



OPEN Enhancing precision agriculture through cloud based transformative crop recommendation model

Gurpreet Singh & Sandeep Sharma

Modern agriculture relies more on technology to boost food production. It aims to improve both the quality and quantity of food. This paper introduces a novel TCRM (Transformative Crop Recommendation Model). It uses advanced machine learning and cloud platforms to give personalized crop recommendations. Unlike traditional methods, TCRM uses real-time data. It includes environmental and agronomic factors to optimize recommendations. The system has SMS alerts for remote farmers. It outperforms baseline algorithms like Logistic Regression, KNN(k-nearest neighbor), and AdaBoost. TCRM empowers farmers with actionable insights, reducing resource wastage while boosting yield. By offering region-specific recommendations, it enhances profitability and promotes sustainable agricultural practices. The model has 94% accuracy, 94.46% precision, and 94% recall. Its F1 score is 93.97%. The fivefold cross-validation score is 97.67%. These findings show that the model can improve precision farming. It can make agriculture more sustainable and efficient.

Keywords Crop recommendation, Internet of things, Machine learning, Precision agriculture

One of the first pillars of human civilization, agriculture has changed a lot over the years. As the earth's population explodes, demand for this vital sector has soared. Agriculture is now a tech-driven industry. It focuses on data-based innovations and decisions. Manufacturers can help meet the urgent need for more food in agriculture. In recent years, many new technologies have emerged since the agricultural revolution. It aims to equip farmers for modern farming^{1,2}. A quickening transformation in the agriculture sector is being driven by the growth of Internet of Things (IoT) technology³. The need for food production is growing quickly because it is predicted that by 2050 there will be 9.7 billion people on the planet⁴. Agriculture is now a tech-driven industry. It focuses on innovations and data-based decisions. Manufacturers must use these to meet the urgent need for food⁵. These technologies are key to high, sustainable agricultural yields to ensure efficient farming. At the heart of this overhaul is the use of data. Researchers collect it with care from various sources and analyze it. This data will inform decisions with high accuracy and speed⁶. A novel, cloud-based crop recommendation platform for precision farming is being designed. At the heart of precision farming is a mantra. It says that we can optimize farming by making it site-specific to each field's uniqueness. Using machine learning⁷ algorithms such as decision trees, random forests, extra trees, logistic regression, SVM (support vector machines), KNN (k-nearest neighbor), DT (decision tree), XGBoost (extreme gradient boosting), Gaussian naive Bayes, AdaBoost, and bagging classifiers. It aggregates and preprocesses raw data to combine several datasets from multiple sources. The researchers want to find the best algorithm. It should give farmers personalized crop suggestions. The recommendations come from a vast database. It includes environmental factors, agronomic variables, and past farm performance data. It is hoped to provide farmers with the information they need. This will help them make timely, data-driven decisions. ⁸ proposed a rise in using advanced ensemble learning in precision agriculture. This was especially true for crop recommendation systems. Traditional machine learning often fails with complex, high-dimensional agricultural data. New methods, like stacking, an ensemble of ensembles, and federated learning, are now strong alternatives. It offers better accuracy, adaptability, and scalability. These methods use diverse models. Methods aim to improve predictions and reduce bias. It also tackles issues like data diversity and privacy. Combining these ensemble techniques with user-friendly platforms like Streamlit boosts their use. It enables efficient, interactive decision-making for sustainable agriculture. ⁹ discussed the lack of labeled data in crop recommendations. It

Department of Computer Engineering and Technology, Guru Nanak Dev University, Amritsar, Punjab, India. email: gurpreetsinghgndu@gmail.com

noted that generative models like VAEs and GANs could help. These models create synthetic data that mimics real-world crop data. This improves the training and performance of agricultural recommendation systems.

¹⁰ proposed a cloud-based crop recommendation platform. It would provide a wealth of real-time recommendations. It has an SMS alert system. It will keep farmers, especially in remote areas. This new technology could revolutionize agriculture. It would improve efficiency and sustainability. Farmers can thus be equipped with actionable, data-based, real-time insights. It can use resources better to boost crop yields. It supports a global goal of safe food for future generations¹¹. The next posts will cover the why, how, and lessons learned from a paradigm-shifting approach to modern agriculture. The growing human population raises the demand for farming. It poses a global food security challenge. With the human population expected to rise by 9.7% in a few decades, our agricultural systems face immense pressure. Arab Spring, a need to feed people, and a youth surge in fast-growing regions. However, we will need new farming methods to feed them. They must produce sustainable food for future generations.

¹² explored using Explainable AI (XAI) in crop recommendation systems. It highlighted XAI's potential to make AI-driven recommendations transparent and reliable. XAI methods, like SHAP and LIME, can help. These methods improve the interpretability of machine learning models in agriculture. These approaches enable farmers to comprehend and trust AI-generated insights, thereby improving decision-making. Also, studies stress the importance of feature importance and visualization techniques, like dependency and summary plots, in explaining model behaviors. Counterfactual explanations, integrated gradients, and layer-wise relevance propagation methods boost these systems' explanatory power. The reviewed literature shows a need for transparent AI tools. They can promote sustainable agriculture and build trust among stakeholders. People must use precision agriculture technology to boost food security. It can optimize crop yields and resource use through data. This paper's focus on a cloud-based TCRM for precision agriculture is very relevant to supercomputing. Advanced machine learning (ML) algorithms and data-driven insights use large, diverse datasets. These algorithms contain environmental and agronomic factors. The Transformative Crop Recommendation Model (TCRM) fills gaps in existing systems. It uses advanced machine and deep learning to provide accurate, tailored crop recommendations. It has a 94% accuracy rate. It can scale for real-time, cloud-based, SMS-enabled farmer assistance. It connects traditional agronomy and modern data-driven insights. It ensures sustainable, context-specific farming practices. To address the challenges and opportunities highlighted, this study proposes the Transformative Crop Recommendation Model (TCRM), designed to bridge gaps in traditional farming with data-driven precision agriculture.

Motivation and contribution

The Transformative Crop Recommendation Model (TCRM) is a novel model that is proposed in this paper. It aims to boost sustainability and productivity in Punjab, India. The region faces soil degradation, water scarcity, and population pressures. Traditional farming and modern solutions, like GMOs (genetically modified organism) and new irrigation, can't meet the demands of food security, climate change, and resource scarcity. The Transformative Crop Recommendation Model (TCRM) uses advanced machine learning and a cloud-based system. It gives real-time, data-driven crop recommendations based on environmental and agronomic factors. TCRM uses IoT sensors, live data, and an SMS alert system. It enables precision agriculture by giving remote farmers insights. It can optimize yields, conserve resources, and promote sustainability. This work helps develop a precision agriculture solution. It uses AI (Artificial Intelligence), IoT (Internet of Things), and cloud computing. The aim is to boost productivity, support sustainable farming, and benefit all in the agricultural value chain. Figure 1 demonstrates the general steps in precision farming, from data collection via IoT sensors to tasks executed by agricultural robots using understandings of the analyzed data. Table 1 sheds light on machine learning algorithms.

Aims and objectives

Empowering precision farming: To empower farmers with data-driven recommendations that enhance the efficiency, sustainability, and productivity of their agricultural practices.

Evaluation of machine learning algorithms: To thoroughly evaluate and compare the effectiveness of machine learning methods for making correct crop recommendations, such as logistic regression, KNN, AdaBoost, and the transformative crop recommendation model (proposed).

Development of a cloud-based crop recommendation platform: To plan and build a cloud-enabled agricultural recommendation tool for farmers.

Integration of real-time data and alerts: To integrate SMS alert generation technology and real-time data sources into the platform to facilitate prompt decision-making.

With these objectives as a guiding framework, the organization of this paper lays out the methodology, results, and discussions systematically to ensure clarity and focus.

Organization of paper

The paper introduces the Transformative Crop Recommendation Model (TCRM). It's a cloud-based platform to improve precision agriculture using machine learning. The study reviews the role of IoT and ML in precision farming. It emphasizes the need for tailored crop recommendations. It then details the methodology. This includes data collection, preprocessing, feature engineering, and model selection. It shows that the TCRM outperforms traditional algorithms on key metrics. The results show TCRM's high accuracy and scalability. It can send personalized SMS alerts, especially to resource-limited areas. Finally, the paper discusses the implications for sustainable agriculture. It outlines future improvements that includes using IoT data and explainable AI for transparency and adoption. Table 2 sheds light on related work in the context of description, methodology,

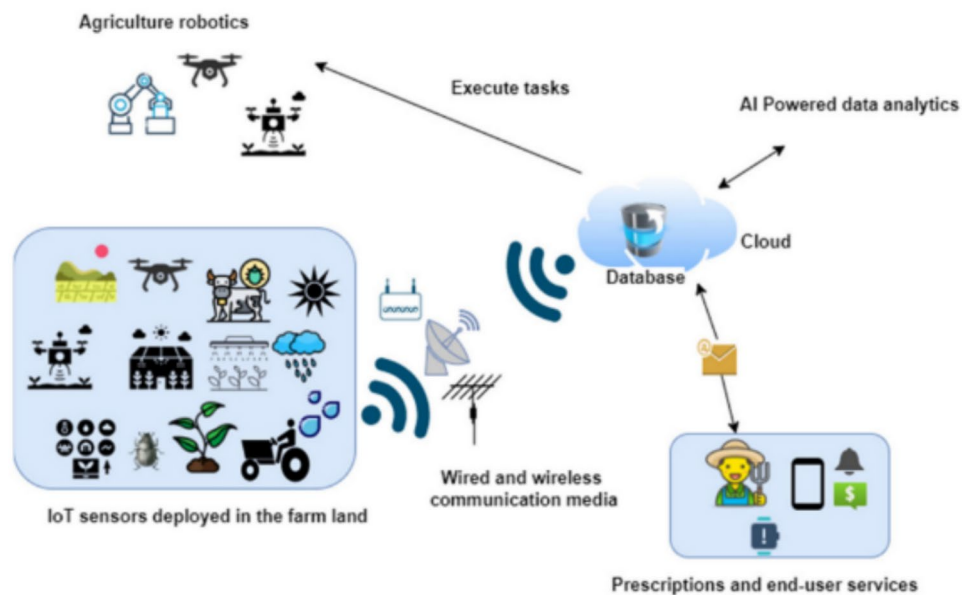


Fig. 1. Steps involved in precision agriculture¹⁰.

findings, benefits, and gaps. The following section delves into the existing body of work, highlighting recent advancements and identifying gaps that this study seeks to address.

Related work

Building upon these findings, this paper introduces a robust methodology for implementing a cloud-based crop recommendation system powered by TCRM.

Role of machine learning in internet of things and precision agriculture in crop recommendation

Machine Learning is key to IoT and precision agriculture. One use case is crop recommendation²⁸. Machine learning provides a single interface to aggregate data from tens of thousands of IoT devices and sensors in remote farms.²⁹ The devices measure many variables, such as soil moisture, temperature, and humidity, as well as crop health. With their integration, ML algorithms help organizations fully understand their data. Feature selection is one of the few machine learning skills. It can identify the key factors that most affect crop yield. They cause the largest variations in changes to gateway-destined crops. Machine learning clears these paths. It helps only with the strongest variables to make recommendations and evaluations. ML algorithms encourage each farmer to decide what crop is to be picked by analyzing previous data related to soil nutrients, climate, and agricultural practices. Recommendations are then adapted to fit local information, like soil type and crop yield potential³⁰. Such algorithms use data on past pest and disease outbreaks, bad weather, and market shifts. This data analysis can help the farmer. It can aid in better, more informed decisions about risks and losses.

A key feature is that ML models can adapt to changing conditions. It is seasonal to provide you some and modify its suggestions. In a dry time, it may suggest drought-tolerant crops or it may recommend crop rotation to improve soil fertility³¹. The model also turns and returns on recommendation systems based on machine learning. The system itself honed its algorithms as farmers then iterated based upon what really happened versus what the original recommendation was. This feedback loop of learning helps to make future suggestions more accurate and relevant. ML-powered crop recommendation systems³² own the traits of scalability and accessibility. This leverages extremely well on its deployable solutions and opens up farmers over the world to the tools². This democratization also allows smaller farmers to have access to data-driven recommendations. Not only this, ML also helps in sustainability and environmentally friendly farming practices.³³ recommended crop rotation, cover cropping, and soil improvement to manage watersheds. This helps to save resources and reduce environmental impact. The above discussion, a brief intro of machine learning and IoT, is revolutionizing precision agriculture by providing advanced crop recommendations. The ML model gives farmers personalized, context-aware recommendations. It helps them make data-driven decisions. This increases crop yields and reduces risks. It also contributes to global agricultural sustainability³⁴. Future food systems could be more resilient, effective, and sustainable due to this tech convergence.

