\Delta z_u \Delta z_y}$

$$n_{\Delta z_u \Delta z_y} = \frac{V_{\Delta z \Delta y}}{\Delta y} \quad (22)$$

The output of both inputs is represented by, the difference between the input and the standard deviation is denoted by y, and the effect of the two sources is z. By using LIME and SHAP into the recommendation system, we ensure that the AI-driven recommendations for precision farming are not only accurate but also understandable. Farmers and stakeholders can see clear reasons behind each recommendation, fostering trust and enabling them to make more informed decisions.

IV. RESULTS AND DISCUSSION

This Unit delves into the results and comparative analysis of recommendation models designed for a smart agriculture system. To gauge the effectiveness of both new and existing models, three core datasets were employed: weather, soil, and crop data. This data encompassed key elements influencing crop forecasts, such as soil characteristics, weather conditions, and agricultural practices, and was collected from 34 districts across South India (Tamil Nadu, Andhra Pradesh, Telangana, Karnataka, and Kerala) over the period from 2001 to 2015. The data was meticulously organized into three distinct datasets, each reflecting the relevant factors. The proposed models were developed using Python on a system equipped with Microsoft Windows 10, an Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz, 16.0GB of RAM, and 512 GB of SSD storage. To scrutinize individual predictions, the SHAP and LIME packages were utilized for in-depth analysis.

A. STATISTICAL CHARACTERISTICS OF DATASET

Table 2 offers a comprehensive summary of soil characteristics across five districts: Anantapur, Chitoor, Guntur, Kadapa, and Nellore. In terms of nitrogen levels, Anantapur and Kadapa exhibit low concentrations, Chitoor shows medium levels, while Guntur and Nellore also present low levels. Phosphorus levels are consistently low throughout all districts. For potassium content, Anantapur has medium levels, whereas Chitoor, Guntur, Kadapa, and Nellore report low levels. The soil types vary among the districts: Anantapur and Chitoor have red soil, Guntur features black soil, Kadapa has a combination of red and black soil, and Nellore is characterized by alluvial soil. Soil depth also differs, with Anantapur and Kadapa having depths ranging from 100 to 300 cm, Chitoor having a very shallow depth of 0 to 25 cm, and both Guntur and Nellore showing a depth of 300 cm. Lastly, the pH levels vary: Anantapur soil is neutral, Chitoor's is slightly acidic, Guntur's and Kadapa's soils are strongly alkaline, and Nellore's soil is slightly alkaline. Table 3 provides a detailed overview of weather characteristics recorded in Adilabad over five years: 2001, 2002, 2003, 2004, and 2005.

- *Pressure*: The atmospheric pressure data shows a slight increasing trend over the years, ranging from 95.623 in 2001 to 97.998 in 2005.
- *Temperature*: The average temperature exhibits a gradual increase from 27.152°C in 2001 to 28.999°C in 2005.
- Humidity: Relative humidity significances are also on the rise, starting from 12.345% in 2001 and reaching 12.985% by 2005.
- *Wind speed*: Wind speed measurements show a gradual increase over the years, from 3.158 m/s in 2001 to 4.125 m/s in 2005.
- *Maximum temperature*: The highest recorded temperatures each year slightly increase, from 47.125°C in 2001 to 47.985°C in 2005.
- *Minimum temperature*: The lowest temperatures recorded have a minor fluctuation, ranging from 8.125°C in 2001 to 8.645°C in 2005.

TABLE 3. Statistical characteristics of weather dataset (Adilabad District).

Year	2001	2002	2003	2004	2005
Pressure	95.623	97.458	97.125	97.859	97.998
Temperature	27.152	28.365	28.565	28.758	28.999
Humidity	12.345	12.089	12.457	12.689	12.985
Wind Speed	3.158	3.247	3.654	3.989	4.125
Maximum temperature	47.125	47.352	47.555	47.689	47.985
Minimum temperature	8.125	8.325	8.457	8.325	8.645
Cloud amount	56.145	48.528	47.135	55.968	53.689
Precipitation	3.225	3.105	3.454	3.565	3.689
UVA Irradiance	14.588	14.254	14.168	14.609	14.578
UVB Irradiance	0.458	0.425	0.498	0.512	0.524
Downward Irradiance	6.145	6.245	6.347	6.458	6.788
PAR Total	100.125	98.562	99.347	99.858	98.858

TABLE 4. Statistical characteristics of agricultural dataset (Anantapur District).

Crop type	Year	Area (1000 ha)	Yield (Kg per ha)	Irrigated area (1000 ha)
Rice	2001	71	2880.561	70.944
	2002	40	2150	39.925
	2003	28.341	2482.361	28.325
	2004	33.588	3176.599	33.504
	2005	48.155	2607.688	48.066

- *Cloud amount*: The cloud cover percentage varies year-to-year, with a high of 56.145% in 2001 and a low of 47.135% in 2003, with some fluctuations in amid.
- *Precipitation*: Annual precipitation amounts show a gradual increase, starting at 3.225 mm in 2001 and reaching 3.689 mm by 2005.
- *UVA irradiance*: UVA irradiance levels vary slightly from 14.588 W/m^2 in 2001 to 14.578 W/m^2 in 2005, indicating a relatively stable level of ultraviolet A radiation.
- *UVB irradiance*: UVB irradiance levels increase marginally from 0.458 W/m^2 in 2001 to 0.524 W/m^2 in 2005.
- *Downward irradiance:* The downward irradiance, which measures the amount of solar radiation reaching the Earth's surface, shows a steady rise from 6.145 W/m^2 in 2001 to 6.788 W/m^2 in 2005.
- *PAR total:* Photo synthetically active radiation (PAR) exhibits some fluctuation, with significances ranging from 100.125 W/m^2 in 2001 to 98.858 W/m^2 in 2005.

