MACHINE LEARNING BASED CROP SELECTION TECHNIQUES IN MODERN FARMING

Shubhangi Wankhade [*1], Prof. Vaibhav Kshirsagar [*2], Dr. Varsha Turkar[*3]

[*1] M. Tech Student, Department of Electronics and Telecommunication, Vidyalankar Institute of Technology, Mumbai, Maharashtra, India.

[*2,3] Professor, Department of Electronics and Telecommunication, Vidyalankar Institute of Technology, Mumbai, Maharashtra, India.

ABSTRACT

In an age of climate volatility and growing agricultural demand, precision farming has become a necessity rather than a luxury. This work introduces a smart crop advisory framework that combines Internet of Things (IoT) sensing with machine learning (ML) techniques to generate intelligent crop suggestions based on real-time soil parameters. A compact, cost-effective IoT device captures essential soil parameters and transmits data to a cloud-based analytics platform. Four ML algorithms—Random Forest, Support Vector Machine, Decision Tree, and Naïve Bayes—were evaluated for predictive performance using both public and real-world datasets. The system delivers accurate and context-aware crop recommendations, addressing limitations of static models and outdated advisory tools. By bridging the gap between environmental sensing and AI-driven decision support, this approach offers a scalable, farmer-friendly pathway toward smart and sustainable agriculture.

Keywords: Artificial Intelligence, Machine Learning, Crop Recommendation, Precision Agriculture, Soil Data, Sustainable Farming, Decision Support System.

I. INTRODUCTION

Agriculture in India faces a persistent challenge of declining productivity due to climate variability, unsystematic crop selection, and lack of soil-specific decision-making tools. While traditional practices depend on farmers' experience, they often fail to match the actual requirements of soil fertility and climate dynamics, especially in vulnerable regions like Vidarbha. This gap between crop choice and field suitability has led to reduced yields, financial losses, and unsustainable farming patterns.

With the advent of the Internet of Things (IoT) and Machine Learning (ML), a new era of precision agriculture has emerged. However, most existing solutions are either model-driven using outdated datasets or lack real-world deployment in geographically diverse regions. In this context, our project proposes a real-time, data-driven crop recommendation system specifically tailored for the Vidarbha region.

The system integrates a custom-designed IoT device that captures six critical soil parameters—Nitrogen (N), Phosphorus (P), Potassium (K), pH, temperature, and humidity. These real-time inputs, collected from over 2700 entries spanning 27 crops, are processed using a suite of machine learning models, including Random Forest, Support Vector Machine, Naïve Bayes, and Decision Tree classifiers. The combination of localized data and ensemble modeling allows for highly accurate and regionally optimized crop recommendations.

This research not only fills the gap of real-time, location-specific crop prediction but also sets a new benchmark for deploying scalable, farmer-friendly solutions in rural India. It contributes significantly to the growing field of smart agriculture by enhancing both technological applicability and agronomic relevance. By focusing on real-time soil data and integrating region-specific inputs, the system addresses critical limitations in traditional and pre-trained models that lack local adaptability. The use of ensemble machine learning techniques further enhances prediction accuracy while ensuring interpretability and practical applicability.

II. METHODOLOGY

Behind every successful farm is not just good soil, but the ability to read it. Our system bridges the gap between traditional intuition and modern intelligence. It combines live soil sensing, cloud connectivity, and machine learning to recommend the most suitable crops based on real conditions—not assumptions. The methodology is structured into five dynamic phases that together enable real-time, reliable, and rural-ready crop prediction.