Materials and methods

To enhance crop recommendation accuracy, a cloud-based TCRM (Transformative Crop Recommendation Model) with SMS alert generation system framework is designed and proposed by integrating state-of-the-art deep learning and traditional machine learning algorithms together. TCRM development and application target context-specific, sound predictions and recommendations for agricultural needs. TCRM architecture fuses tree-

Ref	Machine learning algorithm	Description	Applications	Advantages	Limitations
13	Decision tree	A tree-like model applied to regression and classification issues. It splits data into branches based on criteria to make decisions	Customer churn prediction Fraud detection Medical diagnosis Feature selection	Interpretability Handles both categorical and numerical data Minimal data pre-processing	Overemphasis on certain ML models like Random Forest, with less focus on ensemble approaches Narrow dataset coverage, focusing primarily on Indian crops
14	Random forest	An ensemble learning method that combines different decision trees to improve predicting accuracy and reduce overfitting	Image classification Anomaly detection Recommender systems Credit scoring	High accuracy Robust to outliers Reduces overfitting	Limited coverage of real-world implementation of IoT solutions for smart farming Lack of evaluation metrics for energy and resource savings claimed in the framework
15	Support vector machine (SVM)	A binary classification algorithm that determines the best hyperplane for classifying data points	Text classification Image recognition Bioinformatics Financial forecasting	Works well in high-dimensional spaces Uses kernel functions to handle non-linear data Classification based on margins	High computational complexity due to the use of advanced algorithms like KELM and RF Absence of a comparison with simpler and faster baseline models for practical usability
16	K-nearest neighbors (KNN)	A straightforward classification technique that labels features in the feature space according to the dominant class among its k-nearest neighbors	Handwriting recognition Collaborative filtering Anomaly detection Pattern recognition	Easy to implement No model training required Non-parametric approach	Overfitting in gradient boosting and random forest algorithms Absence of comprehensive testing across diverse farm machinery types and environments
17	Naive Bayes	A Bayes' theorem-based probabilistic method that predicts class labels by calculating conditional probabilities	Spam email detection Sentiment analysis Document classification Medical diagnosis	Works well with text data Is effective with high-dimensional data Needs less training data	Insufficient validation of the system in different geographical and climatic conditions Limited exploration of real-time decision-making mechanisms for farmers
18	Gradient boosting	A method of ensemble learning that combines weak learners (often decision trees) to gradually create a powerful predictive model	Ranking in search engines -click-through rate prediction Anomaly detection Object detection	High predictive accuracy Handles imbalanced datasets Reduces bias and variance	Lack of integration with real-time environmental sensors for continuous monitoring Limited exploration of economic factors in crop recommendation
19	Neural networks (deep learning)	A complex brain-inspired model made up of interconnected nodes (neurons) arranged in layers. Deep learning models have multiple hidden layers	Image recognition Natural language processing Speech recognition Autonomous vehicles	Excellent work in challenging assignments Acquires knowledge of hierarchical structures Big data scalable	Dependency on high-quality and diverse datasets for achieving accuracy in ML models Limited application in smaller farming setups compared to large-scale operations
20	Extreme learning machine (ELM)	A learning method for single-hidden layer feedforward neural networks (SLFN) that provides fast training and effective performance. Weights are assigned analytically without iterative tuning	Crop selection Yield prediction Soil fertility assessment Irrigation management Disease detection Price forecasting Smart automation Climate impact analysis	Precision farming Data-driven decisions Scalability Sustainability Adaptability Economic benefits	Limited adoption by farmers Data dependency challenges Overfitting in models Complex algorithm design High computational cost Insufficient real-time data Accuracy vs. explainability trade-off Scalability issues in implementation Dependence on environmental variables Skill gap for technology utilization

Table 1. Machine Learning Algorithms with their Applications and Benefits.

based machine learning algorithms (i.e., random forests and extra trees) with deep learning layers implemented via TensorFlow and Keras. This deep learning component includes layers such as dense layers, normalization layers, multi-head attention mechanisms, and custom to capture these complicated relationships in the data. By utilizing multi-head attention, the model is able to attend to important interactions and dependencies between features, which improves predictive power.

All the parameters are tuned using all the folds of TCRM, which gets better performance and best accuracy for our model. Training-specific hyper-parameters such as the number of epochs, which are usually set to 200, and the batch size of 32, 64, and 128 represent how many training examples to work through before updating the model parameters in iterations. Learning rate is the most well-known hyperparameter that balances oscillations and convergence, typically set at 0.001 or 0.01. In the case of deep learning layers, it may be set to 0.2 or 0.5 because the dropout is used, which randomly sets a fraction of input units to zero. This is done at training time in order to prevent overfitting.

Activation functions like ReLU (Rectified Linear Unit), Sigmoid, and Softmax help in transforming the outputs of each neuron, and optimizers such as Adam, SGD (Stochastic Gradient Descent), and RMSprop minimize loss during training. Tree-based algorithms such as random forests and extra trees have hyperparameters that are fine-tuned. For example, the number of estimators (trees) in a model is commonly between 100 and 500, the maximum depth of trees will be set at 10, 20, or none, and the minimum samples required to split a node can often accept values such as 2, 5, or even 10. In attention-based models, you can mention the number of heads, e.g., 4 or 8. It can also modify the architecture of the model by specifying a few layers (between three and five is common) to represent complex relationships within our data.

Together, these parameters ensure the robust functioning of the TCRM, enabling accurate and tailored crop recommendations. Figure 2 sheds light on a cloud-based transformative crop recommendation model (TCRM), showcasing four phases: data acquisition, preprocessing & analysis, feature engineering & modeling (including tree-based and deep learning models), and deployment & alerting through cloud integration and SMS notifications to farmers.

Citation	Description	Methodology	Key findings	Benefits	Gaps
15	Multimodal ML for crop recommendation and yield prediction	Used equilibrium optimizer (EO), kernel extreme learning machine (KELM), and random forest (RF). Simulations and benchmark evaluations were conducted	Achieved 97.91% accuracy in reliable wireless environments. Highlights ML potential in improving crop management and yield	High accuracy and suitability for reliable wireless environments	Requires testing in resource-limited or real-world agricultural scenarios
21	Crop prediction using IoT and ML	Analyzed data from IoT architectures with 2200 cases and 8 attributes. Utilized decision table classifiers and multilayer perceptron for classification	Achieved 98.22% accuracy. Demonstrated robust model for precision farming using IoT and ML for crop prediction	Effective integration of IoT for crop prediction and fertilizer recommendations	Limited scalability to larger datasets and field conditions
10	ML-based cloud platform for crop recommendation	Compared KNN, DT, RF, XGBoost, and SVM for building a cloud-based recommendation engine	Proposed a cost-free, open-source platform for precision farming technologies to enhance acceptance in agriculture	Provides a scalable and accessible cloud-based solution for farmers	Limited discussion on real-time implementation and adoption barriers
18	Deep learning for crop and water quality assessment	Proposed a deep learning model considering solar exposure, humidity, soil pH, and water quality. Compared with SVM	Achieved 97% and 96% accuracy for crop and water quality predictions. Demonstrated ML potential in advancing agricultural productivity	High accuracy for water and crop quality assessment with multi-feature inputs	Lacks exploration of real-time or dynamic environmental changes
22	Time-series crop yield modeling	Used datasets across fields and years with auxiliary data (e.g., soil conductivity, rainfall, MODIS images). Random Forest models developed for different growth stages	Predicted crop yields using spatially distributed data. Highlighted benefits of integrating time-series data for improving predictive models	Demonstrates the use of spatial and temporal data integration for yield prediction	Limited application to specific crop types and regions
23	IoT-based NPK sensor for soil nutrient monitoring	Designed NPK sensor with fuzzy logic for nutrient deficiency detection. Integrated data into Google Cloud database with a Raspberry Pi 3 prototype	Demonstrated effective nutrient monitoring with alerts for farmers. Highlighted IoT-based solutions for improving soil management practices	Affordable and portable solution for real-time nutrient monitoring	Needs scalability for larger fields and better integration with other agronomic data
24	Hybrid ML for yield prediction	Proposed a two-tier ML model with adaptive k-Nearest Centroid Neighbor Classifier (aKNCN) and Extreme Learning Machine (ELM). Used metrics like RMSE, R ² , and MAE	Improved accuracy and error reduction using hybrid ML models. Showcased IoT's role in data-driven decisions for agricultural profitability	Combines feature selection and ML techniques for improved prediction accuracy	Lack of validation across diverse geographic or climatic conditions
25	Nutrient profiling of mulberry leaf plantations	Employed Extreme Learning Method (ELM) with alternative activation functions. Classified soil traits like potassium, phosphorus, and boron across zones	Provided key insights into soil fertility and nutrient dynamics in Tamil Nadu, contributing to sustainable agricultural management	Highlights regional nutrient dynamics for sustainable agriculture	Limited scalability to non-mulberry crops and other regions
26	Smart irrigation and IoT in agriculture	Developed a smart irrigation system using Agriculture 4.0 concepts, incorporating soil and climate data for decision-making	Presented a cost-effective urban agriculture solution for localized climate monitoring and decision support	Cost-effective system for urban agricultural monitoring	Needs validation in large-scale and rural agricultural environments
27	CNN-RNN framework for crop yield prediction	Combined CNN and RNN for multimodal deep learning. Applied to decades of maize and soybean yields in the USA	Achieved better performance than Random Forest and other models. Demonstrated deep learning's capability in large-scale yield prediction	Shows the strength of combining CNN and RNN for yield prediction in large datasets	Limited applicability for smaller farms and non-cereal crops

Table 2. Summary of related work on machine learning and IoT applications in precision agriculture.

The proposed work is a cloud-based crop recommendation system based on ML algorithms to solve these problems. The platform aggregates, integrates, and analyzes various data streams, such as soil characteristics, climate records, and past crop behavior to recommend the most precise crop for a specific field. The system determines the best-suited crops for particular areas and growing conditions, maximizing their yield potential while minimizing waste. In addition, there is an SMS alert system for real-time recommendations. Farmers can receive timelier data-driven insights even in the remotest places. This feature maintains scalability and continuous updating and development of its predictive capacity as a cloud-based web platform. This solution combines precision agriculture techniques with modern technologies. It does not just help in better utilization of resources and aids sustainable agriculture, but it is also able to continue the production of major staples like wheat, rice, and maize that are crucial for livelihoods in the region. Getting down to Punjab also enables a regionalized approach to data gathering and analysis, which is de facto in the case of actual precision agriculture solutions. By providing information in a localized area, this approach generates valuable data that can be replicated in similar regions of the world to establish a model for sustainable agriculture in densely farmed and resource-limited areas. Algorithm 1 pseudocode provides a concise summary of the methodology adopted in developing the Transformative Crop Recommendation Model (TCRM), detailing the steps from data collection to deployment and continuous improvement.

Analysis of data and methodology

The database collection process, methodologies, and computing tools are covered in the data analysis and methodology.³⁵ designed to accomplish the study's stated objectives (see Fig. 2).

Data gathering

Sources of data used for the purpose and research include the FAO, USDA NASS, and the World Bank data repository³⁶ seeking to carry out the objectives of this study. ESA Data Archives, World Bank Sustainable Development Metrics, NASA Earth Data Archives, CIAT, and GODAN databases are specifically where the properties of the dataset come from. Additionally, reliable data source websites, including Data.gov, UN Comtrade, and Kaggle, were used to acquire data³⁷.

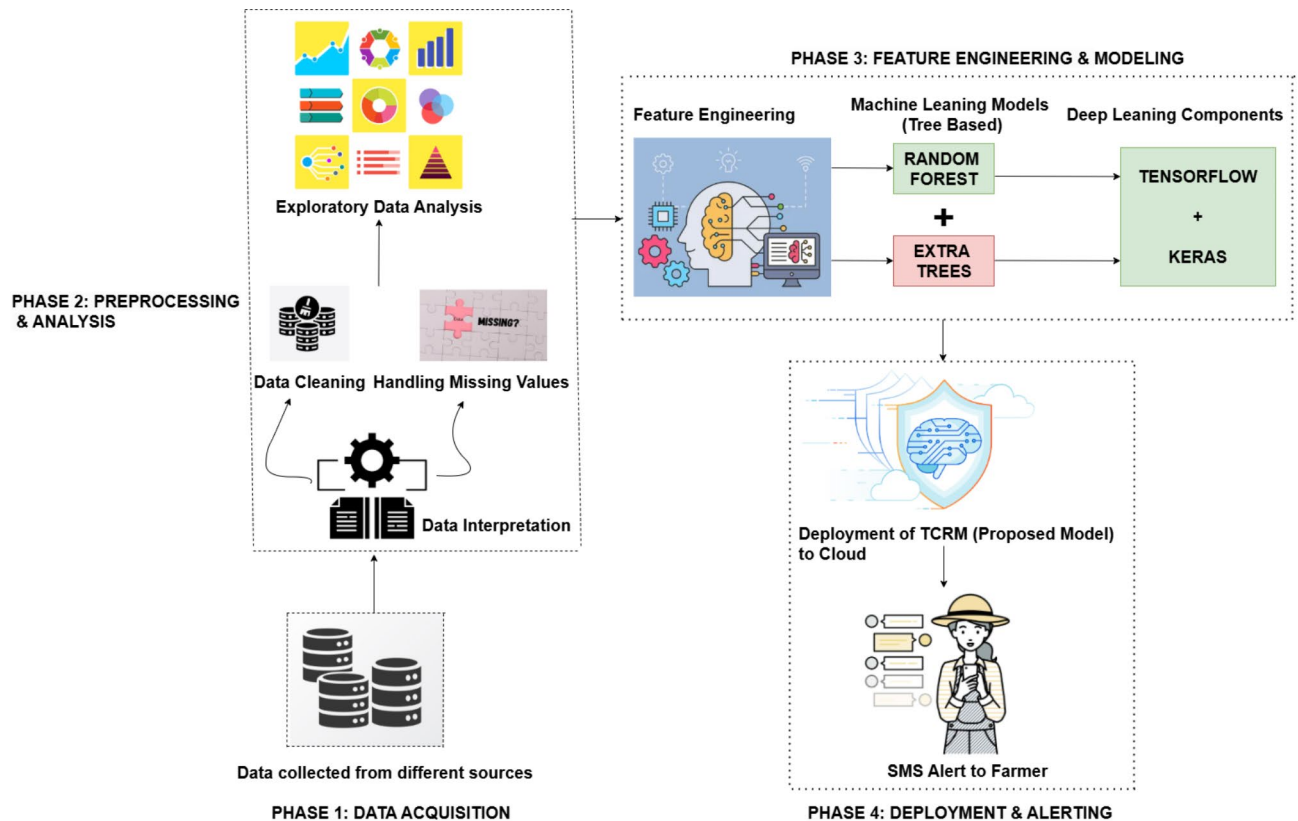


Fig. 2. Cloud-based transformative crop recommendation model.

