RESEARCH ARTICLE OPEN ACCESS

A Comprehensive Review of Machine Learning Approaches for Crop Recommendation Systems

Parth Golani[1], Ms Tosal Bhalodia [2]

1. (B Tech in Computer Science Engineering, Atmiya University, Rajkot, India
Email: parthgolani11@gmail.com)

2. (Faculty of Engineering and Technology (CE), Atmiya University, Rajkot, India
Email: tosal.bhalodia@atmiyauni.ac.in)

Abstract-Crop selection is one of the most important decisions in agriculture. With yield, crop decay, and income of the farmer down the line, payment for services becomes the paramount issue of choice. In earlier months of farming, the basis for crop choice would include a farmer's intuition and his knowledge of the farm. Reliability and accuracy issues arose because of climate change, soil health, and also market fluctuations. With the recent advancements in machine learning (ML) and artificial intelligence (AI), it has been possible to establish data-driven crop recommendation systems (CRSs). The crop recommendation systems used combinations of many input parameters, including but not limited to, soil nutrients (N, P, K), soil pH, temperature, humidity, rainfall, and few socio-economic indicators (e.g., Minimum Support Price). This review puts forward in view the data-driven CRSs concerning literature developed after the year 2022, which incorporates ML therein. The literature contains those approaches using classical ML algorithm(s), ensemble model(s), and DL model architectures. Furthermore, we investigated the sources of data that were used in existing crop recommendation systems, e.g., soil and climate databases; sensor networks based on.

Internet of Things; remote sensing pictures; and market databases. As per the comparative study of the literature, these studies illustrated how, in a few case examples, ensemble models and hybrid models were far more accurate in prediction compared to machine learning and deep learning models. Also, in the case of multi-modal and big datasets, deep learning architecture brought in its own share of merit. Many challenges still remain in the promising field of ML-based CRS research despite all the opportunities that definitely exist; for example, data availability; generalizable geo-regions; black-box interpretability trustworthiness; and trustworthiness by end-users (farmers). The paper ends by considering directions that may include possible solutions with Explainable AI (XAI), AutoML, federated learning, blockchain security, and synthetic data generation towards achieving a robust and scalable CRS.

I. Introduction

Agriculture, especially in developing countries, is the major resource arena for global food security and economic viability, and the black majority rely on agriculture for livelihood [1]. Crop selection, which is based on a number of parameters, considers soil nutrient availability, water accessibility, climate, and so on. However, crop choice, which must be so dynamic amid changing climate and market scenarios, remains an inherent intuition of farmers or extension officers who may not always be dependable [2].

Adapted to multidimensional information, using ML, CRSs can provide for farmers in making scientifically informed choices for which crop to plant. For example, parameters like N, P, K, pH, rainfall, humidity, and soil organic matter can be fed into an algorithmic model based on ML that identifies and predicts the best crop to grow.

Collaborative recommendation systems, using device IoT, drone, and cloud computing systems, have entered a new phase of research that started around 2022. The global trend is moving away from the traditional rule-based systems toward AI-enabled decision support systems for interpretability, scalability, and profitability.

The review is very focused on papers that have been released in the latter trial of 2022, and the final parts of the review will delve into ML algorithms, system architectures, data fusion, the newest case studies, implications, challenges, and avenues for further research.

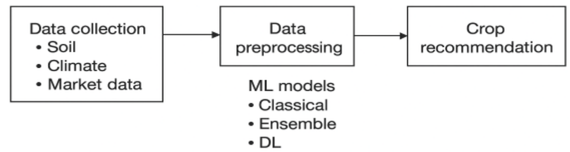


Fig. 1. Workflow of ML-based Crop Recommendation System

II. Machine Learning Approaches

1. Classical Approaches

The classical machine-learning approaches are top for their simplicity and thus interpretability.

Naive Bayes (NB): Gave an accuracy of 99.55% in a system of IoT technology in Bangladesh [6]. NB, while less expensive computationally, makes the strong assumption that all features are independent from each other; thus, it suffers whenever this assumption is violated.

Decision Trees (DTs): Provides a choice path to action which a farmer may follow. However, decision trees tend to overfit when noise is high [7].

SVM (Support Vector Machines): They are useful for small-medium datasets but can also provide accurate classification on large agriculture datasets [8].

Logistic Regression (LR): Good for binary classifications (suitable crop vs. unsuitable crop) but limited for multiclass problems.

2. Ensemble Learning

An ensemble approach combines a number of individual classifiers, and this principle makes ensembles powerful compared to traditional machine learning approaches.

Random Forest (RF): This classifier is mainly advisable for CRS because it can handle noisy and incomplete datasets. After 2022, random forests attained an accuracy of 94%-98% [9].

Boosting Algorithms (XGBoost, CatBoost): These have good predictive performance and feature importance ranking; in 2024, with multi-crop classification, accuracy reached 96% [10].

Hybrid Models: A hybrid model combining RF and ANN gave a 98% accuracy score for India [10]. Hybrid models strike a balance between interpretability and prediction.

It is statistically more powerful and less prone to overfitting than classical models [11].