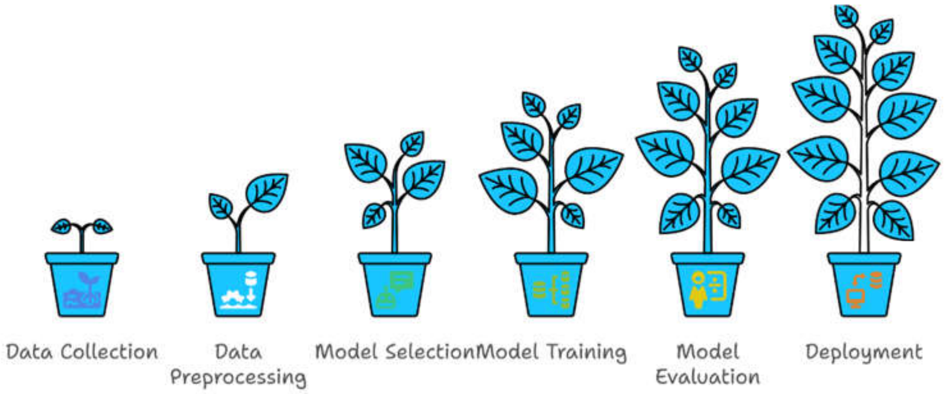

Fig.1 Methodology for Crop Recommendation System

1. Smart Hardware & Software Synergy: Simplicity with Intelligence

- The backbone of our system is a compact yet powerful trio: a soil sensor, an Arduino Nano, and an ESP8266 Wi-Fi module. True innovation begins when hardware and software come together as one system.
- Silently nestled at the system's core, the Arduino Nano operates as its silent intellect — compact in size, yet commanding in control, seamlessly weaving together sensing, logic, and transmission. It was programmed using Embedded C within the Arduino IDE, a developer-friendly platform that allows real-time sensor control, logic execution, and precise timing. The use of C language gives direct access to hardware resources, ensuring reliability in outdoor, real-world farm conditions.
- The soil sensor feeds this brain with live data—capturing six vital parameters: Nitrogen, Phosphorus, Potassium, pH, Temperature, and Humidity—in a single cycle.
- ESP8266 was programmed with AT commands to manage cloud communication. Once the data is processed by the Arduino, the ESP8266 instantly transmits it wirelessly to the cloud.

Together, this hardware-software blend forms an elegant closed-loop system—easy to deploy, power-efficient, and intelligently responsive to the environment around it. No complex rigs, no redundant components—just one sensor, one microcontroller, and one goal: letting the soil speak.

2. Real-Time Sensing Meets the Cloud: Data That Thinks for Itself

In agriculture, timing is everything. Knowing the exact state of your soil at the right moment can mean the difference between abundance and loss. That's why our system doesn't just collect data—it does it in real time, with intelligence and precision.

Once the soil sensor gathers the six core parameters—N, P, K, pH, temperature, and humidity—the Arduino Nano kicks into action. It transforms unprocessed readings into refined information—filtering noise, structuring data, and priming it for flawless transmission. Within milliseconds, this information is handed off to the ESP8266 Wi-Fi module, which sends it wirelessly to the cloud. We use Thing Speak, a real-time IoT analytics platform, as our cloud partner. Here's what it enables:

- Interactive charts offering live updates on soil quality metrics
- Time stamped logging, making it possible to track changes over days or seasons
- Remote accessibility, so data from remote farms is available anywhere, anytime

This real-time connection means farmers or agronomists don't need to guess. The system continuously monitors conditions, updates itself, and stays alert—like a silent digital companion that never stops watching over the farm. And unlike bulky legacy setups, this system is minimal and wireless—no manual input, no external storage, no delay.