Data pre-processing

Data received from multiple sources is first combined. The integration of the information makes it easier to manage it effectively and to accomplish the study's stated goals, which include examining the relationships between the crop recommendation indicators. Filtration is part of the preprocessing of data. Many times filtration and purification of data are done as part of cleaning and transforming the data. This process can be applied to the data set, and removing the outliers becomes easier³⁸. Real-world data is sporadic, inconsistent, and scarce. This results in wrong data collection, which in turn leads to making incorrect models using the data. Data preparation tasks are used to resolve these problems by data mining and configure the information in a way that makes it more familiar. As part of the cleaning phase, it tackles data outliers, consistency issues, noisy data standardization, and the missing values restoration. Faulty, insufficient, or wrong data can potentially lead to the validity of study findings at risk. So, it is a necessary step in the data preparation process³⁹. It prepares our data in two different ways before starting further research.

The dataset used in this study comprises 3000 rows and 8 columns, representing soil and environmental factors relevant to crop recommendations. The features include nitrogen (N), phosphorus (P), and potassium (K) content in the soil, along with temperature (°C), humidity (%), soil pH level, annual rainfall (mm), and the target variable such as crop type (label). The dataset contains no missing values, ensuring completeness and reliability for analysis. All features are numeric, except for the categorical crop type column.

To prepare the dataset for machine learning, several preprocessing steps were undertaken. A thorough data quality check confirmed the absence of missing values and validated the consistency of data types. Features were scaled and normalized to standardize their ranges, enhancing model compatibility. The categorical label column was encoded to numeric format using label or one-hot encoding, depending on the model requirements. Outlier detection and handling were performed to ensure robustness, while imbalances in crop type distribution were addressed through resampling techniques such as oversampling or undersampling. Additionally, derived features like nutrient ratios were considered for feature engineering to improve the model's predictive power. These preprocessing steps ensured the dataset's quality and representativeness, laying a strong foundation for effective machine learning analysis. As indicated in Table 3, the gathered information is divided into the following indicators (I1 to I8).

Handling the missing values

The elimination of missing values is one of the pivotal steps in data preprocessing, particularly during mathematical and statistical analysis. Data with missing values can cause bias, errors in results accuracy, and degradation of a model's performance. There are lots of mathematical equations and methods to deal with this

BEGIN

```

# Step 1: Data Collection
COLLECT data from sources (e.g., FAO, USDA, Kaggle, World Bank)
AGGREGATE datasets into a unified format

# Step 2: Data Preprocessing
CLEAN data:
    REMOVE duplicates, outliers, and missing values
    NORMALIZE features using Min-Max Scaling or Z-Score normalization
    ENCODE categorical variables into numeric format if necessary

# Step 3: Exploratory Data Analysis (EDA)
ANALYZE data:
    GENERATE summary statistics
    PLOT visualizations (correlation heatmap, distribution plots)
    IDENTIFY patterns, trends, and relationships

# Step 4: Feature Selection
COMPUTE feature importance using methods (e.g., Pearson Correlation Coefficient)
SELECT the most relevant features for modeling

# Step 5: Model Training
SPLIT data into training and testing subsets
FOR each model in [Logistic Regression, KNN, AdaBoost, TCRM]:
    TRAIN model on training data
    VALIDATE model using testing data
    EVALUATE model performance (accuracy, precision, recall, F1-score)

# Step 6: Deployment as a Local Web App
SERIALIZE the trained model
DEVELOP a web interface (e.g., Flask, Django)
IMPLEMENT APIs for user interaction

# Step 7: Cloud Deployment
UPLOAD model and preprocessing pipelines to a cloud platform (e.g., AWS)
CONFIGURE the environment for scalability
DEPLOY the model as a web API for live predictions

# Step 8: Integration of SMS Alert System
SET UP SMS alert system using APIs (e.g., Twilio)
SEND real-time crop and fertilizer recommendations via SMS

# Step 9: Monitoring and Continuous Improvement
MONITOR model performance with live data
UPDATE model with new data and fine-tune algorithms periodically

# Step 10: Output Delivery
PROVIDE personalized crop recommendations via SMS and web interface
ENSURE recommendations are region-specific and actionable

```

END

Algorithm 1. Pseudocode representation of the methodology for developing the transformative crop recommendation model (TCRM).

issue. Imputation is one of the common methods to work with missing values³⁹. The numbers are substituted with the predicted values based on available data.

Mean imputation is a very simple method of missing variable imputation. The mean imputation equation is given below:

$$\hat{x}_i = 1/n \sum_{j=1}^n x_j \quad (1)$$

In Eq. (1):

\hat{x}_i represents the corresponding imputed value for the omitted data point x_i . n is the total number of data points with non-missing values. x_j represents the values of the known data.

Mean imputation simply assigns the mean (average score) of all the available values for each missing item. While this method is fast and conceptually simple, it is not always the most accurate due to the non-uniform distribution of data. It might want to consider median imputation when the data distribution is skewed or has outliers.

Indicators	Crop recommendation	Description
I1	N	Shows how much nitrogen is present in the soil. Nitrogen is a crucial nutrient for plant growth since it affects the growth of leaves, the synthesis of proteins, and the general health of plants. Crop recommendations and nitrogen-based fertilizer recommendations are influenced by soil nitrogen levels
I2	P	Reflects the presence of phosphorus in the soil. Phosphorus is vital for root development, energy transfer, and flowering. Understanding soil phosphorus content informs crop selection and phosphorous-specific fertilization recommendations
I3	K	Represents the concentration of potassium in the soil. Potassium is critical for enzyme activation, water uptake, and stress resistance in plants. Soil potassium content guides crop selection and potassium fertilizer applications
I4	Temperature	Describes the prevailing temperature conditions in the area, impacting crop growth stages, flowering, and fruiting. Temperature data helps identify crops suitable for specific temperature ranges
I5	Humidity	Indicates the amount of moisture in the atmosphere. Humidity affects crop transpiration rates, disease susceptibility, and water requirements. It plays a role in recommending crops adapted to local humidity conditions
I6	pH	Reflects the acidity or alkalinity of the soil. Soil pH influences nutrient availability to plants, affecting nutrient uptake and utilization. pH data informs crop selection and soil pH adjustment recommendations
I7	Rainfall	The amount of rain that has fallen per unit of time. Crop irrigation and moisture requirements are met by sufficient rainfall. Rainfall is key to recommending crops appropriate for local precipitation patterns. At last, the final advice for the crop type that suits the place most suitable. This is a consequence of an integrated assessment of all indications, such as soil nutrients, climate, and environmental conditions. For the available data, it tries to suggest the best crop in terms of breeding and resource efficiency
I8	Label	This is the super recommendation regarding the advisable type of crop that might be compatible with the certain region. It is a combination of all indicators, including soil nutrients, climate conditions, and environmental factors. According to the information given, can an ideal crop be suggested that optimizes resource needs?

Table 3. Indicators of crop recommendation for precision farming.

$$xi = \text{median}(x_1, x_2, \dots, x_n) \quad (2)$$

In this case, the imputed value will be the median of the data.

Salgado et al.⁴⁰ use regression imputation, where missing values are imputed by fitting a regression model with other variables as predictors. The mathematical formula for regression imputation will vary depending on the type of regression model used (e.g., linear regression, multiple regression).

The way in which the data are to be utilized and how imputation may affect the study must both be carefully considered.⁴¹ proposed a helpful approach, but it should be used judiciously and there should be an understanding of its constraints. Sometimes more complex imputation techniques such as multiple imputation, K-nearest neighbor, or machine learning models may yield better results and accurate imputations.

Data normalization

A key step in preprocessing is data normalization, which tries to bring all the data into some common range⁴². This kind of strategy is mostly necessary while working with or dealing with several unit types and distribution magnitudes. Normalization creates balance in all features so one feature does not dominate the analysis. Standard Scaler is a preferred type of data normalization that scales down the values of the different features, often values being between 0 and 1⁴³. The Min–Max scaling formula is given below:

$$X_{\text{normalized}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (3)$$

In Eq. (3):

$X_{\text{normalized}}$ represents the normalized value of the data point X . X is the original data point. X_{min} is the minimum value of the feature in the dataset. X_{max} is the maximum value of the feature in the dataset.

Scaling the original data points to a range of 0 to 1 makes the smallest size in the dataset represent 0 and the largest size present as 1. This transformation standardizes the data and makes it easier to compare and analyze the values because all data points will be on the same appreciable scale. Other normalization techniques include Z-score normalization (standardization). This method normalizes the features. The normalized data have a zero mean and unit variance.

The mathematical equation for Z-score normalization is:

$$Z = \frac{x - \mu}{\sigma} \quad (4)$$

Here: The standardized (normalized) value is denoted by Z . The initial data point is x . The feature's mean (average) in the dataset is μ . The feature's standard deviation within the dataset is σ . When the data distribution is substantially normal, Z-score normalization is helpful.

The normalization method is based on the criteria for the analysis and features of the data. These methods assist in preventing the biases or distortions from being introduced into the data after preprocessing and help the data fuel different machine learning algorithms and statistical analysis.

Exploratory data analysis (EDA)

In the process of data analysis, a key step is exploratory data analysis (EDA), where the aim is to understand and summarize the main features of a dataset. The mean is a key ingredient of centrally tended measures and is an important mathematical relation involved in EDA⁴⁴, although this fact is mostly obvious with the use of visualizations and descriptive statistics. The mean (μ) of a dataset is:

$$\mu = 1/n \sum_{i=1}^n x_i \quad (5)$$

There are a number of data points (n), and between and indicates the number of each unique data point. It gives the value of the average in the numbers; it works to understand the basic or core value of your data set. EDA uses the mean, along with different visualizations, to look for patterns, trends, and outliers within your data. In addition, this full cycle enables guided decisions about data preparation, feature selection, and statistical analysis using high-level information.

Feature selection

Feature selection in data analysis and machine learning is the process of selecting those features that are more informative or useful for predictive or classification tasks⁴⁵. Earlier, it explored a subset of features to meet some criteria in ML. However, the feature selection is defined as choosing the best feature from all available features. Common mathematical techniques for feature selection are the Pearson correlation coefficient r (correlation coefficients). The formula for Pearson correlation is given below:

$$r = \frac{\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)}{(n-1)\sigma_x\sigma_y} \quad (6)$$

where r is the linear correlation between x and y , μ_x & μ_y are standard deviations of variables x and y , and \bar{x} , \bar{y} are averages of variables x and y , respectively. One is to make a matrix of correlation between each feature and other features and the target variable as well to find out which are the features that can be a major contribution towards my solution. Possible removal of parameters that induce multicollinearity are weakly correlated during feature selection. This approach is complemented by other mathematical tools, for instance, statistical analyses and machine learning algorithms, which allow the selection of the most useful features to enhance model interpretability and performance⁴⁶.

Machine learning approach

This includes a systematic approach to creating and testing models for prediction, selecting the model with the best performance. It is a strategy that requires machine learning to reach research aims with artificial intelligence powers. The dataset is split into subsets of train and test, which will make it easier to evaluate the model. Subsequently, various machine learning algorithms are applied to develop prediction models with their respective parameters⁴⁷. Those models need to be trained by the training dataset before they can grade their performance through the testing dataset. Thilakarathne et al.¹⁰ assess model efficacy with metrics such as accuracy, precision, recall, and F1-score as well as cross-validation scores. Researchers often play with these hyperparameters to maximize the prediction power of their models. Through this iterative process, the model delivering the best performance will eventually surface. For example, the result of selecting the best model is based on your own study goals while balancing characteristics of models such as complexity, interpretability, and performance of a model on X (functionality) to get good accuracy in function Y (target). To increase the accuracy of prediction, a few assembling methods like bagging and boosting⁴⁸ may also account for these. In general, this machine learning technique⁴⁹ helps in better data-driven decision-making. It helps enable making your research results more effective as well as act as a basis for all the insights and recommendations.