Table 4 provides a detailed account of rice cultivation in Anantapur from 2001 to 2005. It includes key metrics such as the cultivated area, yield, and irrigated area for each year. In 2001, the area under rice cultivation was 71,000 hectares with a yield of 2880.561 kg per hectare and an irrigated area of 70.944 hectares. Over the following years, the area dedicated to rice farming fluctuated, decreasing to 40,000 hectares in 2002 and further to 28.341 hectares in 2003. By 2004, the area increased to 33.588 hectares and further rose to 48.155 hectares in 2005. The yield also varied significantly, starting at 2880.561 kg/ha in 2001, dropping to 2150 kg/ha in 2002, and then increasing to a peak of 3176.599 kg/ha in 2004 before declining to 2607.688 kg/ha in 2005. The irrigated area followed a similar trend, initially at 70.944 hectares in 2001, reducing to 39.925 hectares in 2002, and further decreasing to 28.325 hectares in 2003. It slightly increased to 33.504 hectares in 2004 and then rose again to 48.066 hectares in 2005.

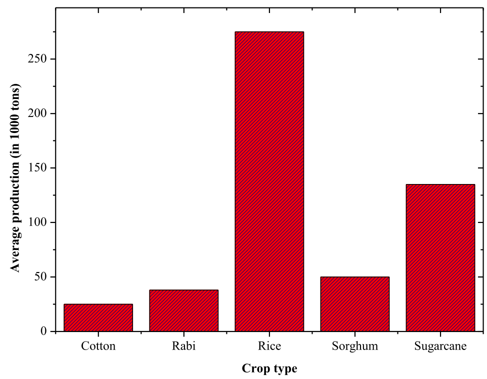

FIGURE 2. Normal manufacture (in 1000 tons) for eachharvest typeS.

Manufacture statistics from 2001 to 2015 across districts in South India—comprising Kerala, Karnataka, Telangana, Andhra Pradesh, and Tamil Nadu—reveals that rice was the predominant harvest through this age. This is largely due to the area's warm and wet weather, which creates ideal situations for rice farming. Other harvests, such as sugarcane and sorghum, also show substantial production levels. In contrast, crops typically grown during the Rabi season exhibited much lower production figures, indicating that their farming was limited likened to other harvests, likely due to their penchant for cooler and drier conditions. Fig. 2 shows the usual manufacture (in 1000 tons) for each harvest type.

TABLE 5. Performance of recommendation models for precision farming.

Models	Performance metrics											
	MAE	MSE	RMSE	R2	MAE	MSE	RMSE	R2	MAE	MSE	RMSE	R2
	Rice				Sugarcane				Sorghum			
Linear	43.766	3178	56.385	0.9215	61.960	7581	87.074	0.7196	38.452	2902	53.845	0.4285
XGB	44.875	3799	61.644	0.9056	34.900	2640	51.395	0.9023	22.011	1070	32.714	0.7893
RF	41.752	3205	56.621	0.9230	31.302	3377	49.972	0.9076	16.695	721	26.864	0.8579
k-NN	50.023	4558	67.528	0.8867	39.574	2063	58.112	0.8751	36.674	2651	51.502	0.4778
LGBM	44.865	3735	61.263	0.9067	26.885	3080	45.421	0.9237	16.354	525	22.936	0.8965
DT	48.652	5253	72.365	0.8699	33.105	2922	55.503	0.8861	20.495	1166	34.152	0.7704
XLNet+NN	39.546	3058	45.285	0.9355	25.125	2541	35.266	0.9258	15.128	508	25.198	0.9158
XLNet+RF	39.124	2985	43.159	0.9398	24.162	2345	35.155	0.9345	14.625	485	24.858	0.9255
XLNet+DT	38.985	2748	42.198	0.9412	23.585	2285	35.089	0.9458	13.255	475	24.455	0.9366
XLNet+SVM	38.562	2566	40.255	0.9455	23.005	2158	34.399	0.9498	13.089	465	24.065	0.9398
	Rabi				Cotton							
Linear	35.515	3276	57.241	0.5012	43.512	4045	63.601	0.1255				
XGB	22.052	2043	45.215	0.6889	16.114	443	21.077	0.5839				
RF	16.408	923	30.385	0.8595	9.245	183	13.546	0.8281				
k-NN	28.975	2489	49.895	0.6211	10.068	242	15.598	0.7723				
LGBM	14.623	768	27.735	0.8832	10.295	238	15.432	0.7768				
DT	14.298	492	22.194	0.9254	9.366	237	15.402	0.7778				
XLNet+NN	13.005	475	21.058	0.9565	8.353	228	13.255	0.9125				
XLNet+RF	12.858	463	20.858	0.9586	7.985	224	13.154	0.9190				
XLNet+DT	12.454	462	20.147	0.9599	7.588	220	13.008	0.9255				
XLNet+SVM	12.085	451	19.658	0.9612	7.465	218	12.989	0.2980				

B. RESULTS COMPARISON

This section discuss the results comparison of proposed models XLNet+NN, XLNet+RF, XLNet+DT and XLNet+SVM with the existing models such as linear, gradient boosting (XGB), random forest (RF), k-nearest neighbor (k-NN), LGBM, decision tree (DT), bagging. Table 5 presents the performance results of various recommendation models for precision farming focused on rice crop cultivation. The models are evaluated based on metrics including mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and R-squared (R^2). For rice, the XLNet+SVM model achieves the best performance with an MAE of 38.562, MSE of 2566, RMSE of 40.255, and R^2 of 0.9455. Compared to the linear model, which has an MAE of 43.766, an MSE of 3178, and an RMSE of 56.385, XLNet+SVM reduces MAE by 12.0%, MSE by 19.3% and RMSE by 28.6%. The R^2 significance also improves by 2.6%. Among other models, XLNet+DT and XLNet+RF also perform notably well, with MAEs of 38.985 and 39.124, respectively, showing decreases of 11.1% and 10.6% compared to the linear model. MSE reductions are 13.5% and 12.0%, and RMSE reductions are 25.1% and 22.6%. The R^2 significances for these models are 2.1% to 2.7% higher than the linear model, indicating better fit.