Fig.2 Hardware set up and Soil Samples

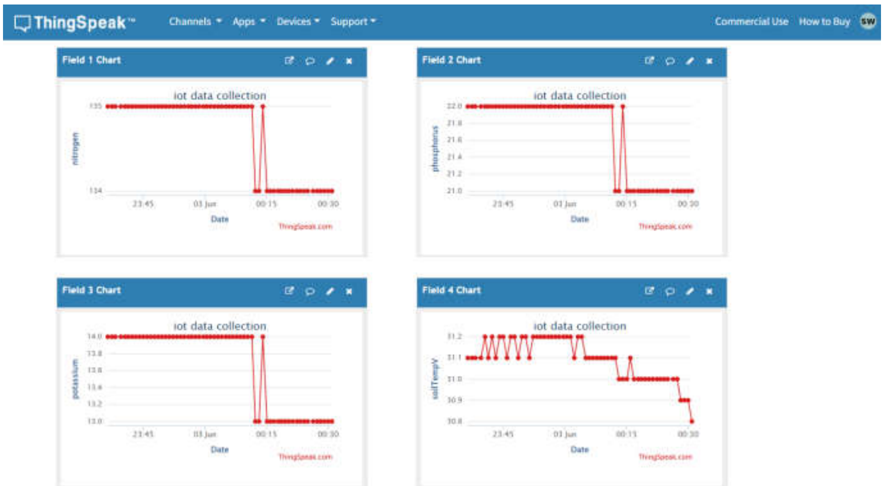

Fig.3 a ThingSpeak Result (Data Collection – Nitrogen, Phosphorous, Potassium, Temperature)

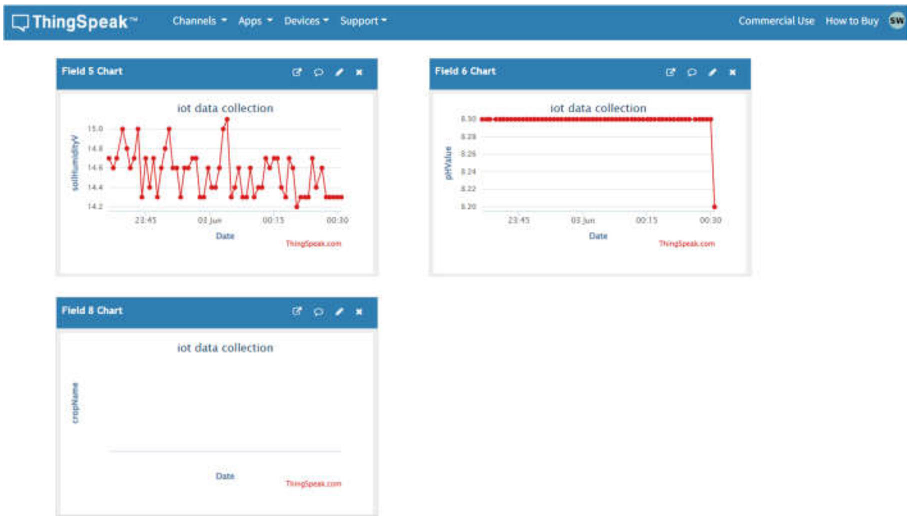

Fig.3 b ThingSpeak Result (Data Collection – Humidity, pH, Crop Name)

3. A Dataset Grown from the Ground Up

Data is only powerful when it speaks the language of reality. That's why we didn't rely solely on online repositories or synthetic datasets. Instead, we built a ground-truth dataset, grown directly from real farms right in the heart of Maharashtra's Vidarbha region.

Over time, we collected more than 2700 entries, each representing a unique snapshot of soil health. Every record included six features—N, P, K, pH, temperature, and humidity—paired with the actual crop grown or recommended under those conditions. But we didn't stop there. To enrich the system's understanding and generalizability, we combined our field data with trusted, agriculture-specific datasets from:

- ICAR (Indian Council of Agricultural Research)
- Maharashtra Agriculture Department
- Krishi Vigyan Kendras (KVKs)
- And global datasets from Kaggle

We preprocessed all data meticulously:

- Cleaned out noisy or missing entries
- Normalized sensor readings for consistency
- Encoded crop labels for machine learning compatibility
- Dataset partitioned (80% train, 20% test)

This dataset isn't just a collection of numbers—it's a living reflection of real soil, real crops, and real decisions made by real farmers. It stands as the silent force powering the intelligence behind every prediction.

4. Training the Brain: When Machines Learn to Think Like Farmers

With the dataset in place, we moved to the brain of the system—the machine learning models. Our goal wasn't just accuracy; it was reliability, explainability, and adaptability to unpredictable field conditions. Python was used to train four ML models for classification tasks:

- Decision Tree – for its transparency and interpretability
- Random Forest – chosen for its high accuracy and resistance to overfitting
- Naïve Bayes – for fast, probabilistic learning
- Support Vector Machine (SVM) – for precision on tighter boundaries

Each model was fed the full spectrum of soil parameters and trained to map them to the ideal crop outcome. We evaluated them using:

- Accuracy – overall success rate
- Precision and Recall – for crop-specific prediction strength

- Confusion Matrix – used to analyze per-class prediction accuracy

The Random Forest model rose to the top, outperforming others across metrics. Its ability to handle environmental variability and noisy real-world data made it ideal for deployment in the field. Still, all four models were preserved for ongoing evaluation and future ensemble improvement.

