National Conference on Behavioural Economics and Intelligent Decision Systems for Climate Change and Sustainable Development
Int. J. Adv. Res. in Artificial Intelligence and Machine Learning Reviews
Volume 2, Issue 1 - 2026

Review Article

AI-Driven Precision Farming: A Scalable Neural Approach for Crop Prediction

Bikramaditya Chakraborty[1], Vibhor Sharma[2], Anurag Singh[3]

[1,2]Research Scholar, [3]Associate Professor, Department of CSE, NIT Delhi, New Delhi, India
DOI: https://doi.org/10.24321/3117.4809.202604

I N F O

Corresponding Author:
Vibhor Sharma, Department of CSE, NIT Delhi, New Delhi, India
https://orcid.org/0009-0002-7837-1277
E-mail Id:
232210011@nitdelhi.ac.in
How to cite this article:
Chakraborty B, Sharma V, Singh A. AI-Driven Precision Farming: A Scalable Neural Approach for Crop Prediction. *Int J Adv Res Artif Intell Mach Learn Rev* 2026; 2(1): 25-31.

Date of Submission: 2025-10-04
Date of Acceptance: 2025-10-28

A B S T R A C T

Population growth, urbanisation, and climate change have continued to put pressure on the world's food demand, creating enormous pressure on agricultural systems to become more productive and sustainable. Precision farming is an application that utilises information to optimise crop production and resource utilisation. The paper uses JK soil analysis data, which has been trained on a machine learning model to provide predictions for the crops. Environmental factors, such as Rainfall and Temperature, were evaluated, along with soil properties (pH, N, P, K, Zn, Fe, and Mn). To achieve high data quality and reliability, sophisticated preprocessing methods were employed. Z-score-based filtering was used to eliminate outliers, and the Min-Max scaling was used to standardise the input space by normalising numerical features. The Chi-Squared test was used to reduce the 10 most significant features for further prediction. The preprocessing pipeline was thorough, and noise and redundancy were reduced to provide a solid dataset for training the model. A neural network with two hidden layers (64 and 32 neurons) is used, achieving 98.45% test accuracy. The method provides a basis for real-time, scalable, and accurate farming frameworks, as well as possible applications in pest and fertiliser management.

Keywords: Precision Agriculture, Crop Recommendation System, Feature Selection, Scalable Framework

Introduction

Traditional farming methods, which in many cases are based on experience and manual decision-making, often fail to address contemporary agricultural problems such as soil erosion, nutrient imbalances, and climatic uncertainty. The shift to a system of precision farming entails the implementation of principles from artificial intelligence (AI), machine learning (ML), and data analytics in the agricultural system to enable informed decisions, optimal crop control, and the sustainable use of resources.[1,2] Precision agriculture enables the assessment of massive soil, climatic, and environmental data to inform crop recommendations and soil fertility management. State-of-the-art work has used ML and deep learning (DL) algorithms to classify the type of soil, measure the content of nutrients, and determine the crops that yield the most profitable results under different conditions.[1,2,3,4] Ensemble models and deep neural networks have shown considerable potential in automating crop recommendation procedures, and IoT-enabled Systems have also provided better real-time monitoring and decision support systems.[3,5,6] Empirical comparisons between various ML models have also indicated that sound preprocessing, feature selection, and incorporation of environmental factors can be vital to stable and scalable prediction systems.[7,8,9,10]

AI-based systems in agriculture have achieved several gains, the current models have various limitations. The majority of frameworks tend to exhibit less generalisation in heterogeneous areas related to agriculture due to the uneven quality of data and the absence of end-to-end preprocessing pipelines. The lack of focus on interpretability and feature optimisation also limits their practical use in supporting decision-making. Additionally, the issue of scalability persists, as most models are trained on smaller datasets that are limited to a local area and do not account for regional differences in soil composition or climatic variations. To overcome these difficulties, there is a strong need for a robust framework that integrates efficient data preprocessing, automatic feature selection, and dynamic learning structures to manage diverse settings within the environment. Major Contributions are as follows:

- **Enhanced Data Preprocessing:** Data quality and model stability: Z-score-based outlier identification and Min-Max normalization.
- **Feature Optimisation:** Statistical test Chi-Squared test was applied to explore important attributes that were predictive, such as pH, nitrogen, phosphorus, potassium, rainfall, and temperature.
- **Architecture as a Data Story:** Creation of a two-layer neural network with ReLU and SoftMax activation functions, with optimization using Adam and categorical cross-entropy loss.
- **Scalability of the system:** Development of a modular ML architecture with the capacity of incorporating IoT-based data from the environment, and capable of being scaled to large-scale agricultural instalments.