Deploy the model as a web app to a local server

Deploying a machine learning model as a web application on the local server makes the developer-friendly version of the same model. Step 1 is one in a chain of steps to serialize the trained model, which means converting it into any easy-saving and known format. Next, this web-based framework is deployed on the backbone of the web framework to construct a web application⁵⁰. Designing the user interface for an engaging and fluid user experience that is created using a combination of HTML, CSS, and JavaScript. The API exposes the ML model, which runs serverless on AWS Lambda, using scalable and often free processing power. Finally, this deployment gives you the ability to go from a simple base ML algorithm configuration to an interactive live front-end tool able to serve predictions or knowledge in real time through a user-friendly online face, greatly enhancing their already accessible utility for more use-cases.

Deploy machine learning model to cloud

A machine learning model has been pushed into the cloud to provide recommendations on crops. It is a strategic move hinging on their Amazon Web Services to abstract the cloud's scalability, availability, and processing power. This deployment process includes a few important steps: The trained model is serialized and prepared to be used. Then, a deployment environment has been chosen in the cloud, be it Google Cloud Platform (GCP), Amazon Web Services (AWS), or Microsoft Azure. The model is in such cases, shipped to the cloud server together with all requisite data pretreatment and post processing pipelines⁵¹. Once the model is deployed, it exposes a web API that consumers can use to interact with the model using other applications like web or mobile apps. For example, a user will input local information like soil qualities or weather conditions into the system and receive instant crop suggestions based on the model predictions. Some of the benefits provided by cloud usage include: unlimited user accommodation, security features, and opportunities for very resource-intensive tasks to be executed quickly and efficiently⁵². On the other hand, cloud-based solutions allow for a far simpler

update and maintenance of systems, as well as ensuring that the crop recommendation model remains correct through time. In common, transferring device finding out fashions into the cloud complements their availability and provides a better person interface that may make crop suggestions information pushed to reinforce farmers and different agricultural stakeholders.

SMS alert generation system

Using Twilio as a bureau to create an SMS alerting system, many applications and services gain a valuable notification/alert layer. One of these platforms that is providing great support for programmatically sending SMS messages in a cloud-based way comes from one of the most robust Twilio. ²³ used Twilio's APIs and services to send out SMS alerts using this system to specific recipients for various reasons or situations. This technique is quite flexible, as it can be applied in many scenarios and businesses. For example, to keep farmers informed about the weather and when their field requires watering or if a disease has broken out on the crops. For healthcare, it can be handled to push emergency notifications or for medication schedules and appointment reminders.

The same can be used in logistics and eCommerce to notify customers about the delivery status of packages, offers on promotions, or status updates related to orders⁵³. This typically includes configuring the provider system that will send messages automatically with event triggers, setting up Twilio with its API and credentials required for authentication, as well as defining what SMS alert body you want to be sent. An SMS Alert Solution is stable and automated to promote communication, reduce Time-To-Action response periods And notify the users in real-time for a better user experience and improved operating efficiency. The methodology outlined has been rigorously applied to evaluate the TCRM, and the results offer valuable insights into its effectiveness and scalability.

Results and discussion

The different measures used to evaluate the performance of these algorithms include accuracy, precision, recall, f1 score, and K-fold cross-validation scores, which provide some insights on their performance and ability to generalize. TCRM (our proposed method) early shows good performance based on key evaluation metrics compared to baseline algorithms such as logistic regression, K-nearest neighbors, and AdaBoost. In particular, this method obtains precision, recall, and F1-score of 94.46%, 94.462%, and 93.97%, respectively. An impressive fivefold cross-validation score of 97.67% is also obtained. This exhibits a clear improvement over the performance of traditional algorithms when understood in statistics quantitatively. These metrics are compared to the relatively lower performance of traditional algorithms, highlighting a clear improvement. However, establishing true novelty in a machine learning context requires more than just superior performance. It may include unique data preprocessing methods, unique architecture designs, or new model improvements that set approaches apart from the existing ones. Moreover, the clear contrast with AdaBoost is that significantly lower scores require careful analysis, providing insights into why your method excels where traditional methods struggle, thereby further emphasizing its impact and innovative nature.

Based on these results, the TCRM (proposed method) model is opted to use in our model. It has presented the highest level of accuracy and generalization capabilities due to disappointing results with a fivefold cross-validation score compared to other models. It is required to deliver the most accurate and dependable crop recommendations to farmers. Moreover, the additional tuning or feature engineering is used. This seems to be a scope for performing more on the classification problem. Our deployed cloud-based platform with the TCRM (proposed method) model is a significant step towards the democratization of precision farming. The platform that connects them includes real-time access to customized crop recommendations, backed by an SMS alert system. These functionalities are most important in resource-limiting regions as they enable farmers to utilize their resources efficiently, reducing waste and increasing the yield of crops.

The transformative crop recommendation model (TCRM) was selected as the final model due to several key advantages that make it highly suited for agricultural datasets and real-time decision-making. TCRM showed an appropriate trade-off between efficiency and accuracy that looked to be well-suited for the rapid analysis of massive agricultural data, allowing quick turnarounds on recommendations. Models such as AdaBoost, K-Nearest Neighbors (KNN), & logistic regression were able to achieve comparable performance, but TCRM had the important benefits of speed (quick processing times are needed for an application in agriculture).

Additionally, the interpretability of TCRM makes it easy to know how different features affecting soil nutrition, weather data, and crop requirements lead to the final recommendations. Such interpretability is important in the context of agriculture since it can help provide farmers and agronomists with meaningful insights to inform their decisions. Moreover, due to the high performance of TCRM, it is able to outstand well with the extreme changes often seen in different agricultural data that reflect varying environmental conditions. The benefits of the crop recommendation model are speed, interpretability, and robustness. It serves precision agriculture well by providing accurate recommendations in a timely fashion to minimize loss of yield or nutritional quality for crops during their critical window between planting and harvesting, where optimal growth must be achieved.

Secondly, the non-field affect of this platform extends into developments that other parts of agriculture can leverage and already enables policy development. It promotes the efficient use of resources and reduces wastage of inputs, contributing to global food security and environmental sustainability while encouraging more sustainable farming by doing so. This data-driven approach has immediate socioeconomic benefits for farmers, providing them with better predictability on yields and policymakers the ability to work with an evidence base to develop sustainability-focused policies aimed at the agriculture sector.

Moreover, the system is capable of being generalized with other datasets. It has good potential as the basis for precision farming research in the future. This work provides applicable solutions and a lasting pathway to address agricultural resilience for present-day impact to farmers, scientific knowledge, as well as policymakers.

Research Objective I: To empower farmers with data-driven recommendations that enhance the efficiency, sustainability, and productivity of their agricultural practices.

To achieve Research Objective I to provide farmers with data-backed recommendations for increased efficiency, sustainability, and productivity of their agricultural practices, the research paper uses a rich dataset consisting of eight important columns. The columns are nitrogen, phosphorus, potassium, temperature, humidity, pH, rainfall, and label. This dataset provides the basic infrastructure for our exploration, allowing us to reach several goals. Again, nutrient levels (N, P, and K) being presented separately in the data set acknowledges soil fertility as crucial to crop cultivation. Nitrogen, phosphorus, and potassium are essential major nutrients for plant growth. The accurate determination of these are critical to optimizing fertilizer recommendations that provide sufficient nutrient content in India. The distribution of features as shown in Fig. 3 in order to determine the type of data count per indication. The correlation matrix reflecting the correlation between various features on the dataset was then presented as a correlation heatmap, as seen in Fig. 4.

The correlation heatmap analysis offers valuable insights into the relationships among the various features within the dataset. One notable observation is the presence of strong positive correlations, where a few features have a correlation score that is near 1. When one increases, the other follows suit, and when one decreases, the other exhibits a corresponding decrease. Such strong positive correlations are indicative of a direct and often predictable relationship between these features.

Conversely, almost all features in the dataset have very low and negative correlations; the correlation values of most features are around zero or below it. Weak correlations indicate that the features are not linearly related, meaning changes in one do not always mean corresponding changes with another feature. Given that the correlations tend to be negative—that is, as one attribute grows, another seems more likely to shrink—it may lead you inevitably towards questions about trade-offs among these qualities.

Visualization of ease can classify the samples in less than 30 member class labels with our crop recommendation model confusion matrix as shown on Fig. 5. A confusion matrix is an important tool to evaluate the performance of a machine learning model, in particular for classification problems. To further optimize the parameters of a model for better predicting crop recommendations to farmers and other agricultural stakeholders, it is crucial that we have this data. It is a good tool to determine the performance of the model and provide direction towards improving its efficiency in tasks related to crop classification and recommendation. As indicated by attributes of this dataset, the soil N, P, and K levels assume an essential role from a biological viewpoint since they are the principle macronutrients that plants take amid the development stage. In general, these macronutrients have the greatest input into subcategories:

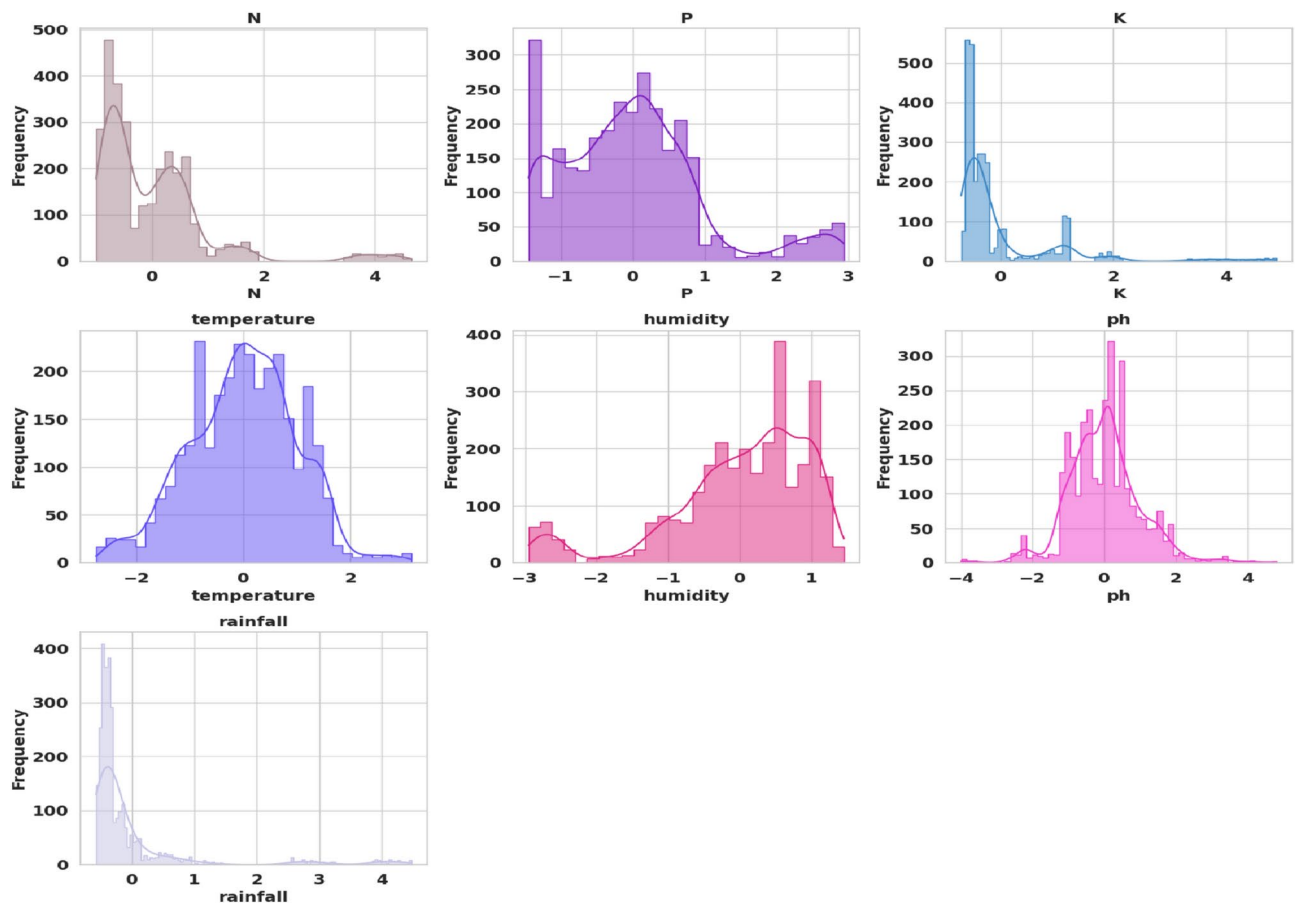


Fig. 3. Count per indicator (I1 to I7) in the dataset.