For sugarcane, XLNet+SVM again leads with an MAE of 23.005, MSE of 2158, and RMSE of 34.399, which represents a significant improvement over the linear model's MAE of 61.960, MSE of 7581, and RMSE of 87.074. XLNet+SVM achieve reductions of 62.8% in MAE, 71.6% in MSE, and 60.5% in RMSE. The R^2 significance increases by 31.9%. XLNet+DT and XLNet+RF also demonstrate substantial performance gains, with MAEs of 23.585 and 24.162, reflecting decreases of 61.8% and 61.0%. MSE reductions are 69.9% and 69.0%, while RMSE decreases are 60.6% and 60.7%. R^2 significances are 31.5% and 29.7% higher compared to the linear model. For sorghum, XLNet+SVM excels with an MAE of 13.089, an MSE of 465, and an RMSE of 24.065, compared to the linear model's MAE of 38.452, MSE of 2902, and RMSE of 53.845. This results in reductions of 66.0% in MAE, 84.0% in MSE, and 55.3% in RMSE, with an R^2 improvement of 119.7%. XLNet+DT and XLNet+RF also show impressive performance, with MAEs of 13.255 and 14.625, respectively, which represent decreases of 65.5% and 62.1%. MSE reductions are 83.7% and 83.3%, and RMSE reductions are 55.0% and 54.9%. R^2 significances improve by 95.6% and 93.4%. For Rabi and cotton, the XLNet-based models again achieve superior performance. For Rabi, XLNet+SVM has the lowest MAE of 12.085 and

MSE of 451, reflecting reductions of 66.0% in MAE and 86.1% in MSE compared to the linear model. For cotton, XLNet+SVM shows a significant MAE reduction of 82.8% and MSE reduction of 83.0% compared to traditional models. Overall, XLNet-based models consistently outperform other models across different crop types, shows improvements in accuracy and error metrics.

Table 6 presents the accuracy of various recommendation models for precision farming across different crop types, including rice, sugarcane, sorghum, Rabi, and cotton. XLNet+SVM emerge as the top-performing model across all crops. It achieves the highest accuracy for each crop type, with scores of 98.565% for rice, 98.098% for sugarcane, 98.656% for sorghum, 98.354% for Rabi, and 98.986% for cotton.

TABLE 6. Accuracy comparison of recommendation models for precision farming.

Model	Accuracy (%)				
	Rice	Sugarcane	Sorghum	Rabi	Cotton
Linear	85.077	84.610	85.168	84.866	85.497
XGB	85.646	85.179	85.737	85.435	86.066
RF	86.214	85.747	86.305	86.003	86.635
k-NN	86.783	86.316	86.874	86.572	87.204
LGBM	87.352	86.885	87.443	87.141	87.773
DT	87.921	87.454	88.012	87.710	88.342
XLNet+NN	98.490	98.023	98.581	98.279	98.911
XLNet+RF	98.515	98.048	98.606	98.304	98.936
XLNet+DT	98.540	98.073	98.631	98.329	98.961
XLNet+SVM	98.565	98.098	98.656	98.354	98.986

As illustrated in Fig. 3, compared to the linear model, which achieves accuracies ranging from 84.610% to 85.497%, XLNet combined with SVM shows a significant improvement. Specifically, it increases accuracy by 16.0% for rice, 13.9% for sugarcane, 13.8% for sorghum, 13.4% for Rabi, and 13.8% for cotton. XLNet+DT and XLNet+RF also demonstrate excellent performance, with accuracies very close to XLNet+SVM. XLNet+DT achieve accuracies of 98.540% for rice, 98.073% for sugarcane, 98.631% for sorghum, 98.329% for Rabi, and 98.961% for cotton. XLNet+RF shows slightly lower but still impressive accuracy levels: 98.515% for rice, 98.048% for sugarcane, 98.606% for sorghum, 98.304% for Rabi, and 98.936% for cotton. Compared to the linear model, XLNet+DT increases accuracy by 13.8% to 14.5%, while XLNet+RF improves accuracy by 13.9% to 14.4%. In comparison, traditional models like k-NN, LGBM, and DT offer lower accuracy rates. For instance, k-NN achieves accuracy scores ranging from 86.783% for rice to 87.204% for cotton, indicating improvements of 1.3% to 1.7% over the linear model. LGBM and DT show better accuracy than k-NN, with LGBM achieving up to 87.773% for cotton and DT reaching 88.342% for cotton. Compared to the linear model, LGBM improves accuracy by 2.7% to 3.5%, while DT shows increases of 3.4% to 4.2%. XLNet-based models consistently outperform all other models in accuracy across all crop types. An improvement over traditional models, especially the linear model, show the significant advancements in predictive performance achieved with ML techniques like XLNet+SVM.