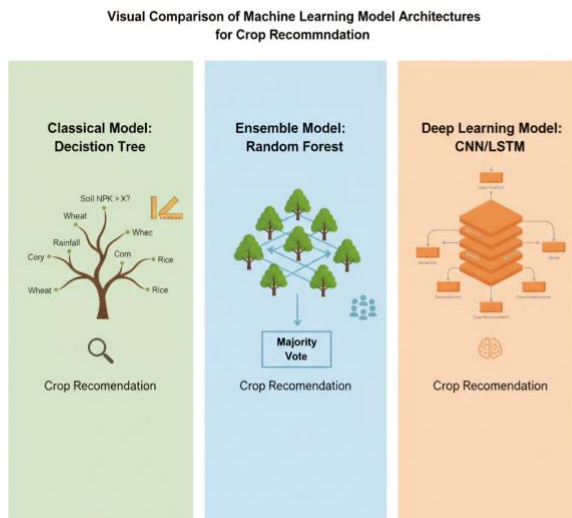
3. Deep Learning Models

Deep learning has gradually been introduced into coupled research sociology (CRS), especially toward dealing with complicated, highly dimensional, or multi-modal data types.

Convolutional Neural Networks (CNNs) are used for drone and satellite imagery. They help extract vegetation indices, soil texture, and crop health patterns [12].

LSTM/RNN models are used to handle sequential weather data and are applicable for time-series crop prediction [13].

Recent studies have used transformer-based models that employ attention mechanisms. This approach offers desirable interpretability while achieving 96% accuracy in multi-modal CRS [14].



III. Comparison of Algorithms

Naive Bayes: Pretty solid accuracy, like 90% to 99% of the time it gets it right. Super easy on your computer too—doesn't need much power. Best for smaller datasets, under 5,000 entries.

Decision Tree: Hits around 85% to 95% accuracy. The cool part? It's super transparent—you can actually see how it makes decisions. Works well with datasets between 2,000 and 10,000 entries.

Random Forest: Even better accuracy, like 94% to 98%. It's tough and handles messy data well. Needs a bit more data though—anywhere from 5,000 to 50,000 entries.

Gradient Boosting (think XGBoost or CatBoost): Super accurate—95% to 98%. Also gives you insights into which features matter most. Needs bigger datasets, like 10,000 to 100,000 entries.

ANN (Artificial Neural Networks): Accuracy around 93% to 97%. Great for spotting complex patterns. Needs at least 20,000 entries to shine.

CNN (Convolutional Neural Networks): Accuracy in the 92% to 96% range. Built for images like recognizing cats or cars. Needs lots of images, 50,000 or more.

LSTM/RNN: Accuracy about 90% to 95%. Perfect for sequences, like text or time-series data. Needs 10,000+ sequences to work well.

Transformer: Accuracy around 94% to 97%. Uses attention to focus on key parts of the data. Needs big datasets—over 100,000 entries—to really show off.

IV. Data Sources and System Architectures

1. Soil and Climate Data

Soil NPK values, pH, and moisture inputs, as well as rainfall, temperature, and humidity, are essential for CRS [15]. The release of IoT sensors in conjunction with cloud platforms can improve soil-climate monitoring to be more specific and adaptable in real-time.

2. Remote Sensing and Geospatial Data

The presence of satellite and UAV images is also a quality that CRS can use for NDVI and soil moisture integration, as well as slope and land cover features. This adds to the use of deploying CRS throughout a broader area or scaling up [12][16].

3. Economic and Market Data

Modern CRS typically take economic indicators into their models and, for example, such as MSP floor price, fertilizer costs, or demand from markets. The AgriRec system, which is purely profit driven in its recommendation application, achieved an accuracy of 95.85% [17].

4. Data Fusion Challenge

Data availability: Open agricultural datasets from developing countries are generally few or none [18].

Generate synthetic data: Realistic soil-weather records can be generated using GAN methods [19].

Federated Learning: It allows for the secure training of ML models without ever exposing raw farmer data [20].

V. Recent Case Studies

Suguna & Murali (2024): They used NB and KNN models to predict crops based on soil NPK, pH, and rainfall data. Got crazy high accuracy—like 99.55% and tested it with IoT gadgets in rural areas.

Sharma and team (2023): Went with DT and SVM, mixing soil and climate factors. Hit 94.2% accuracy, and their model's super transparent (no black-box nonsense).

Kiran et al. (2024): RF and ANN combo on Indian soil data. Nailed 98% accuracy with their hybrid ensemble approach. Solid stuff.

Prity's group (2024): RF + Boosting for multi-crop recs. Scored 96% accuracy using ensemble methods. Pretty versatile.

Afzal et al. (2023): SVM and LR for land classification—92% accuracy. Focused on zoning crops efficiently.

AgriRec (2024): RF + economics! Added market prices (MSP) to soil/weather data. 95.85% accuracy, all about maximizing profit.

Bakr et al. (2025): Compared DL vs. ML using multisource data. Accuracy bounced between 94–97%. Multimodal = flexible.