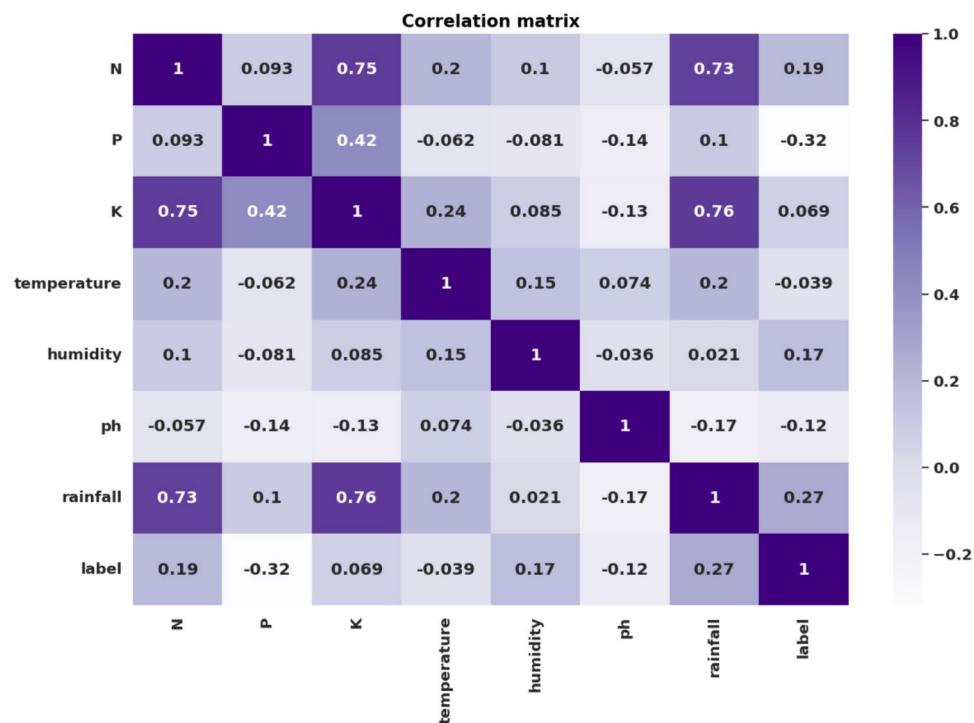


Fig. 4. Correlation matrix displaying the relationship between the dataset's various aspects.

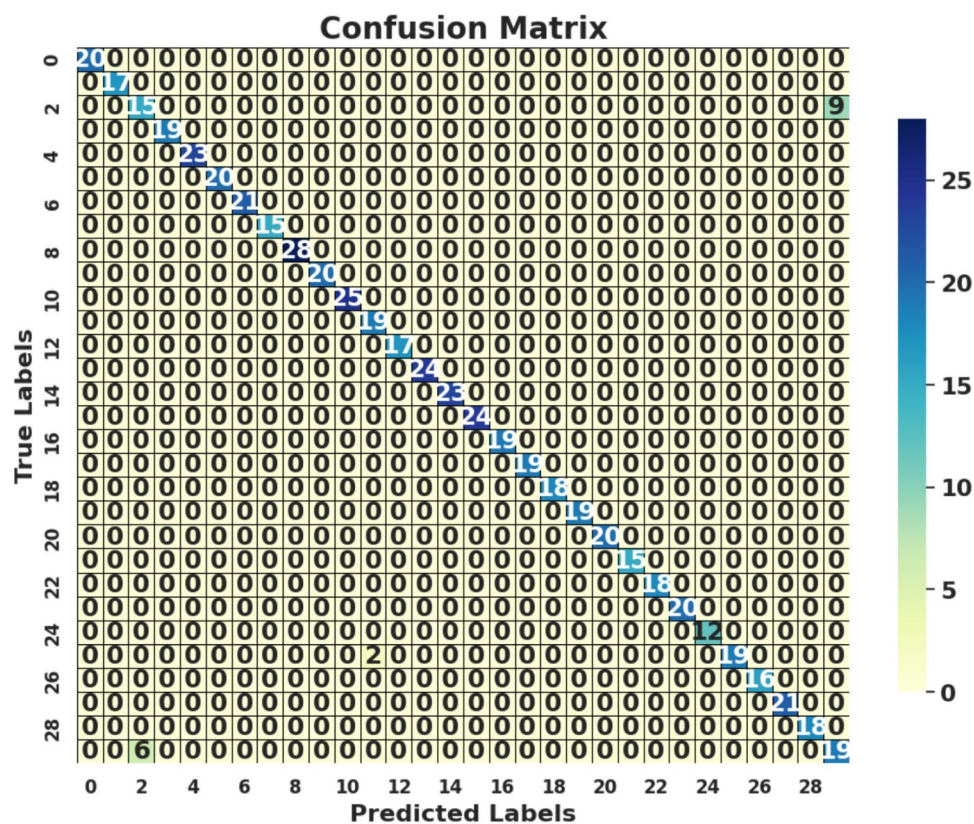


Fig. 5. Confusion matrix.

N—Nitrogen promotes the growth of plant leaves.

P—Phosphorus is mostly needed for the growth of roots, flowers, and fruits.

K—Potassium is important for being able to carry out the general functions of a plant properly.

Fertilizer can supply these macronutrients, and different kinds of crops will benefit from it more depending on the N, P, and K concentrations in the fertilizer. In addition, a lot of N, P, and K are required for crops to grow and thrive, and strong, nourished plants bear more fruit. But in the absence of fertilizer, the soil might not contain enough nutrients to encourage healthy growth. By controlling fertilization, farmers may promote optimal plant development and productivity by understanding the NPK ratios that crops require. The nutrients that the soil needs are delivered to the soil through fertilizer. An analysis of the N, P, and K requirements for various crops in our dataset is shown in Fig. 6. Based on our data in the dataset, it is clear that sugarcane, potatoes, and dates require high nitrogen levels, apples and grapes require high phosphorus levels, and sugarcane, dates, apples, grapes, and litchi require high potassium levels in comparison to all other crops.

Furthermore, environmental factors such as temperature, humidity, pH levels, and rainfall play a significant role in determining crop suitability and performance. These all affect crop impact, disease infection, and yield. The research paper builds on the collection of information relating to these parameters to provide those holistic insights with regards to crop recommendations, ensuring that the climatic and soil conditions were in coherence with one another. The remaining features from this dataset are rainfall, soil pH, air temperature, and air humidity.

The pH of the soil affects the presence of microorganisms, harmful chemicals, and important nutrients. Plants require rainwater to survive on their nature. A plant may require different amounts of water for photosynthesis to occur; the air must be at a certain temperature; when the temperature rises, photosynthesis may also increase; and air humidity is necessary for plant transpiration. For example, when the temperature rises, these characteristics also help determine which crop to harvest. When considered collectively, each of these characteristics is crucial in deciding which crop should be harvested. As the recommended crop type based on the study of all the aforementioned parameters, the 'label' column in the dataset is of utmost significance. The final result of the study is this proposal which can help farmers choose the right crops, plant them, and choose the right management techniques. In essence, this comprehensive and multidimensional information is the foundation of the study, allowing for data-driven analyses and suggestions that enable farmers to improve their farming methods, increase crop yields, and promote sustainability in contemporary farming.

Research Objective II: To thoroughly evaluate and compare the effectiveness of machine learning methods for making correct crop recommendations, such as logistic regression, KNN, AdaBoost, and TCRM (proposed model) (Table 4).

Python programming was used to develop a collection of predictive machine learning models in order to fulfill Research Objective II, which calls for a thorough evaluation and comparison of the efficacy of various machine learning algorithms for making correct crop predictions. These are the models that we built with lots of love on top of a large and diverse dataset consisting of parameters like N-P-K nutrient levels, temperature-humidity-pH-rainfall values, along with the important "label" column, which tells us what crop to recommend. For each method, the performance was thoroughly analyzed and benchmarked with respect to several different parameters like scores of F1-score, recall, accuracy, precision, and fivefold cross-validation. The scores obtained for F1, accuracy, precision, recall, and five fold cross-validation are shown in Table 4 for easier comparison.

Figure 7 sheds light on accuracy achieved by various classifiers. The proposed transformative crop recommendation model (TCRM) achieved the highest accuracy among all compared classifiers, reaching an accuracy rate of 94%. This outperformed K-Nearest Neighbors (KNN) and Logistic Regression, which scored 92.84% and 91.16%, respectively. AdaBoost lagged significantly behind with only an 11.50% accuracy rate,

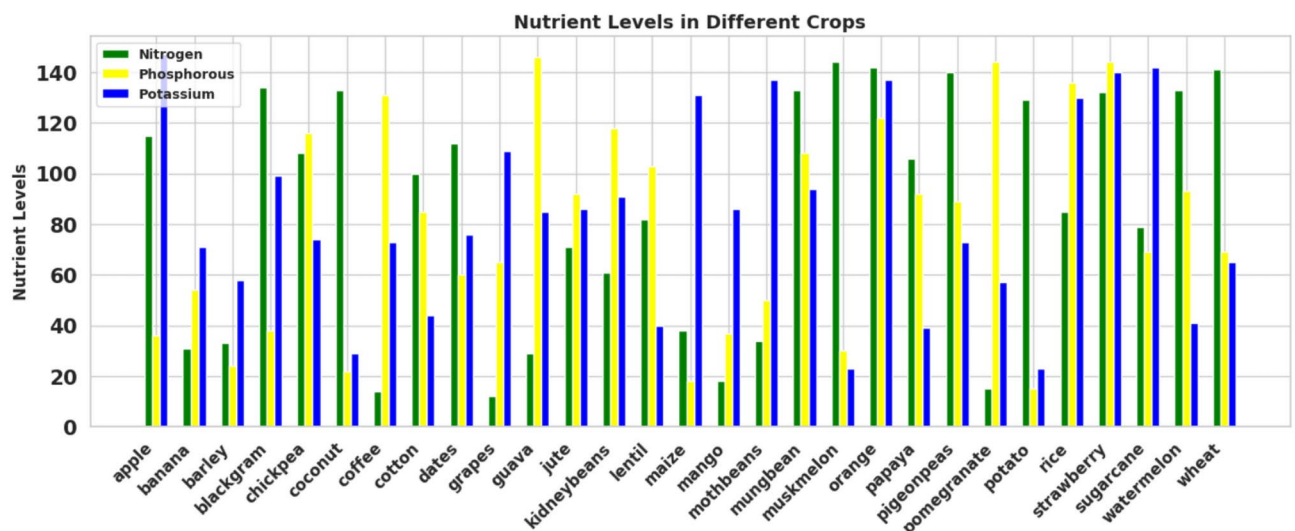


Fig. 6. Comparison of NPK requirements of different crops.

Algorithm	Accuracy	Precision	Recall	F1-score	5-Fold CV Score
Logistic regression	91.17%	91.44%	91.17%	90.99%	84.50%
K-nearest neighbors	92.83%	93.11%	92.83%	92.74%	83.67%
AdaBoost	11.50%	7.92%	11.50%	8.45%	95.00%
TCRM (proposed method)	94.00%	94.46%	94.00%	93.97%	97.67%

Table 4. Scores for accuracy, precision, recall, F1-score, and fivefold cross validation. Significant values are in bold.

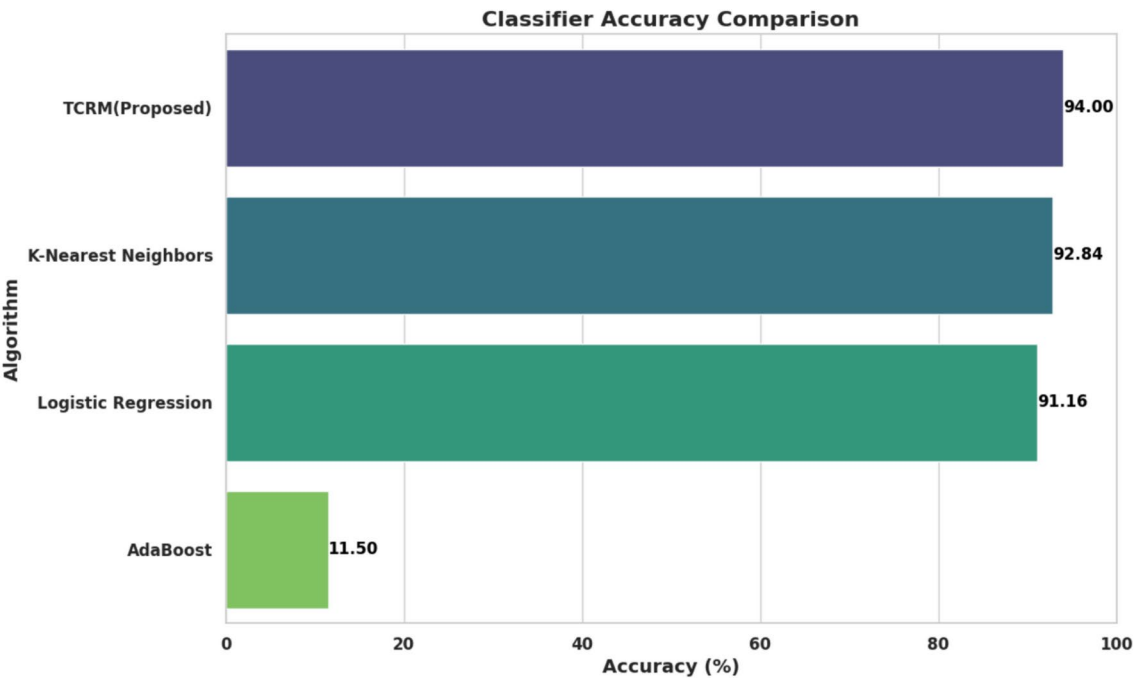


Fig. 7. Accuracy comparison of machine learning algorithms.

indicating that it is less suitable for this application. The high accuracy of TCRM highlights its strength in making precise and reliable predictions tailored to agricultural data.