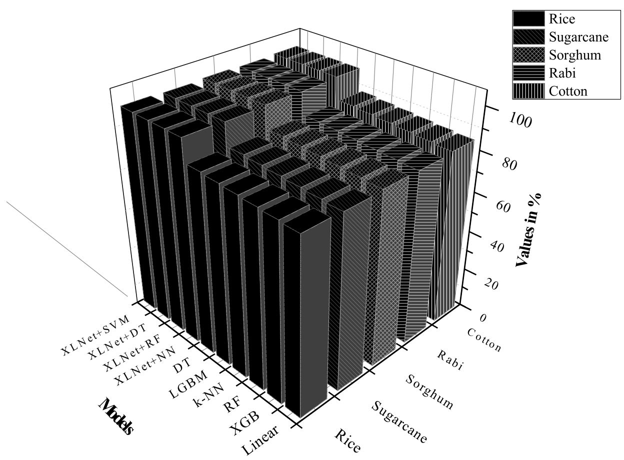

FIGURE 3. Accuracy comparison of recommendation models for precision farming with different crop types.

C. RESULTS INTERPRETATION

From the comparative analysis, we found that XLNet+SVM model achieves the effective results in terms of MAE, MSE, RMSE, R^2 and accuracy. This Unit discusses the interpretation of results of XLNet+SVM model. The explain instance method in the LIME class's tabular explainer, shows in Fig. 4, provides an intuitive and interactive HTML visualization for understanding individual predictions. The visualization is used to explain predictions from XLNet+SVM model, focus on classification tasks such as distinguishing amid rice and non-rice, and sugarcane and non-sugarcane. The HTML output features a comprehensive display of prediction probabilities, offering a clear view of how likely the model considers the instance to belong to each class. Additionally, the visualization shows the coefficients of various features, showing their impact on the prediction outcome. Features that positively influence the prediction are indicated alongside those that have a negative effect, allowing users to grasp which aspects of the data are driving the model's decision. The SHAP summary plot visualizing feature importance, offering insights into how individual features influence the model's predictions on a broader scale as well as for specific classes. For instance, the plot reveals that features such as humidity and PAR are crucial for predicting the "Rice" class. This method also enables a detailed examination of a single class's prediction, allowing us to see clearly how each feature contributes to the model's output.

As depicted in Fig. 5, the summary plot highlights that the XLNet+SVM model's predictions for different crops are

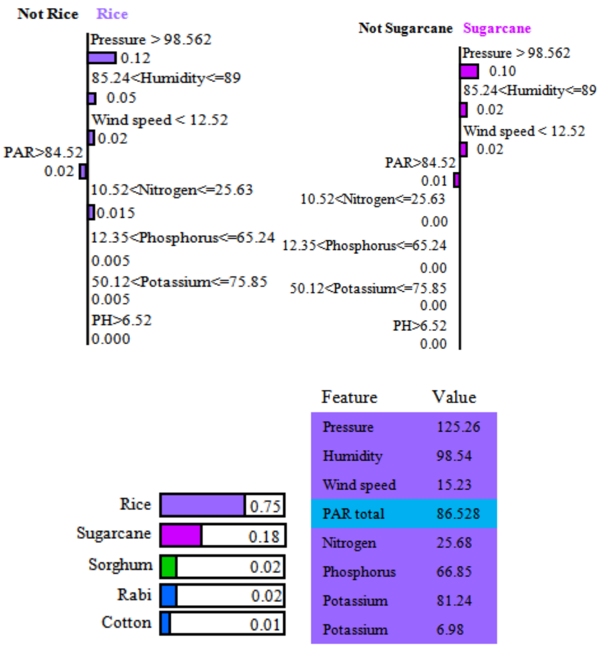

FIGURE 4. XLNet+SVM explainable instance with prediction profanities, classification of rice/not-rice and sugarcane/not-sugarcane.

significantly driven by weather and soil parameters. Specifically, it shows that high significances of both humidity and PAR are strongly associated with the prediction of apples. The plot effectively shows that the XLNet+SVM model relies heavily on the key environmental factors, provides a transparent view into the model's predictive behavior. Using LIME's Sub modular Pick, as detailed in the LIME Unit, we can generate a matrix comprising instances and their associated interpretable features.

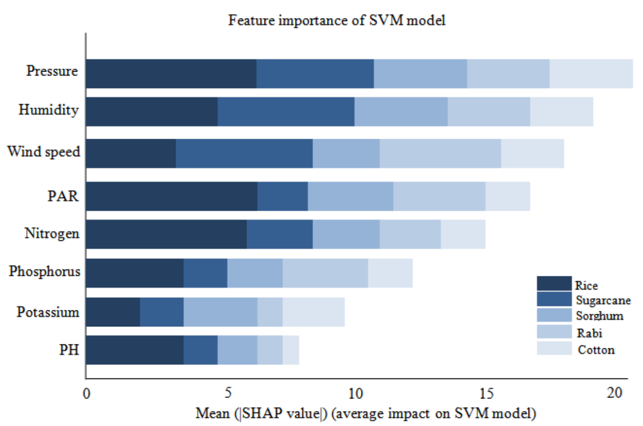

FIGURE 5. SHAP summary plot of XLNet+SVM model for precision farming with different crop types.

