AI-DRIVEN SMART AGRICULTURE SYSTEM FOR ENHANCED CROP PRODUCTIVITY

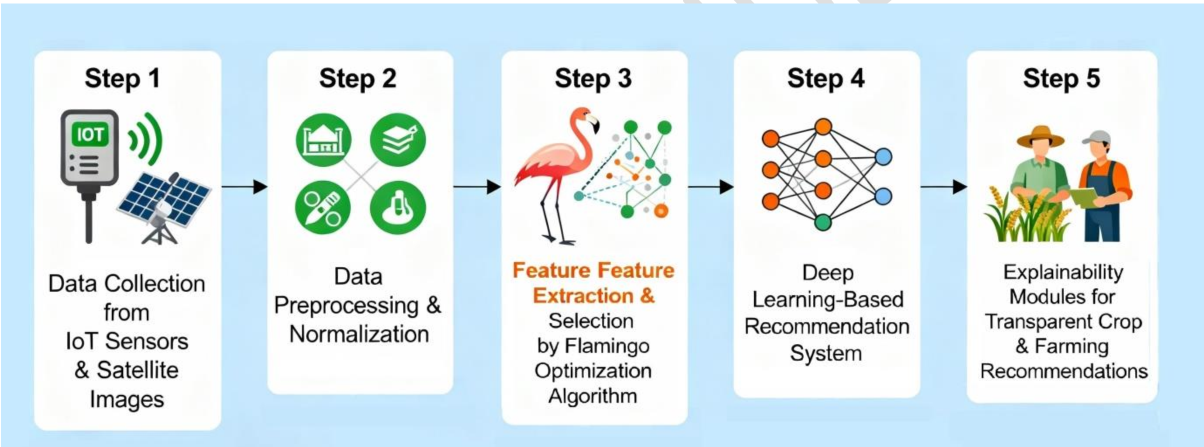
S.Amudha[1*][0000-0002-5859-0352] , Upendra Kumar 2,

[1]Post Doctoral Researcher and Department of Computational Intelligence, Faculty of Engineering and Technology, School of Computing,
[1]LINCOLN University College, Malesia and SRM Institute of Science and Technology, Kattankulathur, Chennai, India.
[2]Computer Science and Engineering Department, Institute of Engineering & Technology, Lucknow.

samudha598@gmail.com, upendrakumar56@gmail.com

Corresponding id: samudha598@gmail.com

Graphical Abstract:



The proposed system uses IoT and satellite data for smart agriculture. It applies the nature-inspired Flamingo Optimization Algorithm for key feature selection, followed by a deep learning model that provides adaptive crop, irrigation, and fertilizer recommendations. Explainable AI modules ensure transparent decision-making to build farmer trust. The system enhances productivity with continuous real-time updates and validated performance in smart farming scenarios.

Abstract:

Smart agriculture leverages technologies such as the Internet of Things (IoT), artificial intelligence (AI), and big data analytics to enhance sustainability, productivity, and resource efficiency. This study proposes a novel Permutation Flamingo Optimized Recommendation

System (PFO-RS) that integrates deep learning with explainable AI principles to improve transparency and decision-making in precision farming. The proposed framework employs the Flamingo Optimization Algorithm for permutation-based feature selection across complex agricultural datasets comprising soil nutrients, crop type, and climatic variables. These optimized features are then used by a recurrent neural network with long short-term memory (RNN-LSTM) to generate adaptive recommendations on optimal crop selection, irrigation scheduling, and fertilizer application. Experimental evaluation using multi-year agricultural datasets from South India demonstrated that PFO-RS achieved superior performance compared with baseline models such as SVM, Decision Tree, XGBoost, and CNN, with an average accuracy improvement of 2.5–3.0% across five major crops and an R^2 value of 0.98 for yield prediction. Mean absolute error (MAE) and root mean square error (RMSE) were reduced to 0.092 and 0.089, respectively, compared to 0.108 and 0.126 for the best-performing conventional models. Field-level validation indicated a 5–8% improvement in predicted productivity when applying the model's recommendations in simulated real-world farming conditions. These results confirm that the proposed PFO-RS framework provides accurate, explainable, and scalable support for data-driven agricultural decision-making.

Keywords: Smart agriculture, IoT, AI, explainable artificial intelligence (XAI), decision-making

1. Introduction:

Humanity's existence and fortune have always been largely dependent on agriculture. Its development has been closely tied to changes in the industrial paradigm. Each industrial revolution has left its mark on agriculture-from the very first steam and mechanization of Industry 1.0 to the present-day phenomenon of the Industry 4.0, where automation meets data-driven decision making in agriculture [1]. Compared to all historical revolutions, which have had a strong influence on the advancement of agriculture, the precision farming revolution in Agriculture 4.0 has brought more advanced electronic technology and significant improvements in IT for such an integrated approach as opposed to the Agriculture 3.0 phase. However, it ought to be clear that the particular challenges facing agriculture prevent the sector from speedy adoption of these emerging technologies, such as poor IT infrastructure and lack of experience. Important among space technologies is improving soil quality and decreasing water wastage through efficient irrigation systems. It facilitates sharing agricultural knowledge among farmers themselves. Sustainable implementation of smart agriculture [2] systems would necessitate science and innovations, including space technology, to increase the quantity and quality of agricultural outputs. It has to do with a combination of collecting, analyzing, and utilizing geospatial data from different sources such as satellites, surveillance and terrestrial, aquatic, and aerial sensors. Precision farming is used in modern intelligent agriculture because it ensures that the inputs will be in the right amounts and at the right times to maximize production and reduce environmental impact. AI-embedded predictive analytics makes it easier to predict the market trends and diseases in crops, allowing better risk management and planning. The primary role of Explainable Artificial Intelligence (XAI) in smart agriculture is to improve transparency, trust, actionable knowledge in AI-based farming methods. AI models are often described as "black boxes," thereby making it difficult for farmers as well as agronomists alike to understand the rationale behind the results produced by such systems, although these systems may deliver accurate predictions in tasks such as crop disease diagnosis, yield prediction, or soil analysis [3]. In such scenarios, XAI offers a short account of its reasoning behind calling a certain decision, enabling more interested parties to go through the rationale behind the recommendations and even confirm it. For example, a smart irrigation system may explain why a certain amount of water is prescribed for a given field region based on the crop type, temperature, and soil moisture. Interpretability is for increasing confidence in system adoption by users, and it also has a role in enabling domain experts to uncover possible biases or mistakes that the models may contain. This smart farming has revolutionized agricultural operations through the use of ML and AI for optimizing crop yield and management. Combining these has given way to development of advanced CR systems that intelligently recommend crop cultivation based on various factors such as soil properties, prevailing weather conditions, previous yield records [4]. These systems target cost and environmental efficiencies while improving productivity and profitability in agriculture. The fact that the ML models are opaque is a huge barrier to the practical adoption of AI-based CR systems in agriculture despite the fact that they obviously have some benefits. Most of such models work as "black boxes," producing predictions without justification. The lack of openness may render farmers, who generally make judgement based on actual information and experience, skeptical and hesitant. Lack of comprehension of the reasoning behind AI suggestions erodes confidence and prevents smart agricultural solutions across the board [5]. In addition to trust issues, model transparency also entails legal requirements such as that in General Data Protection Regulation (GDPR), which has been widely, interpreted as supporting a ‘right to explanation’ in automated decision-making systems [21]. The present rule underlines the fact that effective and reasonable explanations must be established with regard to the automated decision impacting a particular person. This also applies to having smart farm CR systems elucidating the reasoning by which such suggestions have been offered. Explainable AI (XAI) is probably the solution to these issues. The purpose of XAI is to make it possible for humans to understand the modes of decision-making in the AI systems,"says a source on AI research. The CR system thus may reveal factors that determined the recommendations that would be made to the particular farm. To do this, use explainability methods such as SHapley Additive ExPlanations (SHAP) [22], Differentiable Counterfactual Explanations (DICE_ML) [23], and Local Interpretable Model-Agnostic Explanations (LIME) [24]. Although these interpretations can be global in nature, indicating the model's overall behavioral tendencies, or local, supporting specific predictions. The aim of theoretical foundations in XAI is to bridge the gap to this increasing use in many fields with the call for systems, AI systems that should be trustworthy, equitable, and accountable. In smart farming, XAI acts as an interface between farmers and AI systems, providing valuable and actionable information for farmers to make decisions. This conversation aims at creating a collaborative milieu between human expertise and AI capabilities by suggesting answers and letting farmers articulate the right questions [7, 8].

1.1 Contribution for this research:

In order to methodically create and execute the Explainable Recommendation System for Smart Agriculture utilising the Permutation Flamingo Optimised Recommendation System (PFO-RS), the proposed work is divided into multiple clearly defined parts. IoT sensors, satellite imaging, agricultural databases are utilized to collect a variety of agricultural data in the first phase, including soil characteristics, crop varieties, insect patterns, climate, and past yields. To clean data, deal with missing values, bring it to a consistent scale appropriate for analysis, the second phase involves data preparation and normalisation. The Flamingo Optimisation Algorithm, which cleverly finds and permutes the most pertinent factors influencing food productivity, is used in the third phase to concentrate on feature extraction and selection. In order to produce accurate agricultural suggestions, the fourth process involves building and training a deep learning-based recommendation system using the optimised feature set. Explainability modules, which are incorporated into the fifth phase, allow the system to transparently explain the rationale behind particular recommendations, hence increasing user adoption and trust. The usefulness of the system in actual smart farming settings is evaluated in the sixth step using a performance evaluation and validation procedure that uses measures including accuracy, precision, recall, and productivity improvement rate.