In terms of precision, which measures the correctness of positive predictions. In Fig. 8, TCRM led with a score of 94.46%. KNN followed closely with a precision of 93.11%, and logistic regression scored 91.44%. AdaBoost demonstrated a much lower precision at 7.92%. The high precision of TCRM indicates its effectiveness in reducing false positives, which is crucial for accurate recommendations in agriculture, minimizing unnecessary or incorrect interventions.

Recall measures the model’s ability to identify all relevant instances correctly. Figure 9 represents recall achieved by various classifiers. TCRM excelled with a recall score of 94%, showing its proficiency in capturing true positives. KNN and Logistic Regression performed well with scores of 92.83% and 91.17%, respectively, but AdaBoost fell short with a recall rate of only 11.50%. TCRM’s strong recall makes it particularly effective for ensuring critical cases in agricultural recommendations are not missed.

For the F1-score, which balances precision and recall. Figure 10 shows TCRM emerged as the top performer with a score of 93.97%. KNN followed with an F1-score of 92.74%, and logistic regression scored 90.99%. The significantly lower F1-score of AdaBoost at 8.45% highlights its poor balance between precision and recall compared to the other models. The superior F1-score of TCRM further emphasizes its consistent reliability and balanced performance.

In Fig. 11, a fivefold cross-validation test of TCRM demonstrated its robustness and consistency with a score of 97.67%. This was notably higher than KNN (92.74%) and logistic regression (90.99%), while AdaBoost maintained a low score of 8.45%. The strong performance of TCRM across multiple folds underscores its ability to generalize well and maintain stability when presented with varied data, making it ideal for the dynamic conditions of precision agriculture. TCRM excels in all the parameters used such as accuracy, precision, recall, f1-score, and Cross Validation with fivefold.

Research Objective III: To design and construct a cloud-enabled platform for recommending crops to farmers.

This deployment of the TCRM (proposed model) was a significant step in reaching one of the research objectives, Research Objective III to be precise, which involves building and implementing a cloud-based

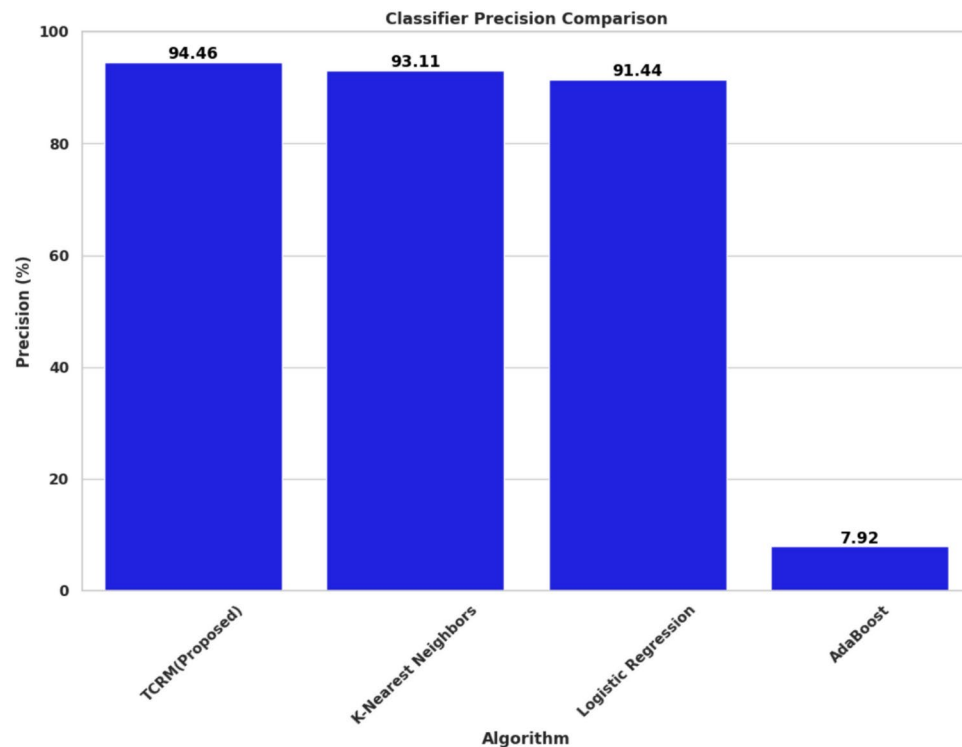


Fig. 8. Precision comparison of machine learning algorithms.

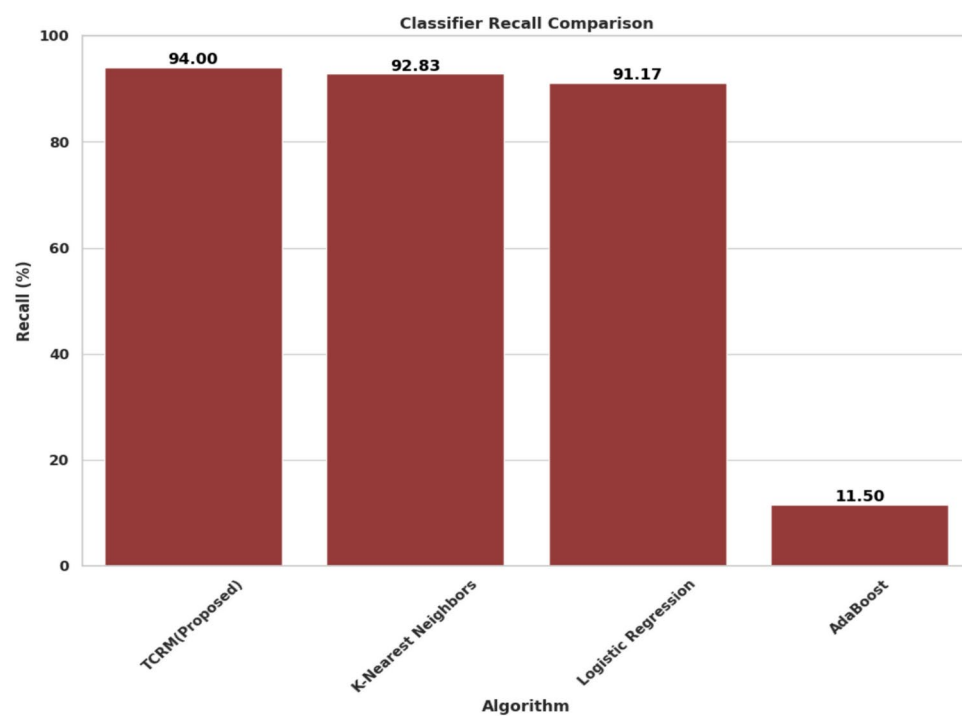


Fig. 9. Recall comparison of machine learning algorithms.

platform for crop recommendation service personnel. This highly sophisticated predictive machine learning model is maintained with a strict curation process to provide accurate and robust crop recommendations. Figure 12 depicts the cloud-hosted crop recommendation interface. In order to provide this technology in an easy-to-access and user-friendly manner, the TCRM (proposed model) was deployed on a cloud architecture (specifically Amazon Web Services (AWS) servers). Because AWS offers a secure and flexible cloud computing

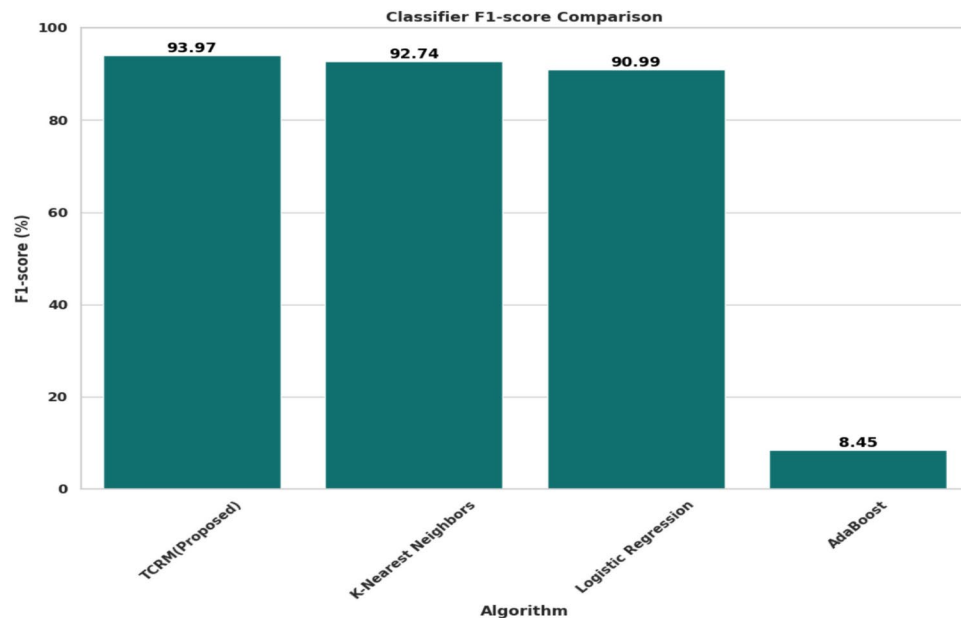


Fig. 10. F1-score comparison of machine learning algorithms.

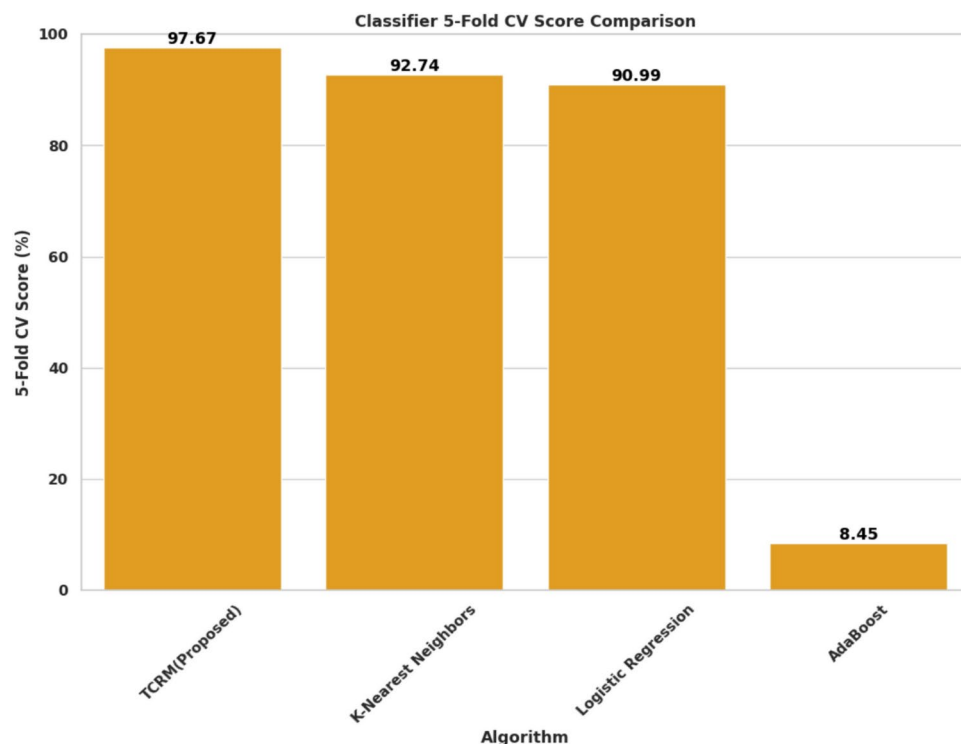


Fig. 11. 5 fold cross validation comparison of machine learning algorithms.

environment, it is an ideal platform for web application or model hosting. Figure 13 shows crop and fertilizer recommendations for a given crop displayed in the cloud-based crop recommendation interface. During deployment, the TCRM (proposed model) needed to be exposed as a web API. This especially allows farmers who can feed soil properties, climatic regions, location, etc. via the interface of this API into the system. The deployed TCRM in turn responds with a prediction that can be based on the specifics of this instance to present appropriate crop suggestions, again providing personalized help and solutions.

With the hosting platform provided by AWS, it ensures that this research project maintains scalability and availability for more users along with data inputs to be processed in our crop recommender system. That provides

Crop Recommendation By Using Machine Learning

Note: N, P & K are in grams per Hectare, Temperature is in Degree Celsius, Humidity in Percentage (%) & Rainfall in Millimeter(mm).

Nitrogen:

Phosphorus:

Potassium:

Temperature:

Humidity:

Ph:

Rainfall:

Fig. 12. Crop recommendation interface hosted in the cloud.

Crop Recommendation By Using Machine Learning

Note: N, P & K are in grams per Hectare, Temperature is in Degree Celsius, Humidity in Percentage (%) & Rainfall in Millimeter(mm).

Nitrogen: 40

Phosphorus: 23

Potassium: 45

Temperature: 37

Humidity: 70

Ph: 6

Rainfall: 35

Soil is fit to grow Pomegranate

Fertilizers Recommendation: You can use Urea, Single Superphosphate, Muriate Of Potash, Zinc Sulfate and Boron for good production of pomegranates.