The suggested framework facilitates the further development of precision agriculture by combining optimized preprocessing methods, statistical feature selection, and adaptive neural modeling. The methodology enhances scalability, interpretability, and generalization, enabling the provision of successful crop advice across a range of soil and climatic regions.

Literature Review

Precision agriculture research has also focused on the application of artificial intelligence and machine learning to enhance crop management, soil analysis, and decision support systems. The most commonly used studies can be divided into three general directions: traditional machine learning, hybrid models, and deep learning frameworks, as well as their integration with Internet of Things (IoT) systems.

Machine Learning-Based Crop Recommendation

The early studies in the area of crop recommendation were mostly based on the classical algorithms of machine learning, including Decision Trees, Random Forest, Naive Bayes, Support Vector Machines, and Logistic Regression to process the data of soil nutrients and environmental factors.[1,2,3] With these models, it was revealed that agricultural decision-making could be automated through the discovery of hidden interactions between soil composition and crop suitability. Their performance, however, was highly reliant on the quality of the data sets, the diversity of features, and the efficiency of preprocessing. Later works emphasised that the mismanagement of data outliers, imbalance of the classes, and a lack of optimisation of feature selection might lead to non-uniform behaviour of the model across regions, which is problematic and should be minimised, whether it is a physical layer or a behavioural layer of the solution under consideration, located between the input and output devices, respectively.[4]

Table I.Summary Of Existing Research On Ai-Based Crop Recommendation Systems

Ref.	Technique Used	Data/Features Considered	Key Contributions
[1]	Ensemble ML (CatBoost, RF, SVM)	Soil and climatic parameters.	Framework for AI-driven crop prediction using hybrid learning.
[2]	Random Forest and Feature Correlation	Soil nutrient dataset	Improved soil classification with correlation-based feature selection.
[3]	Explainable ML (XAI + Ensemble)	Soil fertility data	Enhanced interpretability of fertility assessment.
[4]	IoT-integrated ML system	Real-time soil sensor data	Continuous nutrient monitoring and crop recommendations.
[5]	CNN + ML Hybrid Model	Soil and image-based datasets	Integration of DL with image recognition for precision farming.
[6]	Majority Voting Ensemble	Soil and climate data	Multi-model aggre-integration for improved reliability.

ISSN: 3117-4809
DOI: https://doi.org/10.24321/3117.4809.202604

[7]	Optimized Ensemble (RF, SVM, NB)	Multivariate soil dataset	Enhanced ensemble stability and generalization.
[8]	Random Forest Model	Web-interfaced soil dataset	User-oriented crop recommendation framework.
[9]	CNN + N L P Model	Soil texture and textual data	Combined visual and linguistic data for soil texture-based recommendation.
[10]	Comparative ML Study	Standardized agricultural dataset	Evaluation of mul-Triple algorithms for efficiency and scalability.
[11]	Deep Neural Network + IoT	Regional soil and climate data	Integrated prediction of fertility, crop type, and nutrient requirements.

Hybrid Modeling Approaches

Ensemble and hybrid modelling were presented in the second stream of work to enhance robustness and predictive accuracy. The Random Forest, Gradient Boosting, and stacking-based classifiers were among the techniques that utilised multiple base learners to counteract overfitting and enhance generalisation capabilities.[5,6,7] These methods were effective in complex agricultural settings through the combination of several decision-making routes for soil and crop forecasting.

Deep Learning and IoT-Integrated Systems

The combination of AI, deep learning, and IoT has further increased the prospect of AI in precision agriculture. Nonlinear relationships between soil nutrients, temperature, humidity, and rainfall have been learnt in deep neural networks and convolutional structures, as well as artificial intelligence (AI) systems.[5,8,9,11] Additionally, IoT-based systems have been designed to gather real-time soil and weather information, enabling the creation of adaptive crop recommendation systems.[4,10,11] These models have shown good learning properties but have been found to be weak in scalability as they require region-specific training data and are also computationally intensive. In Table I, a comparative summary of past research is provided, with an emphasis on the methodology used, the type of data utilised, and the main findings of each study. The comparison presents the shift from traditional ML-based approaches to hybrid, deep learning, and IoT-based systems in intelligent crop recommendation.