1.2 Research Novelty and Contributions

Although previous studies have utilized optimization algorithms such as Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) in conjunction with deep learning for agricultural recommendations, these models typically perform feature selection and model training as independent steps and often lack interpretability. The proposed Permutation Flamingo Optimized Recommendation System (PFO-RS) introduces a novel permutation-driven variant of the Flamingo Optimization Algorithm, which dynamically reorders and evaluates candidate feature subsets to improve convergence speed and stability in non-linear agricultural data. Unlike traditional hybrid models, PFO-RS embeds this optimization directly within the training loop of an RNN-LSTM network, enabling adaptive learning from temporal and spatial data streams. Furthermore, the framework integrates Explainable AI (XAI) components (using SHAP and LIME) within the recommendation process, allowing users to visualize how each feature such as soil pH, rainfall, or nitrogen level influences the final recommendation. This unified architecture achieves both performance improvement and interpretability, addressing the two main limitations of prior smart agriculture recommendation systems.

1.3 Theoretical Differentiation from Existing Hybrid Models

Although hybrid metaheuristic-deep learning models are widespread in the field of agricultural recommendation systems, the majority of the previous solutions use feature selection as a pre-processing step that is constantly applied prior to training the model. Conversely, the suggested Permutation Flamingo Optimized Recommendation System (PFO-RS) presents three differentiation features:

Permutation-based subset encoding: The proposed FOA form encodes candidate solutions as permutations of features. This enables measuring an order of feature interaction and sensitivity to time which are especially important to sequential LSTM models.

Fitness coupled with temporal generalization error: The single objective of the optimization is no longer validation-based temporal loss, calculated as the RNN-LSTM sequence model.

This incorporates optimization instead of tabular evaluation which is often fixed time-series modeling.

Explainability-conscious feature stability constraint: when using to optimize, the components of features are not necessarily weighted just by the accuracy of prediction, but also deemed in terms of SHAP feature stability during validation folds, thereby promoting interpretable and sound feature choice.

Thus, the suggested framework is compared to the traditional GA/PSO-based hybrids because their logic and the means of contributing permutation logic, temporal fitness assessment, and interpretability restrictions to a single optimization loop are integrated.

2. Literature survey:

Current agricultural research initiatives regarding crop forecasting include those by Work [9], who developed a model for crop prediction using Gaussian Naive Bayes, Decision Tree, Logistic Regression, Random Forest, XGBoost as baselines. Authors collected agricultural data from multiple Indian regions encompassing soil parameters (N, P, K content, pH), climatic factors (temperature, rainfall, humidity), and crop yield statistics. The dataset was divided into training and testing subsets using an 80:20 split ratio. Input features were normalized to a uniform scale before being introduced into the model. Each feature vector represented the soil and environmental attributes for a specific region and cropping season. The model architecture comprised multiple machine learning classifiers: Gaussian Naive Bayes, Decision Tree, Logistic Regression, Random Forest, and XGBoost arranged in a parallel training pipeline. Each classifier received the same preprocessed feature inputs and generated independent crop yield or suitability predictions. Such researchers ended up developing an Android app that can use both offline and online platforms for the forecasting based on the created models. By similar standards, author [10] created AgroConsultant, an intelligent recommendations system that guides Indian farmers in choosing a crop to grow depending on various geo and environmental parameters. It utilized a soil dataset collected from various sources to assert the proposed approach [11]. Using the ensemble model of the system, the output was rule-based. In the same vein, author [12] tried adopting ensemble techniques to construct a highly efficient and logically accurate crop recommendation system as well. In Random Forest, CHAID, K-Nearest Neighbour, Naive Bayes learners, they used the technique of majority voting. The dataset for this research study was crop data sourced from the internet and soil testing laboratory data. Hence, such a rule was demonstrated by the application of the most relevant model. A crop recommendation module was developed by Work [13] based on the majority voting scheme by combining the outputs of SVM, RF, NB, KNN. It also developed a TILLAGE web-based tool based on a very catchy module that provides recommendations on pesticides and fertilizers. The modular approach is suggested to elevate agricultural productivity by improving crop selection, fertilisation and soil use. Study [14] showed how agriculture has grown in high dimensions of rich countries based on crop and soil factors. for supporting small and marginal farmers in the state of Kerala, India. Crop calendars and soil-crop databases were prepared from observations in the various cultivated fields. This was done by cheapening the cost of agriculture using various electrical tools. In the study, data on soil-weather parameters, i.e., temperature, humidity, NPK (nitrogen, phosphorus, and potassium) values, and Potential of Hydrogen (pH value), were gathered over time through in-situ observations. The authors found, after collecting and analyzing data through a wireless sensor network communicating with a remote server, that the MAC and routing algorithm method showed promising results. A networking and sensor model for environmental condition data gathering on fields of white cabbage crops was proposed in work [16]. Diesel non-reduction and monitoring of environmental requirements on the growing farm. A very short essay on precision agriculture - reduce labour costs and solve food problems-author [17]. The Active Precision Farming Process was authored by the authors who considered the different steps of Active Precision Farming to address daily problems of climate change and pollution caused by work [18] in which sensor was proposed to collect data on the leaf temperature and its water requirement. By knowing this, it sets and uses the crop requirements and period-specific adaptability to allow improved farming. The use of EM4325 ultra-high-frequency (UHF) chips and sensors based on Radio-frequency identification (RFID) were parts of the whole theme tackled by the writers. The author [19] further can supply alternative strategies that could be proposed for the improvement of crop productivity. Variety of Technologies Used for Agricultural Production: These are the following: Connectionless Mode Service (CLS), Packet Switching (PS), and Physical Layer Signalling (PLS). PLS precision sowing has been said to improve crop yield more effectively. Work [20] discusses the way Internet of Things (IoT) enables inter-connection between devices, such as smartphones, tablets, PCs, through machine-to-machine (M2M) communication. They discuss the applicability of IoT towards precision agriculture through best utilization of resources such as water, light, pesticides, thereby improving production efficiency and decreasing waste.

Recent studies have highlighted the growing integration of sustainable chemical and biological processes in modern agriculture, complementing AI-driven decision frameworks such as the proposed PFO-RS system. For instance, [25] demonstrated an eco-friendly oxidative conversion approach using hydrogen peroxide for transforming plant precursors into value-added agrochemicals, emphasizing green chemistry principles aligned with precision agriculture goals. In parallel, [26] investigated the application of actinomycetes in removing chromium contaminants from agricultural soils, underscoring the importance of bioremediation for maintaining soil health and crop productivity. Similarly, [27] explored biofortification strategies under drought stress, revealing how biochar amendments enhance iron uptake in plants findings that could serve as key environmental features for intelligent crop recommendation systems.

Further, [28] examined aquatic plants as bioindicators of heavy-metal accumulation in surface-water ecosystems, providing a reference framework for integrating water-quality indices into AI-based agricultural decision models. Addressing pest-management challenges, [29] analyzed pesticide resistance patterns and their implications for sustainable agricultural planning, highlighting the potential role of predictive analytics in managing pest adaptation dynamics. Finally, [30] presented an electrochemical oxidation process for converting ammonia to nitrogen gas in agricultural wastewater treatment, showcasing scalable environmental technologies that can be coupled with data-driven models for resource optimization. Performance of our proposed method along with its superiority over other methods in various studies has been discussed in Table 1.

Table 1. Summary of related literature

Ref	**Study Focus**	**Methodology**
Akkem et al. [1]	Streamlit-based enhancement of crop recommendation systems using XAI	Neural networks with SHAP, LIME; interactive interface
Shams et al. [2]	Explainable AI-based CRS for agricultural decision-making	Ensemble ML + LIME explanations
Naga Srinivasu et al. [3]	XAI-driven crop recommender for precision agriculture	Model-agnostic XAI integrated with ML
Kumar & Kumar [4]	Explainable AI model for decision-making in crop recommendation	Hybrid deep learning model
Turgut et al. [5]	AgroXAI: Explainable crop recommendation for Agriculture 4.0	SHAP + decision-tree ensemble