Fig. 6 shows the average impact of the selected features within a subset of 50 instances for which 10 explanations were sought. The visualization reveals that the feature with the highest average impact is Total PAR, with a mean effect significance exceeding 84.52. Conversely, features such as

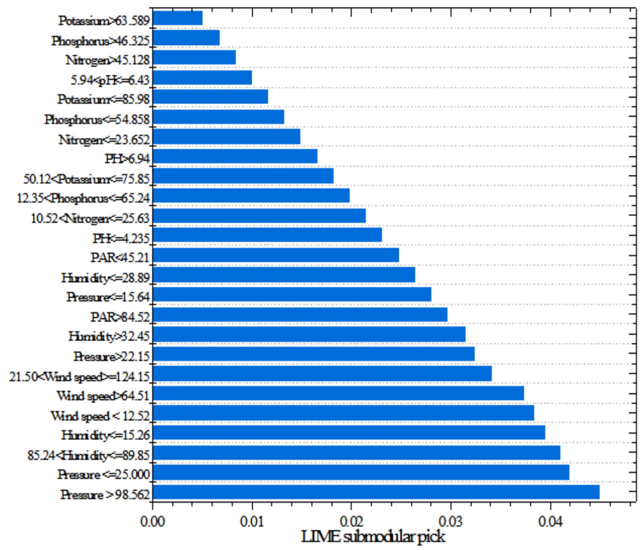

FIGURE 6. Lime sub modular pick of XLNet+SVM model for precision farming with different crop types.

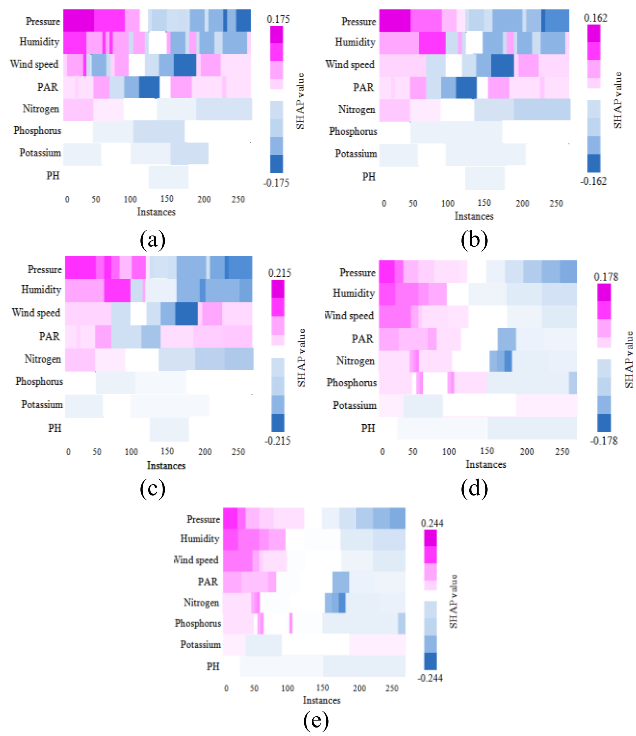

FIGURE 7. Heat map plot results of SVM model for recommendation offarming with different crop types (a) Rice (b) Sugarcane (c) Sorghum (d) Rabi and (e) Cotton.

pressure, humidity, wind speed, and others like nitrogen, phosphorus, potassium, and pH, exhibit minimal or even negative impacts on the predictions. This observation aligns with the insights from the SHAP summary plot, reinforcing the consistency of the findings across different interpretability methods.

An insightful visualization is provided by the SHAP heatmap plot, which illustrates the extent to which each feature affects instances within the test set, as well as the overall importance of each feature for predicting individual classes. Fig. 7 presents the heatmap results from the XLNet+SVM model for different crop types. By comparing the heatmaps, particularly the one for rice with the heatmaps for other classes, we gain a clearer understanding of why certain classes may not always be accurately predicted by the model. The heatmaps reveal that the model uses similar features to classify all crops, demonstrating a generally consistent behavior across different crop types. However, a closer examination shows that the rainfall attribute has a notably positive impact on the classification of rice, sugarcane, sorghum, and Rabi, compared to its effect on cotton.

V. CONCLUSION

The proposed XAI-based smart agriculture framework provides a comprehensive and efficient solution for precision farming, achieving significant enhancements in agricultural productivity while addressing key challenges of data complexity and interpretability. Our method utilizes a well-curated dataset from trusted Indian sources, incorporating data.icrisat.org for crop data, power.larc.nasa.gov for meteorological insights, and geoportal.natmo.gov.in for soil characteristics. By employing XLNet for deep feature extraction, the model successfully identifies intricate patterns that drive accurate predictions. To handle high-dimensional agricultural data and ensure computational efficiency, the Enhanced Barnacles Mating Optimization (EBMO) algorithm was applied for optimal feature selection, enabling improved model performance. Our experiments demonstrated that the XLNet+SVM model achieved superior accuracy rates: 98.565% for rice, 98.098% for sugarcane, 98.656% for sorghum, 98.354% for Rabi crops, and 98.986% for cotton. These results represent notable improvements over existing models, with accuracy gains of up to 2.563%. Importantly, we addressed the necessity for explainable AI in agriculture by integrating XAI tools like SHAP and LIME, which elucidate model predictions. The explainable patterns and interpretable features not only provide clarity into how features influence outcomes but also align AI decisions with agricultural knowledge. This interpretability is crucial for several reasons: In AI-driven systems used in critical fields like agriculture, ethical and legal accountability becomes imperative. XAI methods help ensure transparency, allowing stakeholders to understand decision-making processes and address potential ethical considerations against established knowledge. The ability to interpret and explain AI predictions allows researchers and practitioners to compare model outcomes with existing agricultural literature, thus validating results and enhancing credibility. Stakeholders, including farmers and policymakers, are more likely to adopt AI systems when they can comprehend and trust the recommendations. XAI frameworks play a pivotal role in establishing this trust, making AI applications safer and more reliable for real-world use. The integration of the stability tools not only enhances model transparency but also positions the proposed framework as a robust and trustworthy solution, supporting ethical, verifiable, and widely accepted AI-driven agriculture. Future work could further explore the role of explainable AI in addressing ethical and regulatory concerns, increasing AI model adoption, and strengthening the alignment between AI insights and practical agricultural practices.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing interests.