Research Gaps

Although current agricultural intelligence systems have made significant progress, several challenges remain. Most of them are only trained using localised datasets, so they are not very transferable across areas with different soils and climatic conditions. Preprocessing methods are often not utilised to their full potential, resulting in inconsistencies and reduced interpretability of models. Additionally, deep learning and ensemble algorithms enhance predictive ability, but they require large computational resources and are not easily explainable. The combination of effective preprocessing, feature selection, and adaptive neural modelling within a single framework is necessary to achieve proper, scalable, and interpretable crop suggestions in various settings.

Methodology

The approach combines the use of high-quality data preprocessing, feature optimisation, and neural network modelling to develop a powerful and scalable system for intelligent crop recommendations. Figure 1 illustrates the overall process, which encompasses consecutive steps from dataset acquisition to deployment.

Dataset Details

The JK Soil Analysis Dataset was used to develop the model, comprising 24,747 records with 15 distinct features, including soil macronutrients, micronutrients, and environmental factors. These parameters are Nitrogen (N), Phosphorus (P), Potassium (K), Sulfur (S), pH, Electrical Conductivity (EC), Organic Carbon (OC), and trace elements such as Zinc (Zn), Iron (Fe), Copper (Cu), and Manganese (Mn). Otherwise, Rainfall and Temperature are considered to explain climatic variability. The set of data has a large variety of agricultural settings, which underlies the flexibility of a machine learning system that can be applied to operate on conditionally different soils.

The target variable in the dataset is a categorical variable in which all the items in the dataset are categorised into one of the three classes of crops: Class 0 is Staple crops such as Wheat, Rice, and Maize; Class 1 is oilseeds and cereals such as Mustard, Barley, and Millet; Class 2 is high value crops such as vegetables and Sugarcane. The multi-class structure of this model enables it to make generalisations between nutrient-demanding and drought-resistant varieties, as the agronomic variety of the region is likely to differ.

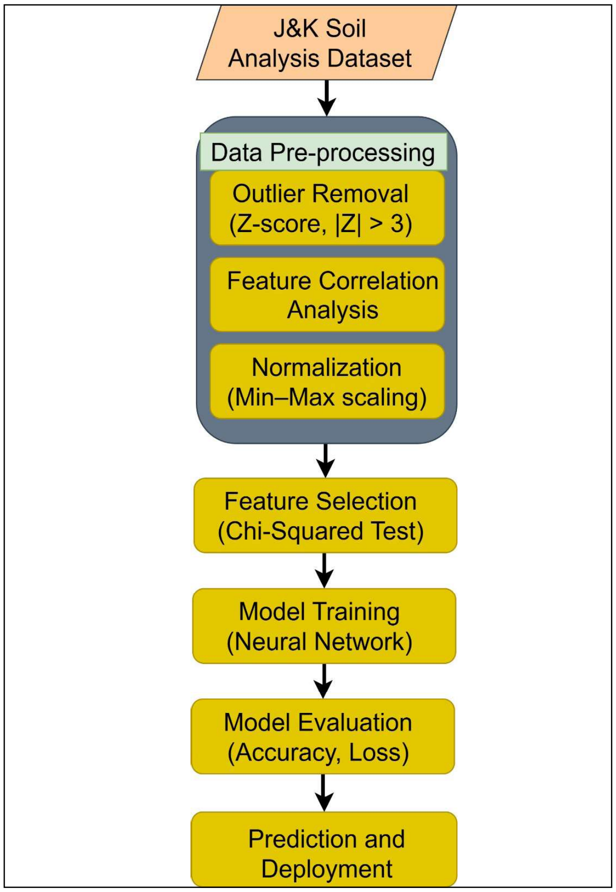

Figure 1.The diagram illustrates the sequential phases, from data preprocessing and feature correlation analysis to feature selection, model training, evaluation, and deployment

The similar structure of datasets and regional adjustments was also utilised[1], focusing on a Smart New Product[2] that improves the product design process. An example of this is that previous research has used local soil and weather information to train machine learning models for crop predictions; however, most of these studies lacked detailed micro-nutrient analysis or feature standardisation, thereby reducing their local predictive capability. The addition of climatic parameters and macro- and micro-nutrient values that follow is another factor that enhances predictive power and generalisation in the present work.