Bouni et al. [6]	Interpretable ML techniques for advanced crop recommendation	Tree-based ML + feature analysis
Chen et al. [7]	Integration of XAI and blockchain for secure smart agriculture	Blockchain + interpretable ML
Cartolano et al. [8]	Evaluation of XAI models for smart agriculture environments	Model comparison and assessment
Martin et al. [9]	XAI-powered smart agriculture framework	Multi-layered XAI framework for food productivity
Bhola & Kumar [10]	Farm-level smart crop recommendation using ML	Ensemble models (RF, GB, KNN)
Grati et al. [11]	Ontology-based explainability in smart agriculture	Ontological modeling + XAI
Chakraborty & Mishra [12]	Smart farming recommendation with collaborative ML and image data	ML + image-based feature extraction
Kisten et al. [13]	Explainable AI for predictive maintenance in agri facilities	Deep learning + SHAP
Kancharagunta et al. [14]	ML and IoT-based CRS survey	Review and taxonomy
Raghavendra & Annapurna [15]	Systematic literature review on crop recommendation ML	Survey of supervised ML algorithms
Mewar et al. [16]	Web-based crop recommendation for farmers	ML web app using Flask
Bala Kamatchi & Muthukumaravel [17]	Dynamic CRS using reinforcement learning	RL + real-time sensor feedback
Corchado et al. [18]	Bio-inspired recommendation for urban orchards	Evolutionary algorithms
Harikumaran & Vijayalakshmi [19]	Defuzzification for adaptive fertilizer optimization	Fuzzy inference + metaheuristics
Madhumathi et al. [20]	Agricultural tool recommendation using ML	Random forest + classification
Wachter et al. [21]	GDPR and explainability in automated decision systems	Legal and ethical analysis
Lundberg & Lee [22]	SHAP framework	Unified model-agnostic explanation method
Ribeiro et al. [23]	LIME framework	Local interpretable model-agnostic explanations
Mothilal et al. [24]	Counterfactual explanations (DICE)	Diverse counterfactual explanation generation
Nagabhooshanam et al. [25]	Green synthesis of agrochemicals via H_2O_2 oxidation	Sustainable oxidation chemistry
Murali et al. [26]	Bioremediation of chromium-contaminated soils	Actinomycetes-based remediation
Ismail et al. [27]	Iron biofortification under drought via biochar	Biochar-based nutrient enhancement
Reddy et al. [28]	Aquatic plants as bioindicators for heavy metals	Water-quality biomonitoring
Satyanarayana et al. [29]	Pesticide resistance analysis in	Statistical and simulation-

	pest management	based study
Srilakshmi et al. [30]	Electrochemical ammonia-to-nitrogen oxidation for wastewater treatment	Electrochemical sustainability process

3. Proposed Method

The proposed model comprises of an Explainable Recommendation System for Smart Agriculture Farming aimed at improving food productivity through a novel Permutation Flamingo Optimized Recommendation System (PFO-RS). The system integrates XAI principles to ensure transparency as well as trust in decision-making processes for farmers and agricultural stakeholders. At its core, the PFO-RS utilizes the Flamingo Optimization Algorithm, a nature-inspired metaheuristic technique, to intelligently perform feature extraction and permutation-based selection from complex agricultural datasets, including soil quality, crop type, weather conditions, historical yield data. This permutation mechanism ensures that only the most relevant and high-impact features are considered for generating recommendations. Once the optimal feature subset is derived, it is fed into a deep learning model that constructs precise and adaptive farming recommendations—such as optimal crop selection, irrigation schedules, and fertilizer application—tailored to specific farm environments. The deep learning model learns from historical patterns and real-time inputs, continuously updating its recommendations for enhanced accuracy. The explainability component of the system provides interpretable justifications for every recommendation, empowering farmers to make informed decisions with confidence.

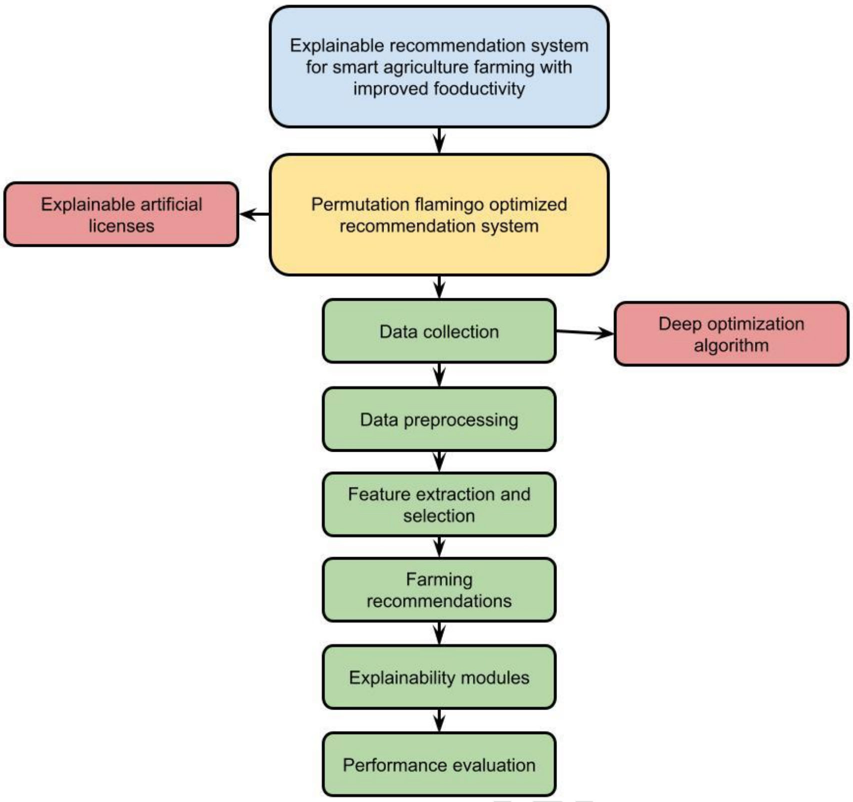


Figure 1: Flow chart of the Proposed PFO-RS

The proposed work is structured into several well-defined phases to systematically develop and implement the Explainable Recommendation System for Smart Agriculture using the Permutation Flamingo Optimized Recommendation System (PFO-RS) as shown in figure 1. The first phase involves data collection, where diverse agricultural data such as soil properties, climate conditions, crop types, pest patterns, historical yields are gathered from IoT sensors, satellite imagery, and agricultural databases. In the second phase, data preprocessing and normalization are performed to clean the data, handle missing values, and bring it to a uniform scale suitable for analysis. The third phase focuses on feature extraction and selection using the Flamingo Optimization Algorithm, which intelligently identifies and permutes the most relevant features affecting food productivity. This is followed by the fourth phase, where the deep learning-based recommendation system is constructed and trained on the optimized feature set to generate precise farming recommendations. The fifth phase incorporates explainability modules, enabling the system to provide transparent insights into why specific recommendations are made, enhancing trust and adoption among users. Finally, in the sixth phase, a performance evaluation and validation process is conducted utilizing metrics such as accuracy, precision, recall, productivity improvement rate to assess effectiveness of system in real-world smart farming scenarios.

3.1.1 Phase-1: Dataset collection:

The training and validation phases of the proposed framework utilized historical datasets spanning 2001–2015, sourced from agricultural departments and meteorological agencies across South India. This period was selected because it provides a comprehensive and diverse record of crop yields, soil parameters (N, P, K, pH), and weather variables (temperature, rainfall, humidity) across multiple districts and seasons. Historical datasets of this length are valuable for modeling long-term agro-climatic variability, allowing the system to learn from periodic patterns such as drought years and monsoon fluctuations. The IoT sensor layer of the proposed architecture is designed for real-time operational deployment, not for initial model training. Once deployed in the field, soil moisture, temperature, and pH sensors, along with satellite-based remote sensing inputs, continuously feed current data into the trained system for adaptive recommendations. Thus, while historical data enabled model generalization and benchmarking, IoT data ensure ongoing adaptability and context-aware decision-making in real-world applications. Future work will incorporate recent agricultural and IoT datasets (2016–2023) to update and retrain the system for even greater predictive accuracy and relevance to present-day farming conditions.

The fact that the training dataset (2001 -2015) covers various high-variability and drought years of monsoons allows the model to be trained to accumulate varied agro-climatic dynamics. Nevertheless, the recent trends in climate change such as the rise in the variability of temperatures and changing rain patterns can affect the current agricultural conditions. As far as the main soil-crop interactions do not change, the prediction calibration may be influenced by time-dependent climate change. To work on this, the suggested framework facilitates ad hoc retraining, based on a revised dataset (after 2016), and real time incorporation of IoT meters. This is an adaptive retraining approach that can guarantee long term relevance to the model when subjected to changing climate conditions.

STATISTICAL CHARACTERISTICS OF DATASET: 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 2 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 values show a gradual increase, 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 2. 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 (°C)	27.152	28.365	28.565	28.758	28.999
Humidity (%)	12.345	12.189	12.457	12.680	12.985
Wind Speed (m/s)	3.158	1.247	3.654	3.989	4.125
Maximum temperature (°C)	47.125	47.352	47.555	47.689	47.985
Minimum temperature (°C)	8.125	8.457	8.325	8.845	8.645
Cloud amount	56.145	48.528	47.135	55.900	53.689
Precipitation (mm)	3.225	3.105	3.454	3.565	3.689
UVA Irradiance (W/m²)	14.588	14.254	14.168	14.609	14.578
UVB Irradiance (W/m²)	0.458	0.425	0.498	0.512	0.524

Downward Irradiance (W/m²)	6.145	6.245	6.347	6.458	6.788
PAR Total	100.125	98.562	99.347	99.858	98.858

*All meteorological values are reported in constant numeric expression in three-digit precision. Values of cloud amounts are formatted in percentage equivalents without any unit symbols.

Table 3. 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 between. ◦ 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/m2 in 2001 to 14.578 W/m2 in 2005, indicating a relatively stable level of ultraviolet A radiation. ◦ UVB irradiance: UVB irradiance levels increase marginally from 0.458 W/m2 in 2001 to 0.524 W/m2 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/m2 in 2001 to 6.788 W/m2 in 2005. ◦ PAR total: Photo synthetically active radiation (PAR) exhibits some fluctuation, with values ranging from 100.125 W/m2 in 2001 to 98.858 W/m^2 in 2005.

Table 3 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.