SMS sent to Farmer's phone number

Fig. 13. Crop recommendation interface hosted in the cloud with recommendations.

farmers with an easy way to use data to make better decisions that will help drive efficiency into agriculture well beyond many other techniques currently in practice today.

Research Objective IV: To integrate SMS alert generation technology and real-time data sources into the platform to facilitate prompt decision-making.

This intends to power the functions of an agriculture platform by integrating live data sources (weather forecasting services) and triggering SMS alerts for real-time assistance in decision-making farm-level suggestions. As an important part of this objective, a system for sending SMS alerts has been added to the platform in order to deliver vital information straight into farmers' mobile phones. The system is aimed at sending SMS alerts containing crop recommendations and fertilizer recommendations specific to the recommended crop.

Using real-time data sources like weather, soil quality, and other ecosystem factors to identify the optimal crop for that field as per season based on our platform algorithms. It can be made to trigger an SMS alert generation system to farmers about a favorable crop on their land and the type of fertilizer they need for its maximum yield in Fig. 14. This real-time tailored assistance provides farmers with insights that they can act upon, allowing for informed and rapid decision-making to adapt to changing weather phenomena and optimize the productivity and sustainability of their agricultural practices.

Comparison with the existing work

Table 5 highlights various methodologies, including neural networks, support vector machines, and deep learning architectures like CNNs. The datasets range from satellite imagery and weather data to field-specific soil reports and crop-specific RGB images. The studies show diverse applications, such as soybean yield forecasting, wheat prediction, and soil nutrient classification, with accuracies ranging from 61 to 94%, showcasing the evolution and advancement in precision agriculture methods.

Figure 15 visually compares the accuracy of these studies, emphasizing the performance of the proposed TCRM (Transformative Crop Recommendation Model), which achieved the highest accuracy of 94%. This indicates the robustness of TCRM in providing personalized, data-driven crop recommendations using advanced ML/DL approaches and real-time data integration.

11:28 AM

Sent from your Twilio trial account

- Soil is suitable for growing
Pomegranate.Fertilizers Recommendation: You
can use Urea, Single
Superphosphate, Muriate Of Potash,
Zinc Sulfate and Boron for good
production of pomegranates..Thankyou For Using Our Crop
Recommendation System..**Fig. 14.** SMS sample to farmer showing crop and fertilizer recommendation.

The findings not only validate the efficacy of TCRM but also pave the way for future innovations, as discussed in the concluding section.

Conclusion and future scope

In this work, the Transformative Crop Recommendation Model (TCRM) is presented that combines scientific machine learning and deep learning to provide temporal, context-aware, and scalable crop recommendations in the era of precision agriculture. By deploying the model on a cloud platform and using SMS alerts to notify farmers regarding the content of the data-driven insights, remote farmers can be served with better information for their farming activities. By leveraging multi-head attention mechanisms, tree-based algorithms, and scalable cloud infrastructure. It not only augments predictive power but also helps in connecting traditional agronomy with modern data-driven approaches. By focusing on data preprocessing, normalization, and feature selection. It provides robustness, scalability, and generalizability to different agricultural settings and input data. TCRM embodies the transformative potential of precision agriculture to optimize yield efficiency and food security by emphasizing sustainable agronomy, resource use, and resilience to stressors such as climate change and water limitation. Punjab, India. The Punjab case study demonstrates the ability of TCRM-driven precision agriculture to enhance water and soil health, optimize crop sectors and realize yield potential under resource-constrained conditions.

In the future, TCRM could be improved by developing region-specific models that include climate, soil, and crop data on a local scale in order to increase its prediction abilities and relevance for different agricultural contexts. Using IoT-based monitoring devices to collect real-time data can be implemented as an extension for accurate and timely recommendations. The addition of proper sustainability measures, including but not limited to measures on water use, carbon footprint, and resource efficiency, will complement the broadband delivery of environmentally friendly farm practices. Multilingual interfaces and mobile applications can make the platform easier to use for farmers in multiple demographics. Utilization of state-of-the-art data augmentation strategies and steady updates to the model will enable TCRM to be proactive with respect to emerging trends and challenges in agriculture. A feedback loop where farmers can provide feedback on suggestions would refine the model even more, making it smarter and more human-centric. Working with public officials, researchers, and extension services can help ensure the expansion and fine-tuning of TCRM for widespread impact. The model can also be fortified by looking at new technologies such as blockchain for secure data sharing and AI-based disease prediction models, ensuring it remains an invaluable instrument based on data to potentially revolutionize agriculture, improve food security, and promote sustainable farming globally. Future work could integrate explainability techniques like SHAP and LIME to make crop recommendation models more transparent. These methods provide insights into feature influence, fostering trust and informed decision-making among agricultural stakeholders.

Ref	Study	Objective	Methodology	Dataset	Key findings	Accuracy
54	Satellite-based soybean yield forecast	Real-time soybean yield prediction in Brazil	LSTM, random forest, OLS regression	Satellite imagery and weather data	LSTM outperformed others; MAE ranged 0.24–0.42 Mg/ha depending on timing	72.74%
55	Wheat yield prediction	Prediction of wheat crop yield using MLP neural networks	Multilayer perceptron with custom activation functions	Weather datasets, 1990–2017	New activation functions improved prediction accuracy compared to Sigmoid functions	90%
56	Soil nutrient classification	Classification of soil fertility indices and pH levels	Extreme learning machine (ELM)	Soil test reports from Kerala, India	Gaussian radial basis function was the most effective, with accuracy for pH classification	90%
57	Soil mineral classification	Ontology-based knowledge base for soil mineral composition	Ontology mining	Expert knowledge and soil datasets	Created structured knowledge for soil composition aiding crop recommendations	91%
58	Nutrient deficiency detection in plants	Detection and classification of nutrient deficiencies in tomato plants	Machine Learning (ANN, image processing)	Tomato plant datasets	Proposed ANN model effectively identified deficiencies using leaf characteristics with high accuracy	88.27%
59	Classification of crop based on macronutrients and weather data	To classify crops (rice, wheat, sugarcane) based on macronutrients and weather data	Support vector machine (SVM) and decision tree	Chhattisgarh state data over 10 years including soil type, rainfall, and temperature	SVM and decision tree algorithms achieved high accuracy in classifying crop types	92%
60	Rice nitrogen nutrition estimation using RGB images	To estimate nitrogen nutrition in rice using RGB images and machine learning	Random forest (RF), backpropagation neural network (BPNN), and simple nonlinear regression (SNR)	RGB images from field experiments over two years	RF performed best with 81% accuracy for nitrogen accumulation estimation, showing strong generalization performance	80.47%
61	Wheat yield prediction using machine learning	To predict wheat yield using advanced sensing techniques	Self-organizing maps (SOMs), counter-propagation ANN, and XY-fusion networks	Data from 22 ha field in Bedfordshire, UK, including NDVI and soil properties	SOMs with supervised learning had 81% overall accuracy, showing effective integration of soil and crop data for yield prediction	81.65%
62	Two-tiered machine learning model for crop yield prediction	To predict paddy yield using nutrient deficiency analysis	Neural networks and K-means clustering with TensorFlow	Images of paddy fields, classified into nutrient deficiencies	Achieved 76%–77% accuracy, effectively combining neural networks and clustering for yield prediction	77%
63	Deep learning classification of land cover and crop types using remote sensing data	To classify land cover and crop types using remote sensing data	Four-level DL architecture with CNNs, MLP, and RF, including preprocessing, supervised classification, and geospatial analysis	Landsat-8 and Sentinel-1A images for Kyiv, Ukraine, 2015, with 11 classes	2-D CNN ensemble achieved the highest accuracy (94.6%), with better discrimination of maize and soybeans	85%
17	Intelligent crop recommendation system	To assist farmers in selecting suitable crops based on environmental and economic factors	Random Forest, SVM, AdaBoost, and Neural Networks	Soil datasets, climate parameters, and economic data from Indian regions	Provided accurate crop recommendations and yield predictions, reducing farming risks	89.80%
64	Maize yield prediction under conservation agriculture	To predict maize yield in Eastern and Southern Africa under conservation agriculture	Logistic Regression, LDA, KNN, CART, Naive Bayes, and SVM	Seven years of maize-legume cropping data across five countries	Linear models (LDA and Logistic Regression) outperformed nonlinear models in accuracy	61%
TCRM (proposed)		To provide personalized, data-driven crop recommendations using a cloud-based platform	TCRM (transformative crop recommendation model) integrates machine learning and deep learning (e.g., random forest, AdaBoost, multi-head attention layers) with real-time data sources and an SMS alert system	Soil and environmental datasets (e.g., NPK levels, temperature, pH, rainfall) with 3000 samples	The model achieved high accuracy and scalability, effectively addressing precision farming needs	94%

Table 5. Comparative evaluation of research about the crop recommendation. Significant values are in bold.

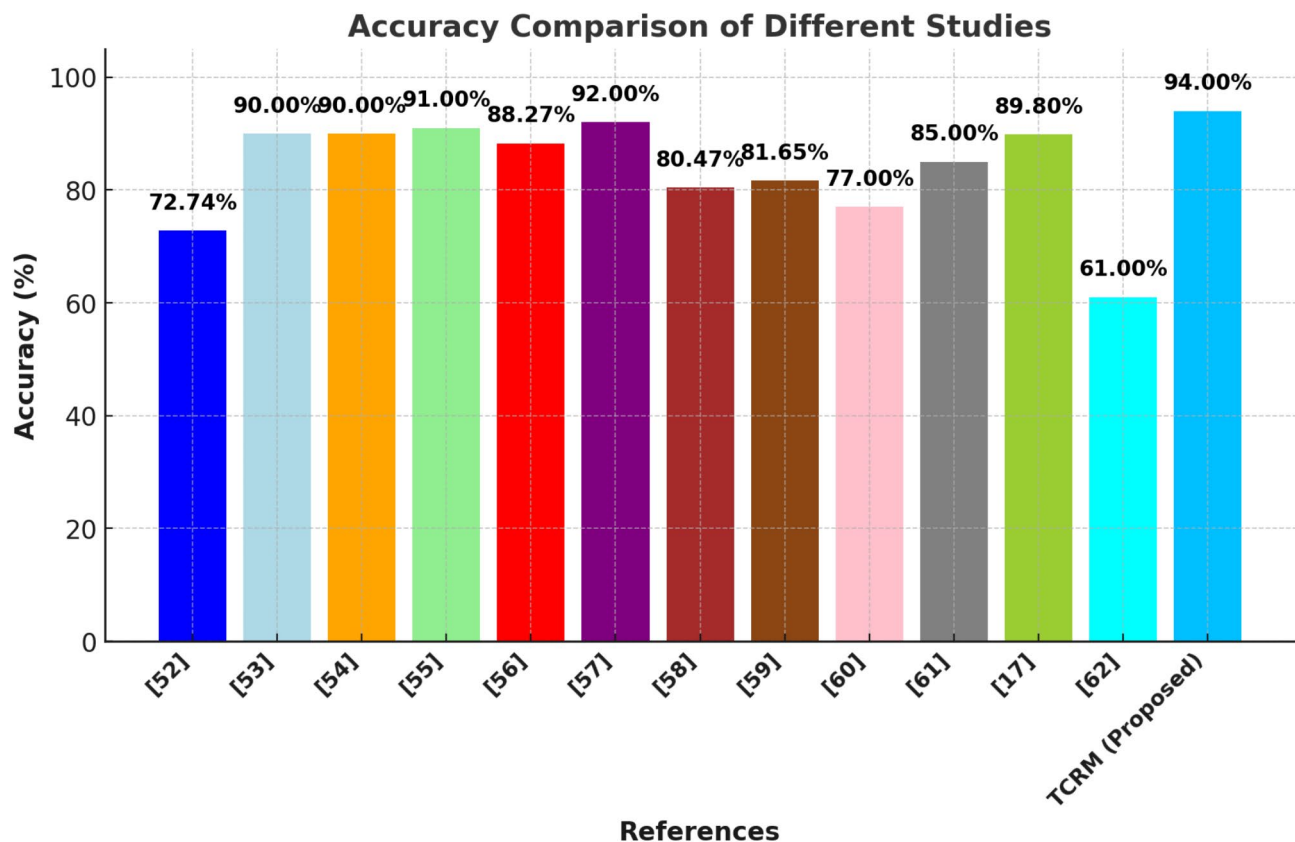


Fig.15. Accuracy comparison of different studies on crop recommendation.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 21 November 2024; Accepted: 6 March 2025