Data Preprocessing

The data quality of the models and their high robustness were ensured through a comprehensive preprocessing pipeline that included outlier removal, normalisation, and correlation analysis. Similar preprocessing methods were, however, applied in the research on ML- and IoT-based precision farming 4– 6, where preprocessing of data before training the model is noted to be crucial.

Outlier Removal: The Z-score was used to identify and delete outliers

$$Z = \frac{x_i - \mu}{\sigma} \qquad (1)$$

The data points with an absolute value of Z greater than 3 were considered anomalies and eliminated. This guarantees the elimination of unrealistic soil or environmental values (e.g., high pH or rainfall). The type of filtering with Z-scores was also employed in ensemble as well as CNN-based systems to improve them [6,7,9] specifically for enhancing data stability before training.

Normalization: The Z-score was used to identify and delete outliers

$$Z = \frac{x_i - \mu}{\sigma} \qquad (2)$$

The data points with an absolute value of Z greater than 3 were considered anomalies and eliminated. This guarantees the elimination of unrealistic soil or environmental values (e.g., high pH or rainfall). This type of filtering using Z-scores was also employed in ensemble and CNN-based systems to improve them.[6,7]

Feature Correlation Analysis: Analysis of correlation between features was done to determine redundancy in soil and environmental features:

$$r_{i,j} = \frac{Cov(f_i, f_j)}{stigma(f_i) stigma(f_j)} \qquad (3)$$

Feature Selection

To define the most significant features associated with the type of crop, the Chi-Squared test was used:

$$X_i^2 = \sum_k \frac{(O_{i,k} - E_{i,k})^2}{E_{i,k}} \qquad (4)$$

The top ten influential features were maintained, including pH, N, P, K, S, Zn, Fe, Mn, Rainfall, and Temperature. It has been demonstrated in other studies that similar strategies, also based on selection by simulating chi-square or correlation, are successful.[2,5,7] In the case of artificial feature subsets. In the case of artificial feature subsets. In the case of artificial feature subsets, In the case of artificial feature subsets. In the case of artificial feature subsets, In the case of artificial feature subsets, cite: In the case of artificial feature subsets, In the case of artificial feature subsets, cite. Although other previous works relied solely on the information about the nutrients to be predicted, the inclusion of climatic parameters, such as Rainfall and Temperature, thus enhances its capability to adapt to various climatic conditions.

ISSN: *3117-4809*
DOI: https://doi.org/10.24321/3117.4809.202604

Algorithm 1 Crop Recommendation System

Input: Dataset D, Feature F, Target T
Output: Predicted crop class P
procedure $PREPROCESSING$
  for all $f_i \in F$ **do**
    for all $samples\ x_i \in D$ **do**
      $Z_{i,j} \leftarrow \frac{x_{i,j}-\mu_i}{\sigma_i}$
      if $|Z_{i,j}| > 3$ **then**
      remove sample x_j
      $x'_{i,j} \leftarrow \frac{x_{i,j}-\min(f_i)}{\max(f_i)-\min(f_i)}$
procedure $FEATURESELECTION$
  $Compute\ correlation\ matrix\ C$
  $X_i^2 \leftarrow \sum_k \frac{(O_{ik}-E_{ik})^2}{E_{ik}}$
  $F_{sel} \leftarrow \mathrm{TopK}(\{X_i^2\})$
procedure $MODELTRAINING$
  $Define\ NN: |F_{sel}| \rightarrow 64 \rightarrow 32 \rightarrow 3\ (ReLU, Softmax)$
  $Compile\ with\ Adam, categorical\ cross\ entropy$
  $Train\ on\ D_{train};\ validate\ on\ D_{test}$
procedure $PREDICTION$
  for all $new\ inputs\ x_{new}$ **do**
    $Normalize\ x_{new}\ as\ in\ PREPROSSING$
    $y \leftarrow NN(x_{new})$
    $P \leftarrow argmax(Softmax(y))$
  $Deploy\ model\ for\ realtime\ recommendations$