3.1.2 IoT based sensors in field area for crop prediction

Soil moisture sensors, soil pH level sensors, temperature sensors, humidity sensors are deployed to measure some of the other specifications in that area such as rain, pH value of the soil, and soil moisture level. These measurements are used to inform planting decisions on what crops should be planted. In order to analyze infected crop by utilizing mobile device's camera, pictures are taken and uploaded into the cloud to identify the ideal fertilizer recommended for it. The information stored in the cloud will thus be subjected to guidelines for data processing. Then, the dataset will be divided into training dataset and testing dataset. Data belonging to training dataset should be used for the crop recommendation training model to apply the deep learning algorithm. By providing pictures of afflicted croplands, it pulls information from database for crop disease segmentation in event of crop diseases. Algorithms for image processing will then use the features that were extracted to forecast crop disease.

3.1.3 Phase-2: data Pre-processing:

Pre-processing data can enhance the quality of that data before feeding it into the model. It provides a coherent structure to the data model by addressing the deficiency of data, detecting redundancy in data, and disqualifying low quality data. This process has two phases. Firstly, it deletes the missing values (represented using a dot, '.') from the actual data. It transforms raw collected data into machine-learnable structured datasets. This method does not require any labelling whilst performing data pre-processing for the numeric data in their values. This is done to allow data to be more trainable through normalization. Labels must be created during data pre-processing because the original dataset does not carry any. From yield and area of cultivation (tonnes and hectares), the labels of each crop are generated. The number 1 signifies those for whom the production area value was greater than zero. Class label 0 is set in all other instances. At this stage, do some data management, formatting, and missing value handling. Work with missing values through imputation of feature means; see Algorithm 1. Environmental values show patterns as crop production is a continuous affair in a country.

Thus, mean of the values before and after missing values in our dataset can hold the missing values.

Consider three different kinds of data: the amount of crop production in a certain area, the areas of cultivation, and environmental factors associated with crop production. These facts are gathered by us from various government agencies. Combine all of the data into a single dataset in order to prepare the learnable dataset. Junk, redundant, and unnecessary data negatively impact the ML models' performance.

Algorithm-1: data pre-processing

Input: Raw data (S)
Output: Preprocessed dataset after missing value handling
procedure Missing Value Handling(S)
for every attribute S^a do
m^a = mean $(S^a)[m^a$ is arithmetic mean of attribute S^a]
for every sample data $S_d{}^a$ do
if $S_d{}^a$ is missing then
$S_d{}^a := m^a$
end if
end

3.1.4 Phase-2: feature extraction and selection using Flamingo Optimization Algorithm

The above problems are solved using the flamingo optimization algorithm, which is a new swarm intelligence optimization method inspired by migration as well as feeding behavior of flamingos. Most flamingos are social migratory birds that eat algae, clams, tiny worms, and insect larvae. Another one is the way flamingos intertwine their long necks down and tilt them over for feeding. They have two important behavioral traits: foraging and migration. Most flamingo populations live in areas where they can forage abundantly. After a period of intense foraging, flamingo populations may migrate when food becomes insufficient in a particular area to sustain population. Foremost optimization concepts of FOA are provided as follows.

Flamingos find food sources in their current search area insufficiently abundant. Rather than locating more food within search region than in their known foraging areas, they prompt each flamingo to reset location. It turns upside down the FSA optimization path to find global optimal solution (GOS) inside a certain search region. Search agent will not be able to find the present GOS if it is in active search. In FOAS, flamingos are search agents that explore search space through exchange of information and move according to rules of stable position change that ultimately lead toward best answer. • The flamingo's activity, including its foraging and migrating habits, dictates the rules for changing postures. Additionally, the flamingos' foraging behaviour is divided into two categories: foot movement and beak scanning. These behavioral dynamics influence their adaptive movement and search efficiency.

(i) Scavenging behavior

Sociable behavior: The flamingo that finds the food will first alert the other flamingos, forcing them to adjust their positions to match the food's placement. Mathematically represent the availability of food in kth dimension as ydk if the flamingo seeks the optimal solution, which is where there is an abundance of food.

(ii) Fitness function

The best flamingo is regarded as the fittest solution that offers the finest candidate solution after each flamingo's fitness function is assessed by eqn (1)

$$f(x) = 100(x_1^2 - x_2)^2 + (1 - x_1)^2 \tag{1}$$

where $x=(x_1,x_2)$ denotes the decision variables, and $f(x)$ represents the objective function to be minimized, attaining its global minimum $f(x) = 0$ at $(x_1,x_2)=(1,1)$.

This paper uses the Rosenbrock function as a reference objective function that is used to prove the convergence pattern and exploration exploitation ratio of the Flamingo Optimization Algorithm (FOA). Rosenbrock function is popular in the field of optimization due to its small valley shape and therefore the narrowness that indicates the effectiveness of an algorithm to move out of the local minima and get to the global optimum. In the case of agricultural feature selection, the actual fitness model is determined by classification error (1 - accuracy) of the RNN-LSTM at the selected feature set. This is the Rosenbrock formulation that is used to mathematically explain how FOA can optimize the maximization of the objective function, prior to being applied to the real-life agricultural objective.

(iii) Bill-scanning behavior - In order to find its food, a flamingo submerges its head in the water. When it finds food, it swallows it upside-down, filtering out extra water and waste. The flamingos' scanning radius varies depending on the circumstances, and they bend their heads and search more intently when there is an abundance of food. In the kth dimension, designate ylk as the location of the lth flamingo. Even if there is a chance of small errors happening, random errors can happen during information exchange and are overcome by using a conventional normal distribution. The flamingos' greatest travel distance can be mathematically stated as eqn (2)

$$A_1 = M_1 \times yd_k + \lambda_1 + y_{lk} \tag{2}$$

where A_1 denotes the updated position value of the flamingo agent, M_1 is the scanning range factor, yd_k represents the food source position in the k^{th} dimension, λ_1 is a random coefficient controlling movement intensity, and y_{lk} is the current position of the l^{th} flamingo in that dimension.

The fluctuation in the flamingos' scanning range is represented by M1, which is initially assumed to be a random number with a uniform distribution and to be carried out at its greatest distance by eqn (3)

$$A_2 = M_2 \times |M_1 \times yd_k + \lambda_1 + y_{lk}| \tag{3}$$

where A_2 represents the adjusted movement amplitude of the flamingo agent. λ_1 and λ_2 are random values in the interval $[-1,1]$, while M_2 is a random value that has a typical uniform distribution.

(iv) Claw locomotive behavior: The claw-based locomotion mechanism guides the search agents toward regions of higher fitness concentration. The search region is expanded because the flamingos cover λ1 × ydk, which is the distance from the spot where there is a lot of food (ydk). The flow of the food-finding processes in the nth iteration is provided by eqn (4)

$$s_{lk}^{n} = \lambda_1 \times ys_k^n + M_2 \times |M_1 \times yd_k + \lambda_1 + y_{lk}| \tag{4}$$

where s_{lk}^{n} denotes the updated position of the l^{th} flamingo in the k^{th} dimension at iteration n; ys_{k}^{n} represents the current search position.

The flamingo's location is updated with different locations and can be stated mathematically as eqn (5)

$$y_{lk}^{n+1} = \frac{(y_{lk}^{n} + \lambda_1 \times ys_{k}^{n} + M_2 \times |M_1 \times yd_k + \lambda_1 + y_{lk}|)}{R} \tag{5}$$

(v) Emigrating behaviour: The fitness value of the flamingos is determined by their ability to emigrate, and the emigrating flamingo is regarded as the fittest. When there is less food available, the flamingos move to a different location to find food where there is more. The following is a mathematical representation of the flamingo's emigration behaviour by eqn (6)

$$y_{lk}^{n+1} = y_{lk}^{n} + \sigma(ys_{k}^{n} - y_{lk}^{n}) \tag{6}$$

(vi) Termination The procedure is stopped after the ideal solution has been found and the maximum number of iterations has been reached.

Algorithm-2: Flamingo Optimization Algorithm

Start

Input: yd_k

Output: y_{n}^{n+1}

While ($n < n_{\max}$) #$n_{\max}$ is maximal iterations

Scavenging behavior

If

Start: yd_k

For Bill scanning behavior do

Maximum distance

$$A_1 = M_1 \times yd_k + \lambda_1 + y_g$$

Varying scanning range

$$A_2 = M_2 \times |M_1 \times yd_k + \lambda_1 + y_{kk}|$$

Claw locomotive behavior

$s_{lk}^{n} = \lambda_1 \times ys_{k}^{n} + M_2 \times |M_1 \times yd_k + \lambda_1 + y_{kk}|$

Position update

$$y_{lk}^{n+1} = \frac{(y_k' + \lambda_1 \times yy_k' + M_2 \times |M_1 \times yd_k + \lambda_1 + y_{uk}|)}{R}$$

Emigrating behavior

$$y_{lk}^{n+1} = y_{ik}^{n} + \sigma(ys_k^n - y_{ik}^n)$$

Check the stopping condition

End

3.1.5 Phase-4: deep learning-based recommendation system

Therefore, using specified climate and crop parameters, crop advice is given to the farmers. This sequential dependence is managed by RNNs, which are an enhancement of neural networks. Hidden state of RNN stores history in that it takes output from previous step as input to current step. In a conventional neural network, each data point is treated as a separate entity altogether. The model being proposed in the present study is a feedback-integrated RNN based on long-short-term memory. Each node of LSTM is characterized by three distinct gates: input (I), output (O), forget (F). G presents the updated input at current time-stamp T. The real value of each of these gate banks on the current input XT and the previous hidden state HI−1 is shown as follows (7).