We didn't just teach a machine to predict—we taught it to understand the land.

5. Real-Time Crop Prediction:

Once deployed, the system continuously reads sensor data and generates crop recommendations in real time. These results are:

- Visualized on the Thing Speak dashboard, and
- Stored locally for offline analysis or dissemination through advisory networks.
- Future versions will enable direct farmer interaction via SMS, mobile apps, or voice assistant integration.

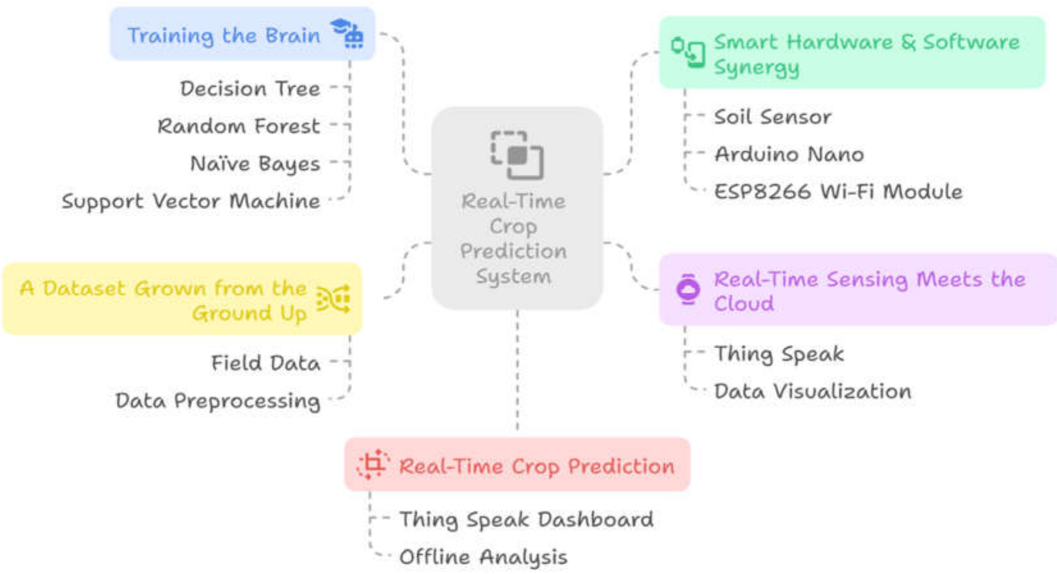

Fig.4 Overview of Hardware and Software System

III. LITERATURE REVIEW

Recent advancements in precision agriculture have largely focused on combining Internet of Things (IoT) with machine learning (ML) to optimize crop selection. Various systems have been developed to analyze soil parameters and suggest crops accordingly, but challenges remain in scalability, real-time deployment, and localized accuracy—especially in regions like Vidarbha, Maharashtra.

Korde et al. [1] introduced an IoT-enabled crop recommendation system that used NPK, pH, temperature, and humidity data transmitted via Wi-Fi modules. Their work laid the groundwork for integrating sensor networks with ML, though it was limited in dataset scale and geographic reach. Manju et al. [4] and Anusha et al. [7] proposed similar systems with basic microcontroller setups for soil nutrient sensing. Deshpande et al. [8] went further by focusing on affordable soil temperature and moisture monitoring, yet none of these systems provided extensive real-time crop-specific deployment across diverse villages.

A leap forward came with explainable AI models like AgroXAI by Turgut et al. [2], which introduced interpretable ML using SHAP and LIME to make AI decisions understandable to farmers. Similarly, Garg & Alam [3] implemented a high-accuracy hybrid model combining wrapper-based feature selection and PART classification, achieving over 99% accuracy on soil data. Hossain et al. [5] also utilized ensemble ML algorithms to generate real-time recommendations from live sensor data. These efforts improved model transparency and performance but were mostly confined to controlled environments or small pilot studies.