Neural Network Architecture : The structure of a multi-class neural network using the selected features as inputs was constructed as a feed-forward neural network:

$$h_1 = ReLU(W_1X + b_1), h_1 \in \mathbb{R}^{64} \quad (5)$$

$$h_2 = ReLU\ (W_2h_1 + b_2), h_2 \in \mathbb{R}^{32} \quad (6)$$

$$y = Softmax(W_3h_2 + b_3), y \in \mathbb{R}^3 \quad (7)$$

The ReLU activation function contributes to non-linear learning, whereas SoftMax provides the probability of classes in the categories of crops. The training of the model was based on the Adam optimiser and the categorical cross-entropy loss:

$$L = -\sum_{i=1}^{c} y_i^{true} log(y_i^{pred}) \quad (8)$$

Network configurations similar to the 64–32–3 architecture was also found in ensemble-hybrid systems[6,7,8] though those models often lacked structured preprocessing and feature ranking. The proposed framework enhances these architectures by integrating preprocessing rigour and data interpretability prior to model learning. Algorithm 1 outlines the overall process, from data cleaning to model deployment.

Training and Prediction: The model was trained with an 80:20 train-test split with a batch size of 64 and 200 epochs. The training network is used to normalise new soil data and process them during the prediction:

$$P = argmax\left(Softmax\left(W_3\left(ReLU\left(W_2\left(ReLU(W_1X + b_1)\right) + b_2\right)\right) + b_3\right)\right) \quad (9)$$

This enables real-time, data-driven crop advisory services, incorporating real-time soil and environmental data. The methodology being proposed builds upon the previous work in the following aspects:

- Similar preprocessing frameworks were observed in works focusing on IoT-assisted soil nutrient monitoring and hybrid ML models.[4,6]
- Ensemble-based systems[5,7] emphasized combining multiple classifiers, whereas the current framework focuses on deep, feature-optimized learning for improved generalization.
- Deep learning models integrated with environmental data[8,9,11] validated the inclusion of climatic parameters, a direction also adopted here to enhance scalability.
- Chi-squared-driven feature selection, as implemented in earlier soil fertility prediction models[2,10] reinforced the effectiveness of statistical selection in agricultural datasets.

This is achieved by employing the entire methodology, which consists of fine-tuning the data preprocessing, optimising features using the Chi-Squared method, and utilising deep neural modelling, thereby ensuring that the trade-off between interpretability and performance is balanced. The current system, in comparison to the previous systems, combines environmental information, detailed preprocessing, and statistical prioritisation of features to provide a scalable and interpretable crop recommendation system based on precision agriculture.

Results And Discussion

The proposed neural network model, developed using advanced preprocessing techniques, achieved robust performance in crop classification. The results demonstrate the effectiveness of the methodology in leveraging soil and environmental data for precision farming.

Model Performance

The neural network model was structured with two hidden layers, designed to effectively handle the dataset's complexity. The model's architecture is summarised as follows: The Input Layer comprises 10 neurons corresponding to soil and environmental features. The first

ISSN: 3117-4809
DOI: https://doi.org/10.24321/3117.4809.202604

hidden layer consists of 64 neurons with ReLU activation. The second hidden layer consists of 32 neurons with ReLU activation. The output layer has 3 neurons (for multi-class classification) with Softmax activation.

The model achieved the following performance metrics:

- Training Accuracy: 98.45%
- Validation Accuracy: 99.00%
- Test Accuracy: 98.46%
- Test Loss: 0.0389

These results demonstrate consistent and robust performance across training, validation, and testing datasets, confirming the generalisability and reliability of the proposed neural network.

Comparative Analysis

A comparative study was conducted between the proposed work and prior research titled Data-Driven Analysis and Machine Learning-Based Crop and Fertilizer Recommendation System for Revolutionising Farming Practices. The results are summarised in Table II.