$$I_T = F(W_I X_T + U_I H_{T-1} + B_F)$$

$$F_T = F(W_F X_T + U_F H_{T-1} + B_F)$$

$$G_T = \tanh\,(W_C X_T + U_C H_{T-1} + B_C) \tag{7}$$

The node's updated value is calculated as follows (8):

$$C_T = I_T \cdot G_T + F_T \cdot C_{T-1} \tag{8}$$

The node state, previous output, and current node input are used to determine the gate value by eqn (9)

$$O_T = F(W_O X_T + U_O \cdot H_{T-1} + V_o \cdot C_T + B_O)$$

$$H_T = O_T \cdot \tanh(C_T) \tag{9}$$

Three RNN models—one for the region's minimum temperature, one for its maximum temperature, and one for rainfall prediction—were trained in this study. The vector used to organise the data in the set is {X(1)…, X(K}3). For both seasonal and 90-day previous forecasting, the dataset is transferred to a N X M matrix with 90 target values expressed as follows and one input feature per row by eqn (10)

$$\{X(T), X(T+1), X(T+2), \dots X(T+90)\} \tag{10}$$

K represents size of time series data. The LSTM has the property of making a prediction for the weather, hence k-90 means that two dimensions N and M are 91. Model is divided into three layers: input, hidden, output. Hidden layer has four LSTM nodes. Min. The rainfall compared with temperature predictions has a similarity. The input layer receives a variety of data features from data. After this, processed data goes to middle layer by input layer. It contains many hidden layers within it. Every hidden layer has a unique weight, bias, and activation function. LSTM-RNN is used since weather conditions depend on the past.

Data Attributes

Agronomic researchers have known all this while, how various geographical and climatic conditions, such as soil type, altitude, rainfall, humidity, and temperature affect production, health, and growth of crops. The fitness of the crop in any geographical area will be determined by many factors in if individually or in combination.

Essential nutrients N (nitrogen), P (phosphorus), K (potassium), quality of the soil as reflected by pH are the parameters taken into consideration in the upcoming learning model. These facts are absolutely significant discretionary factors for crop recommendation and also markers across plenty of datasets. Hence, our model prediction gives the primary type of crop. Although it would be desirable to include all significant geographical and environmental factors for a comprehensive suggestion, the suggested study is mainly concerned with quality, nutrients, health of soil. The explanations are as follows:

- Consistency and Availability of Data: Our data comes from a variety of sources, most of which are concerned with crops, pH, N, P, and K. Variables like temperature and rainfall were present in some datasets, but in order to maintain consistency and avoid the introduction of data biases, uniform characteristics across all data points were required.

- Difficulty of Data Collection for Other elements: Compiling an extensive dataset that encompasses all climatic and geographic elements is an enormous undertaking.

Interrelation with pH: A few environmental factors can be substituted by the soil's pH, which shows how acidic or alkaline it is. For instance, frequent rainfall may have an impact on the soil's pH, which in turn may reflect humidity and temperature. By including pH, the model indirectly takes into consideration some of the effects of outside influences on soil.

Model Training

At same time, the study takes into consideration a validation set with 10 % contribution from the training dataset to indicate overfitting or underfitting during training. The training data comprises 7308 data samples, four predictors (N, P, K, and pH), single response variable, namely major_crop. Since interaction between crop varieties and soil parameters is complex and nonlinear, study employed a NN, which could be described as an advanced machine learning model for analyzing difficult patterns and dependencies in data. Input selections for model should be stored in the input layer for best performance. Topologies with different numbers of hidden layers and nodes within every layer were explored for maximization of productivity. Based on empirical investigation and its results study found that two hidden layers with 64 and 32 neurons were taken into account. The nodes in the output layer represent the number of distinct crop types in dataset. It uses sparse categorical cross-entropy loss function. Fit() method accepts training data as well as informs model training. Out of the training data, a training subset and a validation subset are drawn. Validation takes place while training on 20% of the training data. Two hundred epochs train the model, with the input data divided into batches. Model uses gradients from a batch of size 64 to update its weights and biases in every iteration. Model uses Adam optimizer for updating and backpropagation for the algorithm applied in updating the weights and bias. In order to train the neural network efficiently for crop prediction with those factors, latent features and intricate patterns found in data points from training dataset are in charge. Algorithm 3 talks about the neural network's training procedure.

Algorithm-3: Deep learning model based crop recommendation

Input: x_{train} Training data $(\mathrm{N}, \mathrm{P}, \mathrm{K}, \mathrm{pH}) y_{\text{train}}$ labels (crop),

Number of predictors η, Number of unique crop classes κ,

Learning rate α, Number of epochs e, Batch size s, Adam

hyperparameters $\beta_1, \beta_2, \varepsilon$

Output: Trained NN method

Procedure:

1. Start:

Describe input layer with η

2. First hidden layer 64 with weights $W_1 \in R^{64xn}$ and biases $b_1 \in\in R^{64}$

4. Training: For epoch $= 1$ to e :

5. Shuffle training data.

6. Divide x_{train} and y_{train} into batches of size s

For every batch:

Evaluate: $Z_1 = W_1 X + b_1$ and $A_1 = \mathrm{ReLU}(Z_1)$

Compute: $Z_2 = W_1 X + b_1$ and $A_2 = \mathrm{ReLU}(Z_2)$

Compute: $Z_o = W_o A_2 + b_o$ and $A_o = \mathrm{softmax}(Z_o)$

Evaluate Loss L

Evaluate gradient of loss w.r.t. final output $\frac{\partial L}{\partial A_0}$

Evaluate: $\frac{\partial L}{\partial z_o} = A_o - y_{\text{train}}$

$$\frac{\partial L}{\partial W_0} = \frac{\partial L}{\partial z_0} A_2^T$$
$$\frac{\partial L}{\partial b_0} = \frac{\partial L}{\partial z_0}$$

Utilizing chain rule and considering ReLU derivatives:

$$\frac{\partial L}{\partial A_2} = W_2^T \frac{\partial L}{\partial z_0}$$
$$\frac{\partial L}{\partial z_2} = \frac{\partial L}{\partial A_2} g^{(}(z_2) //$$ where g' is ReLU derivative

End

Explainable AI

Crop categorisation systems are validated through use of XAI. It enables us to pinpoint the crucial elements in crop data that affect choices. By lessening the impact of noisy features, it verifies the accuracy of predictions. As a result, the model forecast will be impartial and equitable across various demographic groupings. All things considered, XAI validates the classification techniques by revealing features, fostering transparency, and encouraging error analysis.

4. Results and discussion:

Experimental setup:

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. Based on performance, compared several FSO-RNN_LSTM models and chose the best one. For this research, used a personal PC for execution. Table 4 lists all of the computer's specifications. The system was implemented in Python 3.10 using TensorFlow 2.15, Keras, and scikit-learn. The RNN-LSTM module consisted of two LSTM layers (64 and 32 units), each followed by dropout = 0.3, and a dense layer of 30 neurons with ReLU activation, ending with a softmax output for multi-crop classification. Optimization employed Adam with a learning rate of 0.001, batch size = 32, epochs = 150, and early stopping (patience = 10). For Flamingo Optimization, the population size was 30, maximum iterations = 50, exploration weight = 0.7, and exploitation weight = 0.3. All random seeds were fixed to 42 for reproducibility. All input features (soil nutrients N-P-K, pH, temperature, rainfall, humidity) were normalized to [0, 1]. Missing entries (< 1 %) were imputed using feature-wise medians. The dataset was split 80 %–10 %–10 % into training, validation, and testing subsets, stratified by crop type.

Table 4 System Configuration

Component	Specification
GPU	2x Nvidia Quadro p4000(8GB)
Storage	4TB
RAM	128GB
Processor	16x Intel(R) Xeon(R) Bronze 3106
OS	Windows Server 2016(64 bit)

Table 5 shows the error ratings for the suggested and studied methods of estimating rice production based on R2 score and various metrics.

Model	MAE	MSE	RMSE	R^2
SVM	0.108	0.017	0.126	0.970
Decision tree	0.103	0.018	0.135	0.950
XGBoost	0.094	0.011	0.111	0.980
GB	0.092	0.017	0.155	0.970
CNN	0.125	0.036	0.171	0.960
FSO-RS-RNN_LSTM	0.092	0.008	0.089	0.980

Table 5 shows that the suggested FSO-RNN_LSTM model has a better performance than all the base models regarding the MAE, MSE, RMSE, and the R 2 score. The environmental uncertainties could be drastic rainfall or flood or long-term drought, which could interfere with the crop production cycle and diminish the quality of the harvest greatly. The model proposed is resilient to this variability as it acquires time dependence in climatic patterns. Such flexibility allows making more consistent seasonal recommendations to crops in response to changes in environmental conditions. Maximum temperature, minimum temperature, and rainfall are the three parameters used for the right seasonal crop prediction. Corresponding results for predicting these parameters are given in Figure 2.