DATA AVAILABILITY

The input data used in the analyses are from published sources that have been cited in the manuscript. The data generated in this study will be made available upon request.

ACKNOWLEDGMENT

The authors gratefully acknowledge the funding of the Deanship of Graduate Studies and Scientific Research, Jazan University, Saudi Arabia, through project number: RG24-S0140.

REFERENCES

[1] A. Corallo, M. E. Latino, M. Menegoli, and F. Signore, "Digital technologies for sustainable development of agri-food: Implementation guidelines toward industry 5.0," *IEEE Trans. Eng. Manag.*, vol. 71, pp. 10699–10715, 2024, doi: 10.1109/TEM.2024.3403251.

[2] M. E. Latino, A. Corallo, M. Menegoli, and B. Nuzzo, "Agriculture 4.0 as enabler of sustainable agri-food: A proposed taxonomy," *IEEE Trans. Eng. Manag.*, vol. 70, no. 10, pp. 3678–3696, Oct. 2021, doi: 10.1109/TEM.2021.3101548.

[3] R. K. Munaganuri and Y. N. Rao, "PAMICRM: Improving precision agriculture through multimodal image analysis for crop water requirement estimation using multidomain remote sensing data samples," *IEEE Access*, vol. 12, pp. 52815–52836, 2024, doi: 10.1109/ACCESS.2024.3386552.

[4] M. R. Mahmood, M. A. Matin, S. K. Goudos, and G. Karagiannidis, "Machine learning for smart agriculture: A comprehensive survey," *IEEE Trans. Artif. Intell.*, vol. 5, no. 6, pp. 2568–2588, Jun. 2024, doi: 10.1109/TAI.2023.3345278.

[5] H. Chen, J. Xie, L. Xu, Q. Feng, Q. Lin, and K. Cai, "Feature selection for portable spectral sensing data of soil using broad learning network in fusion with fuzzy technique," *IEEE Sensors J.*, vol. 24, no. 5, pp. 5644–5653, Mar. 2024, doi: 10.1109/JSEN.2023.3331026.

[6] A. Reyana, S. Kautish, P. M. S. Karthik, I. A. Al-Baltah, M. B. Jasser, and A. W. Mohamed, "Accelerating crop yield: Multisensor data fusion and machine learning for agriculture text classification," *IEEE Access*, vol. 11, pp. 20795–20805, 2023, doi: 10.1109/ACCESS.2023.3249205.

[7] P. Garg, P. Khaparde, K. S. Patle, C. Bhaliya, A. Kumar, M. V. Joshi, and V. S. Palaparthy, "Environmental and soil parameters for germination of leaf spot disease in the groundnut plant using IoT-enabled sensor system," *IEEE Sensors Lett.*, vol. 7, no. 12, pp. 1–4, Dec. 2023, doi: 10.1109/LSENS.2023.3330923.

[8] H. Tong, E. Xia, C. Sun, K. Yan, J. Li, and J. Huang, "Construction and comprehensive evaluation of an index system for climate-smart agricultural development in China," *J. Cleaner Prod.*, vol. 469, Sep. 2024, Art. no. 143216, doi: 10.1016/j.jclepro.2024.143216.

[9] A. R. Mahlous, "Security analysis in smart agriculture: Insights from a cyber-physical system application," *Comput., Mater. Continua*, vol. 79, no. 3, pp. 4781–4803, 2024, doi: 10.32604/cmc.2024.050821.

[10] Z. H. Zul Azlan, S. N. Junaini, N. A. Bolhassan, R. Wahi, and M. A. Arip, "Harvesting a sustainable future: An overview of smart agriculture's role in social, economic, and environmental sustainability," *J. Cleaner Prod.*, vol. 434, Jan. 2024, Art. no. 140338, doi: 10.1016/j.jclepro.2023.140338.

[11] L. Cheng, A. Samah, Y. Zhang, Z. Jia, and F. Li, "Towards urban energy sustainability and resiliency through smart agriculture and effective renewable resource allocation," *Sustain. Cities Soc.*, vol. 107, Jul. 2024, Art. no. 105420, doi: 10.1016/j.scs.2024.105420.

[12] J. Li, M. Xu, L. Xiang, D. Chen, W. Zhuang, X. Yin, and Z. Li, "Foundation models in smart agriculture: Basics, opportunities, and challenges," *Comput. Electron. Agricult.*, vol. 222, Jul. 2024, Art. no. 109032, doi: 10.1016/j.compag.2024.109032.

[13] K. Zidi, K. Ben Abdellafou, A. Aljuhani, O. Taouali, and M. F. Harkat, "Novel intrusion detection system based on a downsized kernel method for cybersecurity in smart agriculture," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108579, doi: 10.1016/j.engappai.2024.108579.

[14] R. Kumar Kasera, S. Gour, and T. Acharjee, "A comprehensive survey on IoT and AI based applications in different pre-harvest, during-harvest and post-harvest activities of smart agriculture," *Comput. Electron. Agricult.*, vol. 216, Jan. 2024, Art. no. 108522, doi: 10.1016/j.compag.2023.108522.

[15] R. Dwivedi, P. Kothari, D. Chopra, M. Singh, and R. Kumar, "An efficient ensemble explainable AI (XAI) approach for morphed face detection," *Pattern Recognit. Lett.*, vol. 184, pp. 197–204, Aug. 2024, doi: 10.1016/j.patrec.2024.06.014.