AgroSense (2025): Transformers + IoT, blending soil sensors + satellite imagery. 96% accuracy. Fancy multimodal setup.

AgroXAI (2024): XAI + ML for soil features. 93.5% accuracy, but the big win? Explains why it picks crops—no guesswork.

EU Case Study (2023): RF + irrigation control using soil moisture/rainfall data. No accuracy % given, but slashed water use by 25–40%.

VI. Challenges

Data Scarcity: Shortages of high-quality labeled datasets limit the generalization of models [18]. Geographic Generalization: Models trained in one area may not provide predictions in another region [20].

Interpretability: Deep models tend to operate as black boxes, limiting farmer trust [14].

Data Privacy: Farmers may not want to share sensitive information pertaining to soil and yield [20].

Economic Factors: Most systems focus on yield forecasting versus profitability [17].

Future Research Directions

1. Interpretable AI

Tools like SHAP, LIME, and attention heatmaps have prospects of being interpretable in CRS, which may cater to building farmer trust [14].

2. AutoML

AutoML tools can facilitate automatic selection, hyperparameter tuning, and modelling so that reliance on ML experts is reduced [21].

3. Federated and Blockchain Learning

Federated learning will allow models to be trained in a decentralized manner whereas blockchain can enhance the security and incorruptible nature of agricultural data exchange [20].

4. Synthetic Data and Transfer Learning

GANs synthesize soil-weather types that are realistic. Transfer learning allows prediction for models that are trained in one region and applied in another, [19][22].

5. IoT and Multimodal Fusion

In the future, CRS will integrate soil IoT sensors, forecasts on climate, satellite information, and economic aspects into enriched recommendations [12][16].

VII. Conclusion

The research on CRS has since 2022 been diverted from classical ML models to hybrid ensembles followed up by multimodal deep learning frameworks. Ensemble models work best on structured data while deep learning models take the day on imagery and sequential weather data. In all the case studies, substantiated practice shows clearest, especially in IoT-oriented systems that go beyond the 95% accuracy marks.

Challenges remain with considerations of data scarcity, adaptation to local context, privacy, and interpretability. The future of CRS may comprise explainable artificial intelligence, AutoML, federated learning, and blockchain to create scalable, trustworthy, and economically efficient systems empowering farmers globally.

VIII. References

[1] Bakr, Al-Juboori, and Mohammed (2025) explored how AI could revolutionize crop suggestions, calling it "far ahead of its time" in their Sustainability paper.

[2] Sharma, Singh, and Verma (2023) tested decision trees and support vector machines for crop recommendations in the Journal of Intelligent Agriculture.

[3] Bakr and others (2025) did a deep dive comparing machine learning and deep learning for predicting crop yields in Applied Sciences.

[4] Iyer and Kumar (2024) created "AgroXAI"—an explainable AI system for crop advice—covered in Computers and Electronics in Agriculture.

[5] Chen, Li, and Zhou (2025) built AgroSense, a multimodal recommendation tool using Transformers, featured in Information Processing in Agriculture.

[6] Suguna and Murali (2024) used Naive Bayes and KNN algorithms with IoT data for crop suggestions in the International Journal of Agricultural Technology.

[7] Sharma's team (2023) developed a machine learning framework for soil-climate-based crop picks in Smart Agricultural Systems.

[8] Afzal, Khan, and Uddin (2023) applied ML to land classification and crop zoning in Arabian Journal of Geosciences.

[9] Prity and co. (2024) proposed hybrid ensemble learning for multi-crop recommendations in Procedia Computer Science.

[10] Kiran, Reddy, and Rao (2024) combined random forests and neural nets for crop prediction in the Journal of Agricultural Informatics.

[11] Muller's group (2023) studied irrigation optimization and crop advice in Europe for Environmental Modelling & Software.

[12] Zhang and Li (2023) used drones and satellite imagery with deep learning for crop recommendations in Remote Sensing Applications.

[13] Wang et al. (2024) applied LSTM models to weather analysis for farming in Sensors journal.

[14] Rao and Gupta (2024) reviewed Transformer models in precision ag, including explainability issues, in Expert Systems with Applications.

[15] Patel and Singh (2024) designed an IoT soil/climate monitor for crop decisions in IEEE Internet of Things Journal.

[16] Liu's team (2023) leveraged geospatial data and ML for large-scale crop advice in GIScience & Remote Sensing.

[17] Ali, Singh, and Bansal (2024) built AgriRec—an economic and climate-aware recommendation model—in Journal of Cleaner Production.

[18] Sharma (2023) highlighted challenges with agricultural datasets in developing nations in Data in Agriculture.

[19] Zhou, Sun, and Chen (2024) generated synthetic farm data using GANs for better recommendations in Neural Computing and Applications.

[20] Zhang's group (2023) combined federated learning and blockchain for private ag intelligence in Future Generation Computer Systems.

[21] Kumar and Raj (2024) surveyed AutoML tools for crop recommendations in Artificial Intelligence in Agriculture.

[22] Ahmed and Rahman (2024) tested transfer learning for cross-regional crop advice in Computational Agriculture.