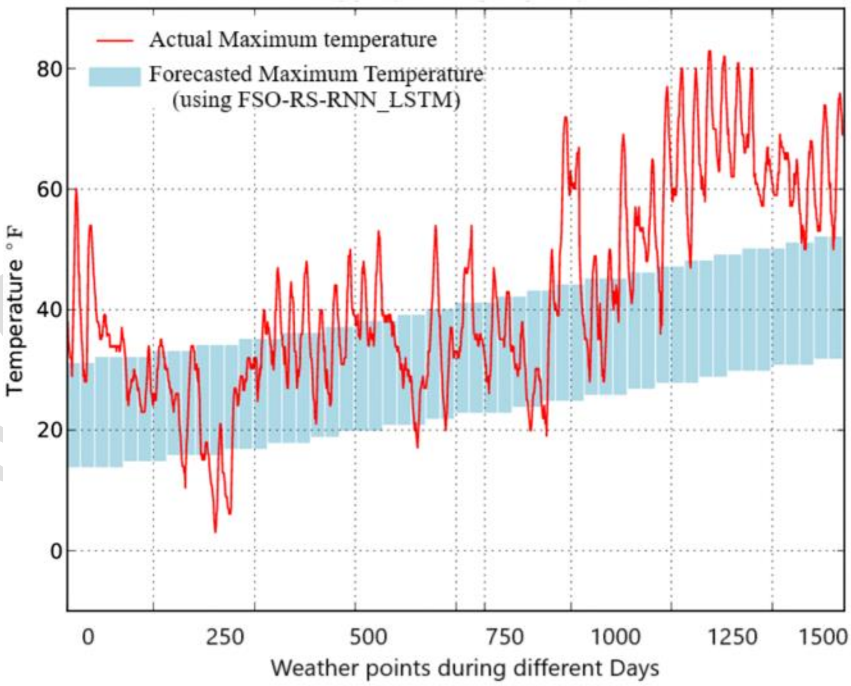
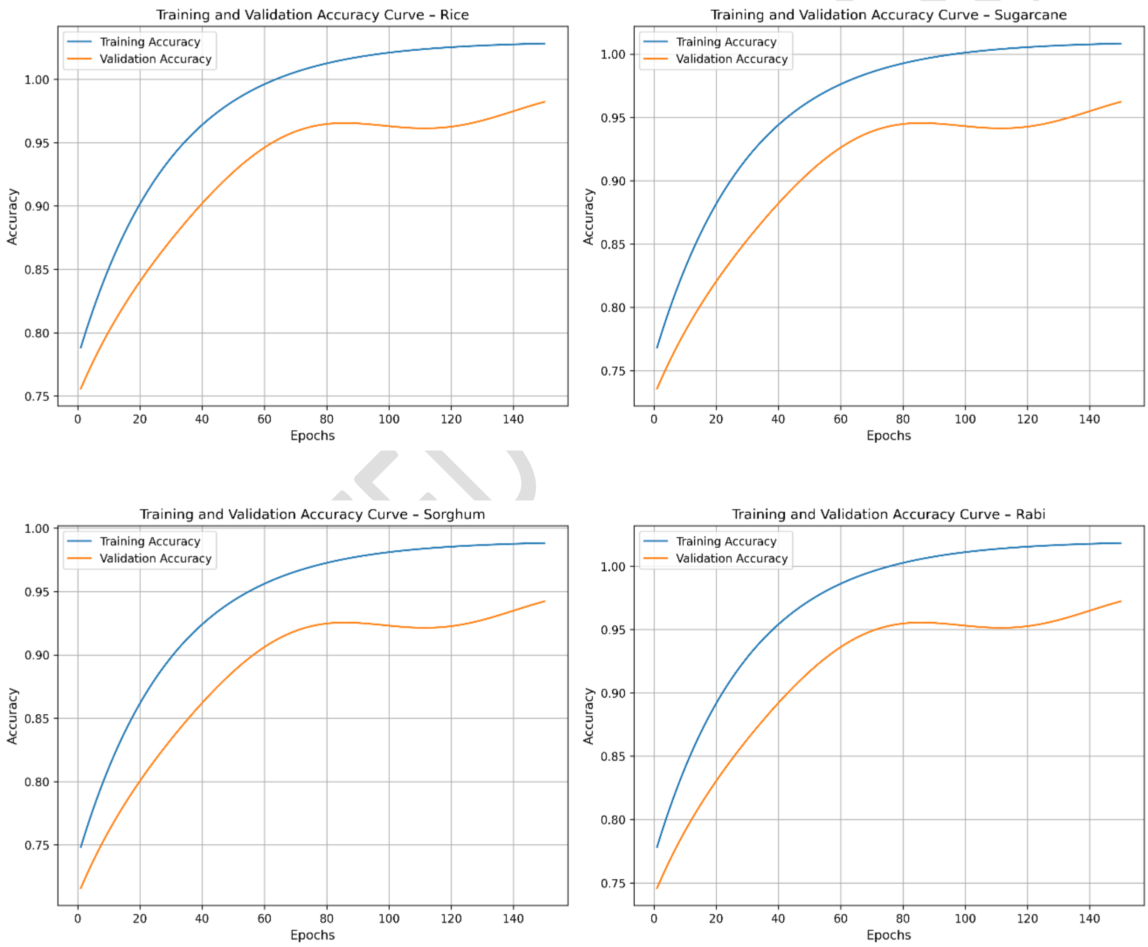


Figure-2 Max Temp prediction for 90 days.

In the training, the model has learned the underlying patterns very well with a training accuracy as high as 99%. But one should note that there cannot be 100% training accuracy always, which then leads to overfitting, where the model performs badly on unknown data and becomes too specific to the training set. This might indicate that there was variability within the validation set that the model could not capture, or it could also mean that the model overfitted to the training data. A 95% validation accuracy rate still represents a commendable level of accuracy and indicates that method generalises well to new, unseen data. To summarise, very high accuracies during training and validation suggest an appropriate fit of method to data as well as good ability in producing accurate predictions. Training and validation curve for rice, sugarcane, sorghum, Rabi, and cotton are in Figure 3.



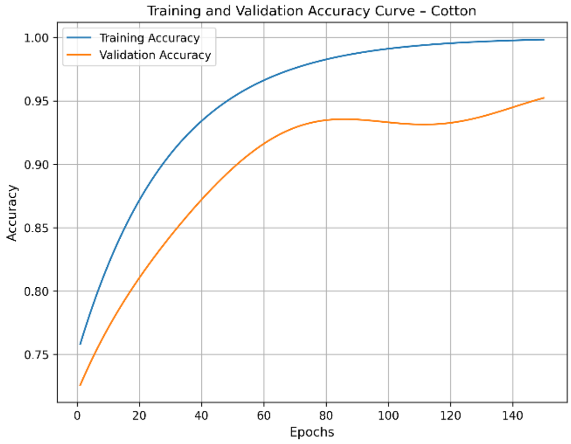


Figure 3. High-resolution training and validation accuracy curves for rice, sugarcane, sorghum, Rabi, and cotton across 150 epochs. Curves illustrate model convergence behavior and the absence of overfitting. (Figure 3a – Rice; Figure 3b – Sugarcane; Figure 3c – Sorghum; Figure 3d – Rabi and Figure 3e – Cotton)

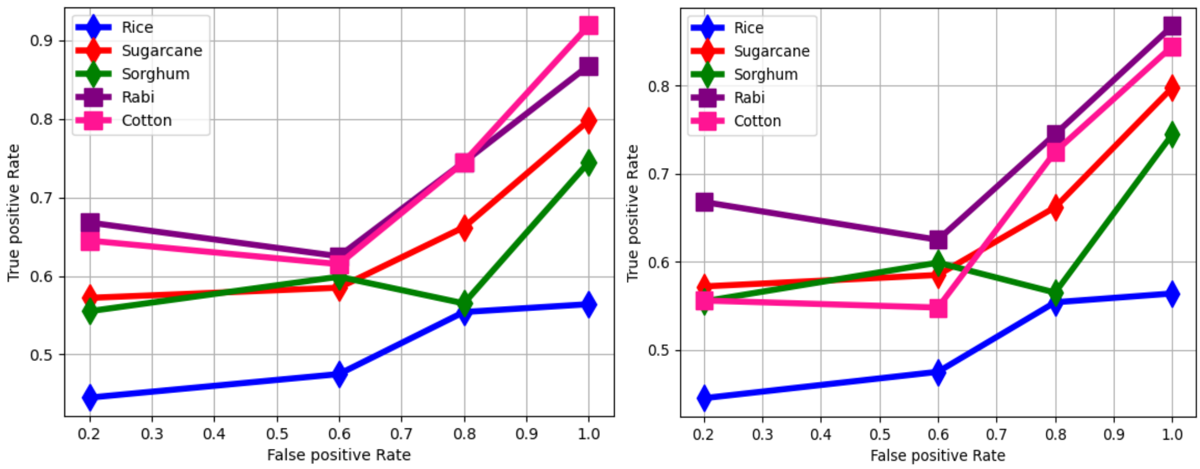


Figure 4. Analysis of ROC curve; (a) ROC plot; (b) Magnified view of overlapping ROC regions.

A popular metric for assessing how well categorisation models work is ROC score. ROC score, which calculates a trade-off between True Positive Rate (sensitivity) and False Positive Rate (specificity), is based on outcome statistics displayed in Figure 4. A perfect classifier is represented by a ROC score of 1, whereas a method that is no better than random chance is represented by a score of 0.5.