[16] H. Jang, S. Kim, and B. Yoon, "An eXplainable AI (XAI) model for text-based patent novelty analysis," *Expert Syst. Appl.*, vol. 231, Nov. 2023, Art. no. 120839, doi: 10.1016/j.eswa.2023.120839.

[17] S. Rajpal, A. Rajpal, M. Agarwal, V. Kumar, A. Abraham, D. Khanna, and N. Kumar, "XAI-CNVMarker: Explainable AI-based copy number variant biomarker discovery for breast cancer subtypes," *Biomed. Signal Process. Control*, vol. 84, Jul. 2023, Art. no. 104979, doi: 10.1016/j.bspc.2023.104979.

[18] M. Jonak, J. Mucha, S. Jezek, D. Kovac, and K. Cziria, "SPAGRI-AI: Smart precision agriculture dataset of aerial images at different heights for crop and weed detection using super-resolution," *Agricult. Syst.*, vol. 216, Apr. 2024, Art. no. 103876, doi: 10.1016/j.agsy.2024.103876.

[19] Á. L. P. Gómez, P. E. López-de-Teruel, A. Ruiz, G. García-Mateos, G. B. García, and F. J. G. Clemente, "FARMIT: Continuous assessment of crop quality using machine learning and deep learning techniques for IoT-based smart farming," *Cluster Comput.*, vol. 25, no. 3, pp. 2163–2178, Jun. 2022, doi: 10.1007/s10586-021-03489-9.

[20] K. Sanogo, I. Touré, D.-D.-A. A. Arinloye, E. R. Dossou-Yovo, and J. Bayala, "Factors affecting the adoption of climate-smart agriculture technologies in Rice farming systems in Mali, West Africa," *Smart Agricult. Technol.*, vol. 5, Oct. 2023, Art. no. 100283, doi: 10.1016/j.atech.2023.100283.

[21] M. Aldossary, H. A. Alharbi, and C. A. U. Hassan, "Internet of Things (IoT)-enabled machine learning models for efficient monitoring of smart agriculture," *IEEE Access*, vol. 12, pp. 75718–75734, 2024, doi: 10.1109/ACCESS.2024.3404651.

[22] E.-S.-M. El-Kenawy, A. A. Alhussan, N. Khodadadi, S. Mirjalili, and M. M. Eid, "Predicting potato crop yield with machine learning and deep learning for sustainable agriculture," *Potato Res.*, Jul. 2024, doi: 10.1007/s11540-024-09753-w.

[23] B. Et-taibi, M. R. Abid, E.-M. Boufounas, A. Morchid, S. Bourhnane, T. A. Hamed, and D. Benhaddou, "Enhancing water management in smart agriculture: A cloud and IoT-based smart irrigation system," *Results Eng.*, vol. 22, Jun. 2024, Art. no. 102283, doi: 10.1016/j.rineng.2024.102283.

[24] S. Mandal, A. Yadav, F. A. Panme, K. M. Devi, and S. Kumar, "Adaption of smart applications in agriculture to enhance production," *Smart Agricult. Technol.*, vol. 7, Mar. 2024, Art. no. 100431, doi: 10.1016/j.atech.2024.100431.

[25] A. Morchid, I. G. M. Alblushi, H. M. Khalid, R. El Alami, S. R. Sitaramanan, and S. M. Muyeen, "High-technology agriculture system to enhance food security: A concept of smart irrigation system using Internet of Things and cloud computing," *J. Saudi Soc. Agricult. Sci.*, Feb. 2024, doi: 10.1016/j.jssas.2024.02.001.

[26] K. V. Reddy, B. S. K. Reddy, V. Goutham, M. Mahesh, J. S. Nisha, G. Palanisamy, M. Golla, S. Purushothaman, K. R. Reddy, and V. Ramkumar, "Edge AI in sustainable farming: Deep learning-driven IoT framework to safeguard crops from wildlife threats," *IEEE Access*, vol. 12, pp. 77707–77723, 2024, doi: 10.1109/ACCESS.2024.3406585.

[27] M. F. Shahid, T. J. S. Khanzada, M. A. Aslam, S. Hussain, S. A. Baowidan, and R. B. Ashari, "An ensemble deep learning models approach using image analysis for cotton crop classification in AI-enabled smart agriculture," *Plant Methods*, vol. 20, no. 1, Jul. 2024, Art. no. 104, doi: 10.1186/s13007-024-01228-w.

[28] L. Yu et al., "Near surface camera informed agricultural land monitoring for climate smart agriculture," *Climate Smart Agricult.*, vol. 1, no. 1, Aug. 2024, Art. no. 100008, doi: 10.1016/j.csag.2024.100008.

[29] U. Nikhil, A. Pandiyan, S. Raja, and Z. Stamenkovic, "Machine learning-based crop yield prediction in South India: Performance analysis of various models," *Computers*, vol. 13, no. 6, p. 137, May 2024, doi: 10.3390/computers13060137.

[30] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," 2019, *arXiv:1906.08237*.

[31] M. Onat Topal, A. Bas, and I. van Heerden, "Exploring transformers in natural language generation: GPT, BERT, and XLNet," 2021, *arXiv:2102.08036*.

[32] M. H. Sulaiman, Z. Mustaffa, M. M. Saari, and H. Daniyal, "Barnacles mating optimizer: A new bio-inspired algorithm for solving engineering optimization problems," *Eng. Appl. Artif. Intell.*, vol. 87, Jan. 2020, Art. no. 103330, doi: 10.1016/j.engappai.2019.103330.