Published online: 17 March 2025

References

1. Khanna, A. & Kaur, S. Evolution of internet of things (IoT) and its significant impact in the field of precision agriculture. *Comput. Electron. Agric.* **157**, 218–231. <https://doi.org/10.1016/j.compag.2018.12.039> (2019).
2. Boursianis, A. D. et al. Internet of things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: A comprehensive review. *Internet Things* **18**, 100187. <https://doi.org/10.1016/j.iot.2020.100187> (2022).
3. Ouhami, M. et al. Computer vision, IoT and data fusion for crop disease detection using machine learning: A survey and ongoing research. *Remote Sens.* **13**, 2486. <https://doi.org/10.3390/rs13132486> (2021).
4. Singh, G. & Sharma, S. A comprehensive review on the internet of things in precision agriculture. *Multimed. Tools Appl.* <https://doi.org/10.1007/s11042-024-19656-0> (2024).
5. Abo-habaga, M. et al. Development of an automated cartesian arm for planting seeds in pots. *J. Soil Sci. Agric. Eng.* **15**, 217–224. <https://doi.org/10.21608/jssae.2024.305809.1238> (2024).
6. Raj, M. et al. A survey on the role of internet of things for adopting and promoting agriculture 4.0. *J. Netw. Comput. Appl.* **187**, 103107. <https://doi.org/10.1016/j.jnca.2021.103107> (2021).
7. Paudel, D. et al. Machine learning for large-scale crop yield forecasting. *Agric. Syst.* **187**, 103016. <https://doi.org/10.1016/j.agry.2020.103016> (2021).
8. Akkem, Y. et al. Streamlit application for advanced ensemble learning methods in crop recommendation systems – A review and implementation. *Indian J. Sci. Technol.* **16**, 4688–4702. <https://doi.org/10.17485/ijst/v16i48.2850> (2023).
9. Akkem, Y. et al. A comprehensive review of synthetic data generation in smart farming by using variational autoencoder and generative adversarial network. *Eng. Appl. Artif. Intell.* **131**, 107881. <https://doi.org/10.1016/j.engappai.2024.107881> (2024).
10. Thilakarathne, N. N. et al. A cloud-enabled crop recommendation platform for machine learning-driven precision farming. *Sensors* **22**, 6299. <https://doi.org/10.3390/s22166299> (2022).
11. Sinha, B. B. & Dhanalakshmi, R. Recent advancements and challenges of internet of things in smart agriculture: A survey. *Future Gener. Comput. Syst.* **126**, 169–184. <https://doi.org/10.1016/j.future.2021.08.006> (2022).
12. Akkem, Y. et al. Streamlit-based enhancing crop recommendation systems with advanced explainable artificial intelligence for smart farming. *Neural Comput. Appl.* **36**, 20011–20025. <https://doi.org/10.1007/s00521-024-10208-z> (2024).
13. Shripathi Rao, M. et al. Crop prediction using machine learning. *J. Phys. Conf. Ser.* **2161**, 012033. <https://doi.org/10.1088/1742-6596/2161/1/012033> (2022).
14. Phasinam, K. et al. Applicability of internet of things in smart farming. *J. Food Qual.* **2022**, 1–7. <https://doi.org/10.1155/2022/7692922> (2022).

15. Gopi, S. S. et al. Multimodal machine learning-based crop recommendation and yield prediction model. *Intell. Autom. Soft Comput.* **36**, 313–326. <https://doi.org/10.32604/iasc.2023.029756> (2023).
16. Waleed, M. et al. Classification of agriculture farm machinery using machine learning and Internet of Things. *Symmetry* **13**, 403. <https://doi.org/10.3390/sym13030403> (2021).
17. Priyadarshini, A. et al. Intelligent Crop Recommendation System using Machine Learning. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). <https://doi.org/10.1109/iccmc51019.2021.9418375> (2021).
18. Alkhudaydi, T. et al. Deep learning for combined water quality testing and crop recommendation. *Int. J. Adv. Comput. Sci. Appl.* **14**, 2023. <https://doi.org/10.14569/ijacsa.2023.0140450> (2023).
19. Mekonnen, Y. et al. Review—Machine learning techniques in wireless sensor network-based precision agriculture. *J. Electrochem. Soc.* **167**, 037522. <https://doi.org/10.1149/2.0222003jes> (2019).
20. Akkem, Y. et al. Smart farming using artificial intelligence: A review. *Eng. Appl. Artif. Intell.* **120**, 105899. <https://doi.org/10.1016/j.engappai.2023.105899> (2023).
21. Bakthavatchalam, K. et al. IoT framework for measurement and precision agriculture: Predicting the crop using machine learning algorithms. *Technologies* **10**, 13. <https://doi.org/10.3390/technologies10010013> (2022).
22. Filippi, P. et al. An approach to forecast grain crop yield using multi-layered, multi-farm datasets and machine learning. *Precis. Agric.* **20**, 1015–1029. <https://doi.org/10.1007/s11119-018-09628-4> (2019).
23. Lavanya, G. et al. An automated low-cost IoT-based fertilizer intimation system for smart agriculture. *Sustain. Comput. Inform. Syst.* **28**, 100300. <https://doi.org/10.1016/j.suscom.2019.01.002> (2020).
24. Gupta, A. & Nahar, P. Classification and yield prediction in a smart agriculture system using IoT. *J. Ambient Intell. Hum. Comput.* **14**, 10235–10244. <https://doi.org/10.1007/s12652-021-03685-w> (2022).
25. Blesslin Sheeba, T. et al. Machine learning algorithm for soil analysis and classification of micronutrients in IoT-enabled automated farms. *J. Nanomater.* **2022**, 1–7. <https://doi.org/10.1155/2022/5343965> (2022).
26. Singh, D. K. et al. LoRa-based intelligent soil and weather condition monitoring with Internet of Things for precision agriculture in smart cities. *IET Commun.* **16**, 604–618. <https://doi.org/10.1049/cmu2.12352> (2022).
27. Khaki, S. et al. A CNN-RNN framework for crop yield prediction. *Front. Plant Sci.* <https://doi.org/10.3389/fpls.2019.01750> (2020).
28. Shaikh, T. et al. Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Comput. Electron. Agric.* **198**, 107119. <https://doi.org/10.1016/j.compag.2022.107119> (2022).
29. Pagano, A. et al. A survey on LoRa for smart agriculture: Current trends and future perspectives. *IEEE Internet Things J.* **10**, 3664–3679. <https://doi.org/10.1109/jiot.2022.3230505> (2023).
30. O'Sullivan, C. A. et al. Strategies to improve productivity, product diversity, and profitability of urban agriculture. *Agric. Syst.* **174**, 133–144. <https://doi.org/10.1016/j.agsy.2019.05.007> (2019).
31. Franzluebbers, A. J. & Gastal, F. Building agricultural resilience with conservation pasture-crop rotations. *Agroecosyst. Divers.* <https://doi.org/10.1016/b978-0-12-811050-8.00007-8> (2019).
32. Bandara, P. et al. Crop recommendation system. *Int. J. Comput. Appl.* **175**, 22 (2020).
33. Reeves, D. W. Cover crops and rotations. *Crops Residue Manag.* <https://doi.org/10.1016/j.cropro.2018.07.007> (2018).
34. Roos, E. et al. Risks and opportunities of increasing yields in organic farming. A review. *Agron. Sustain. Dev.* **38**, 0489. <https://doi.org/10.1007/s13593-018-0489-3> (2018).
35. Assarroudi, A. et al. Directed qualitative content analysis: The description and elaboration of its underpinning methods and data analysis process. *J. Res. Nurs.* **23**, 42–55. <https://doi.org/10.1177/1744987117741667> (2018).
36. Gebru, T. et al. Datasheets for datasets. *Commun. ACM* **64**, 86–92. <https://doi.org/10.1145/3458723> (2021).
37. Lu, Y. & Young, S. A survey of public datasets for computer vision tasks in precision agriculture. *Comput. Electron. Agric.* **178**, 105760. <https://doi.org/10.1016/j.compag.2020.105760> (2020).
38. Fatima, A. et al. Data cleaning in data warehouse: A survey of data pre-processing techniques and tools. *Int. J. Inf. Technol. Comput. Sci.* **9**, 50–61. <https://doi.org/10.5815/ijitcs.2017.03.06> (2017).
39. Obaid, H. S. et al. The impact of data pre-processing techniques and dimensionality reduction on the accuracy of machine learning. In *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*. <https://doi.org/10.1109/iemecon.2019.8877011> (2019).
40. Salgado, C. M. et al. Missing data. *Sec. Anal. Electron. Health Rec.* https://doi.org/10.1007/978-3-319-43742-2_13 (2016).
41. Austin, P. C. et al. Missing data in clinical research: A tutorial on multiple imputation. *Can. J. Cardiol.* **37**, 1322–1331. <https://doi.org/10.1016/j.cjca.2020.11.010> (2021).
42. Mishra, P. et al. New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC Trends Anal. Chem.* **132**, 116045. <https://doi.org/10.1016/j.trac.2020.116045> (2020).
43. Borkin, D. et al. Impact of data normalization on classification model accuracy. *Res. Pap. Fac. Mater. Sci. Technol. Slovak Univ. Technol.* **27**, 79–84. <https://doi.org/10.2478/rput-2019-0029> (2019).
44. Camizuli, E. & Carranza, E. J. Exploratory data analysis (EDA). In *The Encyclopedia of Archaeological Sciences*. 1–7. <https://doi.org/10.1002/9781119188230.saseas0271> (2018).
45. Cai, J. et al. Feature selection in machine learning: A new perspective. *Neurocomputing* **300**, 70–79. <https://doi.org/10.1016/j.neucom.2017.11.077> (2018).
46. Urbanowicz, R. J. et al. Relief-based feature selection: Introduction and review. *J. Biomed. Inform.* **85**, 189–203. <https://doi.org/10.1016/j.jbi.2018.07.014> (2018).
47. Sajid, S. S. et al. County-scale crop yield prediction by integrating crop simulation with machine learning models. *Front. Plant Sci.* **13**, 100022. <https://doi.org/10.3389/fpls.2022.1000224> (2022).
48. Gonzalez, S. et al. A practical tutorial on bagging and boosting-based ensembles for machine learning. *Inform. Fusion* **64**, 205–237. <https://doi.org/10.1016/j.inffus.2020.07.007> (2020).
49. Attaluri, S. S. et al. Crop plantation recommendation using feature extraction and machine learning techniques. *J. Appl. Technol. Innov.* **4**, 4 (2020).
50. Chen, A. et al. Developments in MLflow. *Proc. Fourth Int. Workshop Data Manag. End-to-End Mach. Learn.* <https://doi.org/10.1145/3399579.3399867> (2020).
51. Bisong, E. Building machine learning and deep learning models on Google Cloud Platform: A comprehensive guide for beginners. Apress (2019).
52. Butt, U. A. et al. A review of machine learning algorithms for cloud computing security. *Electronics* **9**, 1379. <https://doi.org/10.3390/electronics9091379> (2020).
53. Allen-Sader, C. et al. An early warning system to predict and mitigate wheat rust diseases in Ethiopia. *Environ. Res. Lett.* **14**, 115004. <https://doi.org/10.1088/1748-9326/ab4034> (2019).
54. Schwalbert, R. A. et al. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agric. Forest Meteorol.* **284**, 107886. <https://doi.org/10.1016/j.agrformet.2019.107886> (2020).
55. Bhojani, S. H. & Bhatt, N. Wheat crop yield prediction using new activation functions in neural network. *Neural Comput. Appl.* **32**, 13941–13951. <https://doi.org/10.1007/s00521-020-04797-8> (2020).
56. Suchithra, M. S. & Pai, M. L. Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters. *Inf. Process. Agric.* **7**, 72–82. <https://doi.org/10.1016/j.inpa.2019.05.003> (2020).
57. Elumalai, M. & Anuncia, S. M. Development of soil mineral classification using ontology mining. *Arab. J. Geosci.* **14**, 1371. <https://doi.org/10.1007/s12517-021-07651-w> (2021).

58. Jose, A. et al. Detection and classification of nutrient deficiencies in plants using machine learning. *J. Phys. Conf. Ser.* **1850**, 012050. <https://doi.org/10.1088/1742-6596/1850/1/012050> (2021).
59. Dash, R. et al. Classification of crop based on macronutrients and weather data using machine learning techniques. *Results Eng.* **9**, 100203. <https://doi.org/10.1016/j.rineng.2021.100203> (2021).
60. Shi, P. et al. Rice nitrogen nutrition estimation with RGB images and machine learning methods. *Comput. Electron. Agric.* **180**, 105860. <https://doi.org/10.1016/j.compag.2020.105860> (2021).
61. Pantazi, X. E. et al. Wheat yield prediction using machine learning and advanced sensing techniques. *Comput. Electron. Agric.* **121**, 57–65. <https://doi.org/10.1016/j.compag.2015.11.018> (2016).
62. Shidnal, S., Latte, M. V. & Kapoor, A. Crop yield prediction: Two-tiered machine learning model approach. *Int. J. Inf. Technol.* **13**, 1983–1991. <https://doi.org/10.1007/s41870-021-00649-2> (2021).
63. Kussul, N. et al. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geosci. Remote Sens. Lett.* <https://doi.org/10.1109/LGRS.2017.2681128> (2017).
64. Mupangwa, W. et al. Evaluating machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa. *SN Appl. Sci.* **2**, 952. <https://doi.org/10.1007/s42452-020-2711-6> (2020).

Acknowledgements

Special thanks to Dr. Sandeep Sharma for his great support and guidance.

Author contributions

A wrote the main manuscript text and prepared figures. B supervised the work All authors reviewed the manuscript. A: Gurpreet Singh B: Sandeep Sharma.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to G.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025