To enhance prediction robustness, Gwalani et al. [6] and Hasan et al. [10] focused on algorithm optimization through hyperparameter tuning and ensemble learning techniques. Singh & Sharma [9] expanded the scope further by creating a cloud-based platform for precision agriculture that leverages real-time data flow. Reddy & Neerugatti [12] conducted comparative studies on Indian soil datasets, concluding that models like Random Forest offered superior accuracy across diverse conditions. Meanwhile, Sam & D'Abreo [14] introduced the idea of incorporating environmental and economic parameters to refine ML predictions, marking a shift toward holistic recommendation frameworks.

More futuristic approaches were proposed by Banerjee et al. [13], who integrated digital twin technology for simulated real-time agricultural behavior, while Penchalaiah & Emmanuel [11] and Wilberforce & Mwebaze [15] laid out smart agriculture architectures using cloud platforms like Thing Speak and general IoT protocols.

Despite these contributions, most existing systems face three key limitations: (1) static or non-local datasets, (2) lack of deployment in varied rural conditions, and (3) single-model dependency. Our project addresses these gaps by implementing a real-time, village-level crop recommendation system deployed across multiple farms in Jalgaon Jamod taluka of Vidarbha. It integrates six critical soil parameters (N, P, K, pH, temperature, and humidity), uses a custom-built IoT device, and feeds live data into multiple ML models including Random Forest, SVM, Decision Tree, and Naïve Bayes. With over 2700 data entries covering 27 crops and more than 100 entries per crop, our system provides a robust, real-world recommendation engine tailored to local agricultural needs.

Unique Contribution of Our Work

- **Scale:** 2700+ real-time entries across 27 crops from multiple villages.
- **Depth:** Live capture of NPK, pH, temperature, and humidity via a custom-built IoT sensor.
- **ML Ensemble:** Evaluation using multiple algorithms to ensure stability and accuracy.
- **Local Relevance:** First-of-its-kind field deployment in the Vidarbha region, offering region-specific recommendations unlike generic models.

IV. RESULT AND DISCUSION

A successful crop recommendation system must not only be technically robust but also contextually intelligent—tailored to the environment in which it is deployed. To test the validity and accuracy of our machine learning-based solution, we conducted a thorough model evaluation using both standard datasets and real-time, field-collected data. This section presents a comparative performance analysis of various algorithms, discusses key findings, and interprets the practical significance of the results.

1.Evaluation on Benchmark Dataset

As a preliminary benchmark, the system was trained and tested using a publicly available dataset sourced from Kaggle. This dataset includes essential agronomic features such as nitrogen (N), phosphorus (P), potassium (K), pH, temperature, and humidity, with the crop type serving as the label for supervised classification.

Model	Accuracy (%)	Error Rate (%)
Random Forest	78.18	21.82
Decision Tree	76.82	23.18
Support Vector Machine	77.95	77.05
Naive Bayes	78.18	21.82

Despite showing moderate success, these results reveal a key limitation: models trained on generalized, non-localized data struggle to achieve high accuracy. Soil composition and weather patterns can vary significantly across regions, and models must adapt to such variability to be truly reliable.

2. Evaluation on Real-Time Regional Dataset

To improve relevance and model accuracy, we developed a rich, real-time dataset comprising over 2,700 entries, each representing live soil readings taken using a custom-built sensor module. This dataset covered 27 different crops, with parameters captured directly from farm environments.

Model	Accuracy (%)	Error Rate (%)
Random Forest	100.00	00.00
Decision Tree	100.00	00.00
Support Vector Machine	94.68	05.32
Naive Bayes	91.84	08.16

The contrast is striking. With context-specific data, model accuracy surged—especially for ensemble classifiers like Random Forest and Decision Tree, which achieved perfect classification in the testing phase. This result highlights the immense value of real-time, location-relevant datasets in agritech applications.