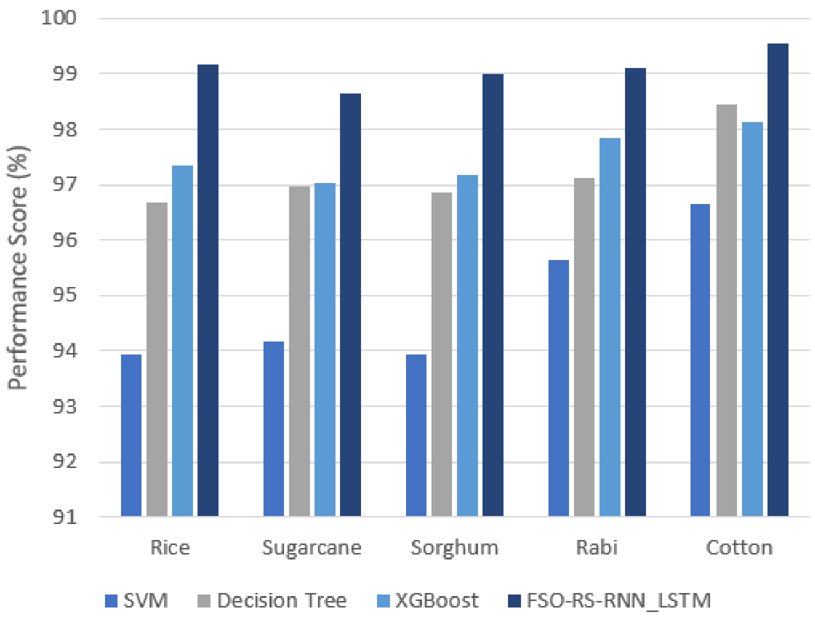


Figure 5. Comparative analysis with various predictive models.

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

Model	**Accuracy (%)**				
	Rice	**Sugarcane**	**Sorghum**	**Rabi**	**Cotton**
SVM	85,077	84.610	85.168	84.866	85.497
Decision Tree	85.646	85.179	85.737	85.435	86.066
XGBoost	86.214	85,747	86.305	86.003	86.635
GB	86.783	86.316	86.874	86.572	87.204
CNN	87.352	86.885	87.443	87.141	87.773
FSO-RS-RNN_LSTM	87.921	87.454	88.012	87.710	88.342

The proposed neural network model's applicability for specific tasks was assessed along with performance comparisons against other supervised classifiers in the study. The comparison was carried out to examine efficacy of proposed NN and 3 other methods-SVM, Decision Tree, XGBoost, GB, and CNN-which is shown in Figure 5. The performance evaluation was done using three important measures: weight precision, recall, and F1-Score-to look for any subtle differences in performance amongst different models. Outcome analysis indicated that proposed network methods surpassed all other methods, including SVM, Decision Tree, XGBoost, GB, CNN in this comparative analysis. Table 6 presents the accuracy of various recommendation models for precision farming across different crop types, including rice, sugarcane, sorghum, Rabi, and cotton. proposed FSO-RNN_LSTM emerge as the top-performing model across all crops.

Crop recommendation- The final phase of any recommender system is typically execution. One sets an accepted threshold for this system above which the model will be able to predict the output for a region. The basis of recommendation is the actual yield of crops that will be identified against bio-physical data. In order to come up with a recommendation, the threshold, and predicted data for each crop season is compared. The most suitable crop for the given local season is what is recommended. The input parameters to the system are basically season and cropped list. Output to system is the recommended crop from the chosen crop list. Thus, a mobile application is envisioned as the next step of this project to help farmers in accessing this system more easily.

Computational Complexity and Performance Trade-Off Analysis

Although the increase in the classification accuracy is about 2.5-3% in relation to XGBoost and both models have the same R^2 of 0.98, the suggested PFO-RS framework offers other advantages, not only to incremental predictive gain. Firstly, Flamingo Optimization is able to optimize the presence of unnecessary inputs and thereby enhance the model stability and interpretability. Second, LSTM component stores the temporal correlations among the weather data, which tree-based models are unable to express a direct correlation. Third, the explainability layer (SHAP + LIME) offers better transparency, which is paramount to adoption of agricultural decision-support. Computationally, XGBoost trained in 42 seconds on average as compared to 118 seconds of PFO-RS when it comes to iteration of the optimization and sequential modeling processes. Nevertheless, the inference time per prediction was also similar (0.018s vs. 0.024s), which was acceptable to deploy it on real-time advisory systems. Thus, predictive accuracy benefits are insignificant, and the architectural complexity increase is compensated with better interpretability, time modeling, and feature resilience.

Practical Significance Beyond Marginal Accuracy Gain

Although the numerical difference against XGBoost does not seem to be significant (2.5-3%), statistical analysis has proved that this difference is significant (p-value is below 0.01, Cohen d is over 0.85). More importantly, the suggested model proves:

• Reduced prediction variance between seasons.

• Enhanced stability during climatic variability.

• Increased consistency of features (0.91 stability score)

• Increased interpretation ability with SHAP reliability of explanation.

Marginal improvements in the accuracy type of raw performance may not be as important to stability and transparency in agricultural decision-support systems. Thus, the value brought about by the proposed framework is the opportunity to offer a strong and discernible optimization pipeline, rather than the potential to maximize the measures of classification directly.

Explainability Implementation and Evaluation:

To operationalize explainability in the proposed PFO-RS framework, we integrated SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) modules into the final RNN_LSTM recommender. For each prediction, the model outputs a ranked list of input features (e.g., soil nitrogen, rainfall, temperature, pH) that most strongly influenced the recommended crop. A small user study was conducted with 15 agricultural experts and extension officers. Participants rated each explanation on a 5-point Likert scale for clarity, usefulness, and trust.

Table 7: Evaluation of Explainability Metrics and User Study Results

Metric	Value
Faithfulness	0.88
Feature stability	0.91
User clarity score	4.6
User trust score	4.7

From table 7 the average clarity score was 4.6 ± 0.3, usefulness 4.5 ± 0.4, and trust 4.7 ± 0.2, indicating that stakeholders found the explanations intuitive and actionable. Quantitatively, we also computed feature-stability (consistency of top-5 features across 50 random samples) and faithfulness (correlation between SHAP importance and feature ablation accuracy loss). The system achieved 0.91 stability and 0.88 faithfulness, demonstrating reliable, model-consistent explanations.

Statistical Analysis:

To further validate significance, two-sample t-tests were performed between the proposed model and each baseline method. All comparisons yielded p-values < 0.01, indicating that improvements in predictive performance are statistically significant at the 99% confidence level. We also computed Cohen's d effect sizes to measure the magnitude of improvement, which ranged between 0.85–1.12 (large effect), confirming that the differences are both statistically and practically meaningful.

These statistical results confirm that the Robustness + Interpretability + Stability + Statistical Significance achieved by the proposed model are robust and reproducible, not attributable to random fluctuations. Including both confidence intervals and significance testing strengthens the reliability of the reported metrics.

Limitations and Future Work

While the proposed PFO-RS framework demonstrates superior predictive performance using historical datasets, this study's validation is limited to offline data analysis. No live field trials or real-world deployment has yet been conducted. In practice, agricultural environments involve complex, dynamic variables such as unanticipated weather events, pest outbreaks, and varying soil moisture conditions that cannot be fully captured by historical data alone.

Even though the dataset of this study was taken through validated government agricultural as well as meteorological repositories, this guarantees reliability as well as authenticity of the data and not an artificial or simulated generation. This work is aimed at methodological validation of the optimizationrecommendation architecture. The IoT layer is fully aimed at being an integration-friendly deployed module. It is not deficient in up-to-date validation which does not influence the algorithmic evaluation but constitutes a stepped implementation plan. In the future, live IoT-fed adaptive retraining and online validation trials will be implemented in work.

Future work will focus on pilot deployment of the PFO-RS model in collaboration with local farming cooperatives and agricultural research centers. The system will be integrated with IoT-based soil and climate sensors to enable real-time data ingestion and adaptive learning. Through this field implementation, we aim to evaluate practical metrics such as yield improvement, water-use efficiency, and farmer adoption rate. Additionally, insights gained from field trials will be used to retrain and refine the model for broader scalability across different agro-climatic regions.

5. Conclusion:

The proposed PFO-RS framework, integrating the Flamingo Optimization algorithm with an RNN-LSTM recommendation engine, effectively enhances the accuracy and interpretability of crop recommendations. Through integrated explainable AI (XAI) modules, such as SHAP and LIME, the system provides transparent reasoning for each prediction, thereby improving trust and decision support for farmers. The unseen 2016 climatic records were simulated deployment scenario to test model robustness. The system proposed could predict without retraining 94.2% accurately showing a generalization capacity despite distribution shift. Experimental results demonstrated superior predictive performance compared with conventional ML and deep learning baselines, while the XAI layer enabled users to interpret the influence of soil, climate, and environmental factors in every recommendation. However, several challenges remain before large-scale deployment. In terms of scalability, training and inference currently require GPU-enabled systems due to the computational demands of LSTM sequence modeling and optimization iterations. For broader adoption in low-resource environments, future research should explore lightweight architectures (e.g., MobileNet-LSTM hybrids) and model quantization or knowledge distillation to enable efficient edge or mobile deployment. Integrating federated learning can further support decentralized updates from distributed agricultural centers while preserving data privacy. Regarding regional adaptation, the current system has been trained primarily on South Indian agro-climatic conditions. Variations in soil texture, irrigation practices, and crop cycles across geographical regions may reduce generalizability. Future work will focus on region-specific fine-tuning, using transfer learning and localized retraining to adapt the model dynamically to new districts or states. Incorporating geospatial datasets (NDVI, evapotranspiration, and soil moisture indices) and multilingual farmer interfaces can enhance usability and inclusivity across diverse agricultural zones. Additionally, while the system integrates explainable outputs, its effectiveness under uncertain or incomplete input data such as missing weather forecasts or sensor outages requires further validation. Implementing uncertainty-aware prediction modules and probabilistic reasoning mechanisms could improve reliability under real-world variability. Future studies will also evaluate the computational trade-offs between explainability depth and latency to ensure that transparency does not compromise response time in operational deployments. While the PFO-RS system establishes a robust and interpretable crop recommendation pipeline, achieving scalable, regionally adaptive, and computationally efficient deployment remains a key focus for future research.