Table 2.Comparative Analysis Of Model Performance

Approaches	Features Used	Architecture	Accuracy (%)
Data-Driven Analysis and ML-Based Crop System [5,12,13]	pH, N, P, K	64-32-3 (two hidden layers)	97.00
Proposed Work	pH, N, P, K, Zn, Fe, Mn, Rainfall, Temperature	64-32-3 (two hidden layers)	98.46

The work made a significant improvement in the accuracy of the tests by incorporating additional soil and environmental characteristics, as opposed to the work used as a reference for this work. The enhancement underscores the importance of incorporating various and useful functionalities to enhance predictive accuracy in precision farming.

The combination of soil characteristics, environmental conditions, and state-of-the-art pre-processing into the designed neural network model resulted in a highly accurate and reliable crop recommendation system. The comparison with existing research highlights the impact of incorporating new features, demonstrating the effectiveness of the proposed methodology in precision agriculture applications.

Conclusion

The paper proposes a precision farming system based on AI, which incorporates sound preprocessing, statistical feature optimisation, and neural network modelling to precisely predict crop yields. Outlier rejection using Z-scores and Min-Max normalisations enhanced the reliability of the data, whereas Chi-Squared feature selection improved the interpretability and efficiency of the model. The proposed neural network, with two hidden layers, achieved a test accuracy of 98.46%, surpassing similar models trained on limited features. The framework had good generalisation to a variety of agricultural conditions by integrating soil macronutrients, micronutrients, and environmental variables. The findings suggest the potential of scalable, data-driven AI systems to support sustainable and informed crop management. Future directions will involve incorporating real-time IoT data and adaptive learning to continually model and update it in dynamic field environments.

References

1. S. K. Apat, J. Mishra, K. S. Raju, and N. Pady, “An artificial intelligence-based crop recommendation system using machine learning,” Journal of Scientific & Industrial Research (JSIR), vol. 82, no. 05, pp. 558–567, 2023.
2. M. Suchithra and M. L. Pai, “Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters,” Information processing in Agriculture, vol. 7, no. 1, pp. 72– 82, 2020.
3. M. R. Islam, K. Oliullah, M. M. Kabir, M. Alom, and M. Mridha, “Machine learning enabled iot system for soil nutrients monitoring and crop recommendation,” Journal of Agriculture and Food Research, vol. 14, p. 100880, 2023.
4. S. K. S. Durai and M. D. Shamili, “Smart farming using machine learning and deep learning techniques,” Decision Analytics Journal, vol. 3, p. 100041, 2022.
5. S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, “Crop recommendation system for precision agriculture,” in 2016 Eighth International Conference on Advanced Computing (ICoAC), pp. 32–36, IEEE, 2017.
6. N. H. Kulkarni, G. Srinivasan, B. Sagar, and N. Cauvery, “Improving crop productivity through a crop recommendation system using ensembling technique,” in 2018 3rd international conference on computational systems and information technology for sustainable solutions (CSITSS), pp. 114–119, IEEE, 2018.
7. S. K. Upadhyay et al., “Intelligent crop recommendation using machine learning,” in 2024 International Conference on Automation and Computation (AUTOCOM), pp. 330–335, IEEE, 2024.
8. N. Mittal and A. Bhanja, “Implementation and identification of crop based on soil texture using AI,” in 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1467–1471, IEEE, 2023.
9. K. El Barrak, S. Lakhal, and O. Abdoun, “Crop

recommendation using machine learning algorithms," in Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security, pp. 1–5, 2024.
10. D. Varshitha and S. Choudhary, "An artificial intelligence solution for crop recommendation," Indonesian Journal of Electrical Engineering and Computer Science, vol. 25, no. 3, pp. 1688–1695, 2022.
11. M. Sujatha and C. Jaidhar, "Machine learning-based approaches to enhance the soil fertility—a review," Expert Systems with Applications, vol. 240, p. 122557, 2024.
12. F. S. Prity, M. M. Hasan, S. H. Saif, M. M. Hossain, S. H. Bhuiyan, M. A. Islam, and M. T. H. Lavlu, "Enhancing agricultural productivity: a machine learning approach to crop recommendations," Human-Centric Intelligent Systems, vol. 4, no. 4, pp. 497–510, 2024.
13. P. S. Kiran, G. Abhinaya, S. Sruti, and N. Padhy, "A machine learning- enabled system for crop recommendation," Engineering Proceedings, vol. 67, no. 1, p. 51, 2024.