[33] A. Kaveh, *Applications of Artificial Neural Networks and Machine Learning in Civil Engineering* (Studies in computational intelligence). Cham, Switzerland: Springer, 2024, doi: 10.1007/978-3-031-66051-1.

[34] Z. Sun, G. Wang, P. Li, H. Wang, M. Zhang, and X. Liang, "An improved random forest based on the classification accuracy and correlation measurement of decision trees," *Expert Syst. Appl.*, vol. 237, Mar. 2024, Art. no. 121549, doi: 10.1016/j.eswa.2023.121549.

[35] Y. Ma, H. Zhang, Y. Cai, and H. Yang, "Decision tree for locally private estimation with public data," in *Proc. 37th Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Red Hook, NY, USA: Curran Associates, 2024, pp. 43676–43705.

[36] H. Wang and Y. Shao, "Fast generalized ramp loss support vector machine for pattern classification," *Pattern Recognit.*, vol. 146, Feb. 2024, Art. no. 109987, doi: 10.1016/j.patcog.2023.109987.

[37] V. Vimbi, N. Shaffi, and M. Mahmud, "Interpreting artificial intelligence models: A systematic review on the application of LIME and SHAP in Alzheimer's disease detection," *Brain Informat.*, vol. 11, no. 1, Dec. 2024, Art. no. 10, doi: 10.1186/s40708-024-00222-1.

[38] D. Gaspar, P. Silva, and C. Silva, "Explainable AI for intrusion detection systems: LIME and SHAP applicability on multi-layer perceptron," *IEEE Access*, vol. 12, pp. 30164–30175, 2024, doi: 10.1109/ACCESS.2024.3368377.

R. JOHN MARTIN (Member, IEEE) received the Ph.D. degree in computer science from Bharathiar University, India, with a focus on an innovative machine-learning model for epileptic seizure detection using EEG signals. He is currently a Distinguished Educator in computer science and holds expertise in computational intelligence, machine vision, healthcare AI, sustainable computing, and signal processing, spanning a 27-year career in higher education, research, and educational leadership. At Jazan University, Saudi Arabia, he's known for his pedagogical talent and administrative leadership. His impactful works feature in the prestigious journals of IEEE and the ACM. Notably, he holds international patents and copyrights and showcases ground-breaking healthcare AI innovations.

RUCHI MITTAL is currently a professional with over two decades of work experience in teaching, training, research, and academic administration. Her teaching expertise comprises DBMS, machine learning, social networking analysis, and data mining. She has published more than forty papers in Scopus/SCI-indexed journals, book series, and conference proceedings. She received Chitkara University Research Excellence Award in author with the highest H-index category (Computer Applications Discipline), in 2020. She reviews for a number of leading international journals, such as *Internet Research* (Emerald), *World Review of Science, Technology and Sustainable Development* (Inderscience), *International Journal of Computational Science and Engineering* (Inderscience), and *International Journal of Intelligent Systems Technologies and Applications* (Inderscience), and a number of IEEE and Springer conferences. She serves on the TPC for a number of conferences and is an active Ph.D. Advisor.

VARUN MALIK (Member, IEEE) received the Ph.D. degree in computer science and engineering from I. K. Gujral Punjab Technical University, Punjab, India. He holds more than ten Indian patents/copyrights. His current research interests include blockchain, artificial intelligence, cloud computing, cyber security, the Internet of Things, data mining and warehousing, and method engineering. He is serving as a reviewer for many international journals.

FATHE JERIBI received the M.S. degree in computer science and information technology from Sacred Heart University, CT, USA, and the Ph.D. degree in information technology from Towson University, MD, USA. He is currently an Associate Professor with the College of Computer Science and Information Technology, Jazan University, Saudi Arabia. His areas of research interests include machine learning, software engineering, computer networks, SDN, wireless ad hoc networks, the IoT, and distributed computing.

SHAMS TABREZ SIDDIQUI received the Ph.D. degree in computer science from Aligarh Muslim University, with research expertise in areas, such as wireless communication networks, AI-enabled systems, and blockchain technologies. He is currently an Assistant Professor with the Department of Computer Science, Jazan University, Saudi Arabia. He has contributed extensively to the fields of cybersecurity, the Internet of Things (IoT), and machine learning, with numerous publications in reputable journals. His recent work focuses on AI and data science applications for next-generation communication frameworks and secure computing. He is actively involved in international conferences and has authored several books and book chapters.

MOHAMMAD ALAMGIR HOSSAIN received the Ph.D. degree in computer science and engineering from the National Institute of Technology, Durgapur, with research interests in computer vision, image processing, machine learning, and artificial intelligence. He is currently a Senior Lecturer with the Department of Computer Science, Jazan University, Saudi Arabia. He has over 26 years of academic experience and has published extensively in leading international journals and conferences. His recent work includes advancements in facial emotion recognition, deep learning applications in healthcare, and AI-based predictive models. He has also served as the session chair and a keynote speaker for prestigious international conferences.

S. L. SWAPNA received the master's and M.Phil. degrees in computer science from Manonmaniam Sundaranar University, India, and the Ph.D. degree in computer science from Bharathiar University, Coimbatore, India. She is currently an Academician and a Technical Trainer. She is a Professional IT Trainer with Applexus Technologies, TechnoPark, Thiruvananthapuram. Previously, she was a Faculty Member with the School of Engineering and Computer Science, Jazan University, Saudi Arabia. Her research interests include big data, machine learning, predictive data analytics, and privacy-preserving data security.