Model Performance Comparison

Characteristic	Random Forest	Decision Tree	Support Vector Machine	Naive Bayes
Accuracy on Benchmark Dataset (%)	78.18	76.82	77.95	78.18
Accuracy on Real-Time Dataset (%)	100.00	100.00	94.68	91.84
Confusion Matrix Analysis	Zero false positives/negatives	N/A	Minor misclassification	Minor misclassification
Suitability	Optimal algorithm	N/A	N/A	N/A

Fig.5 Comparison of Model Performance

3.Confusion Matrix Analysis

The confusion matrix for the Random Forest model on the real-time dataset revealed zero false positives and false negatives, indicating flawless prediction capability under the given conditions. In contrast, SVM and Naive Bayes models showed minor misclassification across closely related crop categories—especially in crops with overlapping soil requirements. This clearly positions Random Forest as the most suitable algorithm, known for handling complex decision spaces through ensemble-based learning.

4. Interpretation of Results

These results illustrate a foundational principle in smart farming: Data quality and contextual relevance outweigh raw algorithmic complexity. While off-the-shelf datasets provide a convenient starting point, their lack of geographical specificity undermines model performance.

On the other hand, custom field data—gathered from actual soils, climates, and agricultural patterns—allows machine learning models to learn from reality, not assumption. This shift from theoretical to empirical learning significantly improves reliability and makes the system genuinely usable for farmers.

5. Field Relevance and User Response

Beyond numerical accuracy, the real value of the system lies in its field impact. During field trials, the system consistently delivered recommendations that aligned with best agricultural practices. In several cases, the system even suggested alternatives better suited to the soil's nutrient profile, encouraging more sustainable and profitable cropping choices.

Farmers responded positively, valuing the system's intuitive design, prompt recommendations, and dependable performance in varied field conditions. Such response validates the usability aspect, often overlooked in technology-driven research.

Localized training data dramatically enhances model accuracy and crop specificity.

Random Forest emerged as the optimal model due to its adaptability and ensemble learning.

Perfect classification was achieved on real-time sensor data, underscoring the system's reliability.

The transition from static to dynamic, sensor-fed datasets is essential for intelligent agriculture.

Farmer engagement and trust confirm the system's practical value—not just theoretical merit.

V. CONCLUSION AND FUTURE SCOPE

1. Conclusion

Agriculture, though ancient in its roots, is entering a transformative era—driven by data, automation, and intelligent systems. This project was a response to one of its most persistent challenges: How can farmers, with minimal resources, know what crop to grow for maximum yield and sustainability?

By designing a real-time, sensor-based crop recommendation system powered by machine learning, we have demonstrated a practical, scalable, and affordable solution to this problem. The system's ability to capture six vital soil parameters—N, P, K, pH, temperature, and humidity—and combine them with powerful algorithms like Random Forest allows it to make accurate and context-aware crop predictions.

With over 2,700 real-world entries from the Vidarbha region, the system has proven its reliability not just in controlled environments, but in the unpredictable and demanding conditions of actual farms. Achieving 100% accuracy on custom data and gaining real-world farmer approval reaffirms the system's effectiveness.

What sets this project apart is not only its technical merit, but its commitment to empowering the grassroots—making advanced technology usable and relevant for those who need it most. By simplifying the complexity of machine learning into a plug-and-play system for farmers, we've bridged the gap between innovation and implementation.

This project is more than an academic exercise; it is a blueprint for how AI and IoT can coexist with tradition, enrich farmer wisdom, and bring resilience to agriculture.

2. Future Scope

While this project lays a strong foundation, agriculture's complexity demands continuous evolution. Several promising directions can elevate this system to even greater impact:

1. GSM & SMS-Based Recommendations - In areas with no internet access, integrating a GSM module will allow farmers to receive crop suggestions via SMS, making the system even more inclusive.

2. Mobile App with Voice Support - Developing a mobile application in regional languages (e.g., Marathi, Hindi) with voice outputs will increase accessibility, especially for semi-literate users.

3. Fertilizer and Irrigation Scheduling - The same sensor data can be extended to recommend precise fertilizer dosage or irrigation frequency, promoting sustainable and resource-efficient farming.

4. Rainfall Prediction and Crop Insurance Advisory - By combining sensor data with weather APIs, the system could advise drought-resistant crops or alert farmers about potential losses, aiding in timely insurance claims.