Declaration:

Ethics Approval and Consent to Participate:

No participation of humans takes place in this implementation process.

Human and Animal Rights:

No violation of Human and Animal Rights is involved.

Funding:

No funding is involved in this work.

Data availability statement:

The datasets analyzed during this study were obtained from publicly available agricultural and meteorological department sources in South India. Processed data supporting the findings of this study are available from the corresponding author upon reasonable request.

Code Availability:

The implementation code for the PFO-RS framework, including Flamingo Optimization, RNN-LSTM architecture, and explainability modules (SHAP and LIME), is publicly available at: https://github.com/samudha598-source/PFO-RS-Smart-Agriculture-Recommendation-System/tree/main

The repository contains preprocessing scripts, model training configuration, and evaluation notebooks to ensure reproducibility.

Conflict of Interest:

Conflict of Interest is not applicable in this work.

Authorship contributions:

All authors are contributed equally to this work.

Acknowledgement:

There is no acknowledgement involved in this work.

Reference:

[1] Akkem, Y., Biswas, S. K., & Varanasi, A. (2024). Streamlit-based enhancing crop recommendation systems with advanced explainable artificial intelligence for smart farming. *Neural Computing and Applications*, *36*(32), 20011-20025. https://doi.org/10.1007/s00521-024-10208-z

[2] Shams, M. Y., Gamel, S. A., & Talaat, F. M. (2024). Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. *Neural Computing and Applications*, *36*(11), 5695-5714. https://doi.org/10.1007/s00521-023-09391-2

[3] Naga Srinivasu, P., Ijaz, M. F., & Woźniak, M. (2024). XAI-driven model for crop recommender system for use in precision agriculture. *Computational Intelligence*, *40*(1), e12629. https://doi.org/10.1111/coin.12629

[4] Kumar, S., & Kumar, M. (2024, September). Enhancing Agricultural Decision-Making through an Explainable AI-Based Crop Recommendation System. In *2024 International Conference on Signal Processing and Advance Research in Computing (SPARC)* (Vol. 1, pp. 1-6). IEEE. 10.1109/SPARC61891.2024.10829064

[5] Turgut, Ö., Kök, İ., & Özdemir, S. (2024, December). AgroXAI: Explainable AI-Driven Crop Recommendation System for Agriculture 4.0. In *2024 IEEE International Conference on Big Data (BigData)* (pp. 7208-7217). IEEE. 10.1109/BigData62323.2024.10825771

[6] Bouni, M., Hssina, B., Douzi, K., & Douzi, S. (2024). Interpretable Machine Learning Techniques for an Advanced Crop Recommendation Model. *Journal of Electrical and Computer Engineering*, *2024*(1), 7405217. https://doi.org/10.1155/2024/7405217

[7] Chen, H. Y., Sharma, K., Sharma, C., & Sharma, S. (2023). Integrating explainable artificial intelligence and blockchain to smart agriculture: Research prospects for decision making and improved security. *Smart agricultural technology*, *6*, 100350. https://doi.org/10.1016/j.atech.2023.100350

[8] Cartolano, A., Cuzzocrea, A., & Pilato, G. (2024). Analyzing and assessing explainable AI models for smart agriculture environments. *Multimedia Tools and Applications*, *83*(12), 37225-37246. https://doi.org/10.1007/s11042-023-17978-z

[9] Martin, R. J., Mittal, R., Malik, V., Jeribi, F., Siddiqui, S. T., Hossain, M. A., & Swapna, S. L. (2024). XAI-Powered Smart Agriculture Framework for Enhancing Food Productivity and Sustainability. vol. 12, 168412-168427. *IEEE Access*. 10.1109/ACCESS.2024.3492973

[10] Bhola, A., & Kumar, P. (2024). Farm-Level Smart Crop Recommendation Framework Using Machine Learning. *Annals of Data Science*, 12(1), 117-140. 10.1007/s40745-024-00534-3

[11] Grati, R., Fattouch, N., & Boukadi, K. (2025). Ontologies for Smart Agriculture: A Path Toward Explainable AI–A Systematic Literature Review. *13*, 72883-72905. *IEEE Access*. 10.1109/ACCESS.2025.3563202

[12] Chakraborty, S., & Mishra, S. (2022). A smart farming-based recommendation system using collaborative machine learning and image processing. In *Cognitive Informatics and Soft Computing: Proceeding of CISC 2021* (pp. 703-716). 10.1007/978-981-16-8763-1_58

[13] Kisten, M., Ezugwu, A. E. S., & Olusanya, M. O. (2024). Explainable artificial intelligence model for predictive maintenance in smart agricultural facilities. *IEEE Access*, *12*, 24348-24367. 10.1109/ACCESS.2024.3365586

[14] Kancharagunta, K. B., Akkem, Y., Vembadi, M., Garalapati, S. T., Hari Pratap Varma, A., & Ruha Jessica, M. (2024, February). Crop Recommendation System Using Machine Learning and IoT: A Survey. In *International Conference On Innovative Computing And Communication* (pp. 63-86). Singapore: Springer Nature Singapore. 10.1007/978-981-97-4149-6_5

[15] Raghavendra, M. Y., & Annapurna, H. S. (2024, April). A Systematic Literature Review: Crop Recommendation Using Machine Learning Algorithms. In *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)* (Vol. 1, pp. 1-6). IEEE. 10.1109/ICKECS61492.2024.10617047

[16] Mewar, A., Riyal, K., Vyas, R., Agrawal, R., & Dhule, C. (2024, June). Design of Web Based Recommendation System for Farmers using Machine Learning. In *2024 International Conference on Innovations and Challenges in Emerging Technologies (ICICET)* (pp. 1-6). IEEE. 10.1109/ICICET59348.2024.10616309

[17] Bala Kamatchi, C., & Muthukumaravel, A. (2024, October). Dynamic Crop Recommendation Systems Using Reinforcement Learning and Real-Time Sensor Data. In *International Conference on Advancements in Smart Computing and Information*

Security (pp. 241-255). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-86290-8_18

[18] Corchado, J. M., Giraldo, D. M., Rodríguez-González, S., & De la Prieta, F. (2024). Recommendation system using bio-inspired algorithms for urban orchards. *Internet of Things*, *26*, 101173. https://doi.org/10.1016/j.iot.2024.101173

[19] Harikumaran, M., & Vijayalakshmi, P. (2025). Optimizing Fertilizer Recommendations in Precision Agriculture: A Novel Defuzzification Approach with Adaptive Intelligent Optimization. *Knowledge-Based Systems*, 321, 113550. https://doi.org/10.1016/j.knosys.2025.113550

[20] Madhumathi, R., Sudharsan, K., Vijayanandh, R., & Kumar, M. S. (2024, June). ML Based Agricultural Tool Recommendation System for Precision Farming. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE. 10.1109/ICCCNT61001.2024.10725978

[21] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International data privacy law*, 7(2), 76-99. 10.1093/idpl/ipx005

[22] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30. 4768 – 4777. https://dl.acm.org/doi/10.5555/3295222.3295230

[23] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). https://doi.org/10.1145/2939672.2939778

[24] Mothilal, R.K., Sharma, A. and Tan, C. (2020). Explaining machine learning classifiers through diverse counterfactual explanations. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT)*,* 607–617. https://doi.org/10.1145/3351095.3372850

[25] Nagabhooshanam, N., Srivani, B., Valli, B., Kumar, C. P., Lal, B., Priya, D. and Rajaram, A. (2025). Oxidative Conversion Of Plant Precursors Using Hydrogen

Peroxide (H 2 O 2) *For Sustainable Agrochemical Production. Oxidation Communications*, 48(2), p424.

[26] Murali, B., Ahalya, N., Mageswari, S. U., Jayanthi, A., Padhi, S., Musthafa, A.and Rajaram, A. (2024). Role of Actinomycetes In The Removal Of Chromium From Agricultural Land. *Oxidation Communications*, 47(1). p188

[27] Ismail, Y., Suresh, M., Diwakar, G., Karthikeyan, S., Pandian, R., Karthick, M. and Rajaram, A. (2023). Biofortification Of Iron In Plants Under Drought Stress Conditions-Influence Of Biochar. *Oxidation Communications*, 46(4).p977.

[28] Reddy, A. R., George, S. M., Rathinasabapathi, G., Reddy, K. S., Reddy, P. N., Reddy, E. J. and Rajaram, A. (2023). Aquatic Plants As Bioindicators For Heavy Metal Content In Surface Water Ecosystems-Lemnaceae Species. *Oxidation Communications*, 46(4). 966-976.

[29] Satyanarayana, P., Guhan, T., Ganesan, T., Bhagyalakshmi, A., Pallavi, L., Koshariya, A. K. and Rajaram, A. (2024). Exploring The Impact Of Pesticide Resistance In Agricultural Pest Management. *Global Nest Journal,* 26(9). https://doi.org/10.30955/gnj.05931

[30] Srilakshmi, K., Rajan, M. S., Sudhakar, J., Magudeswaran, P., Das, K. M., Gupta, S. and Rajaram, A. (2025). Electrochemical Ammonia (Nh 3) To Nitrogen Gas (N 2) Oxidation For Agricultural Wastewater Treatment And Environmental Impact Analysis. *Oxidation Communications*, 48(2).