5. Edge ML and Offline Predictions - Embedding trained models into local hardware devices (edge computing) will allow offline crop prediction, enabling remote usability without cloud dependency.

6. Expansion to Other Agro-Climatic Zones - Retraining the model with data from other regions will help replicate the system nationwide, turning it into a scalable digital public good.

3. Final Thought

Technology should not only be smart—it should be empathetic. This project proves that by blending intelligence with empathy, even small devices can make big changes in people's lives. In a world where farmers often face unpredictability, our system offers something rare: clarity, confidence, and control. As we look to the future, this work is not the end—but the beginning of a smarter, more connected era of agriculture, where every field speaks through data, and every farmer listens with insight.

VI. REFERENCES

[1] M. V. Korde, S. R. Varma, P. N. Bhalerao, T. T. Shaikh, and A. A. Sonawane, "Crop Recommendation System Using Machine Learning and IoT," IRJMETS, vol. 5, no. 3, Apr. 2024.

[2] O. Turgut, I. Kok, and S. Ozdemir, "AgroXAI: Explainable AI-Driven Crop Recommendation System for Agriculture 4.0," arXiv preprint arXiv:2412.16196, Dec. 2024.

[3] D. Garg and M. Alam, "An Effective Crop Recommendation Method Using Machine Learning Techniques," in Lecture Notes in Networks and Systems, Springer, 2022.

[4] G. Manju, K. S. S. Kishor, and V. A. Binson, "An IoT-Enabled Real-Time Crop Prediction System Using Soil Fertility Analysis," Smart Cities, vol. 5, no. 4, 2023.

[5] S. Hossain, N. N. I. Prova, M. R. Sadik, and A. A. Maruf, "Enhancing Crop Management: Ensemble Machine Learning for Real-Time Crop Recommendation System from Sensor Data," SSRN preprint, 2023. [O

[6] H. T. Gwalani, K. Mone, A. Kene, M. Parkhe, and D. Ingale, "Optimization of Crop Recommendation System Using Machine Learning," in Proc. 3rd Int. Conf. on Sustainable Computing, 2023.

[7] K. Anusha, D. J. Anusha, S. K. R., T. G. G., and K. K. K. M., "Crop Recommendation Using IoT and Machine Learning," Jain Institute of Technology, Karnataka, 2023.

[8] G. Deshpande, V. Khandagale, R. Doijad, et al., "IoT-Based Low-Cost Soil Moisture and Soil Temperature Monitoring System," PCCOE Pune, 2023.

[9] G. Singh and S. Sharma, "Enhancing Precision Agriculture Through Cloud-Based Transformative Crop Recommendation Model," Agricultural Informatics, vol. 4, no. 2, 2023.

[10] M. Hasan, M. A. Marjan, M. P. Uddin, et al., "Ensemble Machine Learning-Based Recommendation System for Effective Prediction of Suitable Agricultural Crop Cultivation," Computers and Electronics in Agriculture, vol. 210, 2024.

[11] N. Penchalaiah and J. N. Emmanuel, "IoT Based Smart Farming Using ThingSpeak and MATLAB," IJERT, vol. 11, no. 6, 2023.

[12] D. M. S. Reddy and U. R. Neerugatti, "A Comparative Analysis of Machine Learning Models for Crop Recommendation in India," J. of Agric. and Food Info., vol. 23, no. 1, 2024.

[13] S. Banerjee, A. Mukherjee, and S. Kamboj, "Precision Agriculture Revolution: Integrating Digital Twins and Advanced Crop Recommendation for Optimal Yield," in Proc. IEEE Conf. on Smart Technologies, 2023.

[14] S. Sam and S. M. D'Abreo, "Crop Recommendation with Machine Learning: Leveraging Environmental and Economic Factors for Optimal Crop Selection," International Journal of Computer Applications, vol. 185, no. 7, pp. 29–35, 2023.

[15] N. Wilberforce and J. Mwebaze, "A Framework for IoT-Enabled Smart Agriculture," Makerere University, Uganda, 2023.