Volume 9, Issue 3, pp. 81-84, 2025.

International Journal of Scientific Engineering and Science
ISSN (Online): 2456-7361

Real-Time IoT Sensor Networks Enhancing Precision Agriculture Data Collection

Seema J. Patil[1], Dr. Pallavi V. Baviskar[2]
[1]School of Computer Science and Engineering, Dr. Vishwanath Karad MIT-World Peace University, Pune, India - 411038
[2]Sandip Institute of Engineering and Management, Sandip Foundation, Nashik, India - 422213

Abstract—*The main difficulty in agriculture is picking the best crop depending on soil conditions in order to increase output and sustainability. This work solves this issue by using modern sensor technologies and machine learning algorithms to forecast the best crops for certain soil conditions. We used a 7-in-1 soil sensor linked to an Arduino UNO to capture real-time soil data, including pH, temperature, and moisture, across several geographic areas. The obtained data were evaluated using four different machine learning algorithms: Support Vector Machine (SVM), Decision Tree Classifier, Random Forest Classifier, and Gradient Boosting Classifier. Each model was assessed based on its ability to anticipate acceptable crops. The Random Forest Classifier had the highest accuracy (95%), followed by the Gradient Boosting Classifier (94%), the Decision Tree Classifier (93%), and the SVM (90%). Our findings indicate the viability of using machine learning techniques in precision agriculture to propose crop selections based on the soil's unique qualities. This technique can help farmers make better decisions, optimize agricultural yields, and promote sustainable farming practices. This paper also examines the early hurdles, such as merging sensor data with predictive models, and how they were solved in order to create a reliable prediction system.*

Keywords— *Internet of things, Soil Moisture, Agriculture, Real-Time data, Machine Learning*

I. INTRODUCTION

Maximizing agricultural yields through optimal crop selection based on soil parameters is a key problem in precision agriculture. The capacity to forecast which crops are most suited to given soil conditions may dramatically boost agricultural output and sustainability. This issue is highlighted by rising food demand as the world's population grows, as well as the pressing need for ecologically friendly farming techniques.

Current crop selection procedures are mostly based on traditional agronomic practices that rely on broad parameters and farmer experience. These approaches, while useful, frequently lack the accuracy needed to account for micro-variations in soil qualities across various plots of land. Furthermore, conventional techniques [1] fail to take use of current advances in sensor technology and data analytics, which may give extensive insights about soil health and crop compatibility.

Recent efforts to incorporate technology into agriculture have resulted in the adoption of basic computer models and simple data analysis tools. However, these efforts are sometimes impeded by a failure to integrate real-time soil data collecting with modern analytical models. Furthermore, present models frequently ignore the intricate interdependence of multiple soil characteristics, resulting in inefficient crop recommendations that may not maximize production or sustainability.

In response to these restrictions, our study seeks to close the gap by utilizing a sophisticated collection of machine learning algorithms capable of interpreting complicated information collected from modern soil sensors. This work not only improves crop forecast accuracy, but it also advances the area of smart farming by incorporating data-driven insights directly into actual agricultural decision-making processes.

II. MATERIALS AND METHODS

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it.

A. Data Collection

The study was conducted across various agricultural fields characterized by diverse soil types to ensure comprehensive data representation. Soil samples were collected using a systematic grid-sampling method to cover different plots uniformly. For real-time data acquisition [2], we employed a 7-in-1 soil sensor module connected via a TTL to RS485 module to an Arduino UNO. This setup enabled the measurement of critical soil parameters including pH, moisture content, temperature, and more, which are crucial for determining soil health and suitability for crop cultivation.

Figure 1. 7-in-1 soil sensor

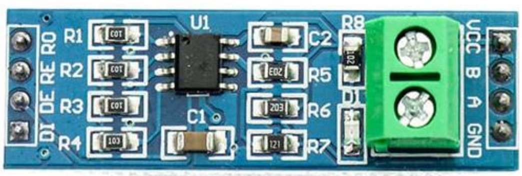

Figure 2. TTL to RS485 module

Volume 9, Issue 3, pp. 81-84, 2025.
ISSN (Online): 2456-7361

B. Implementation part

The connection flow is something like:

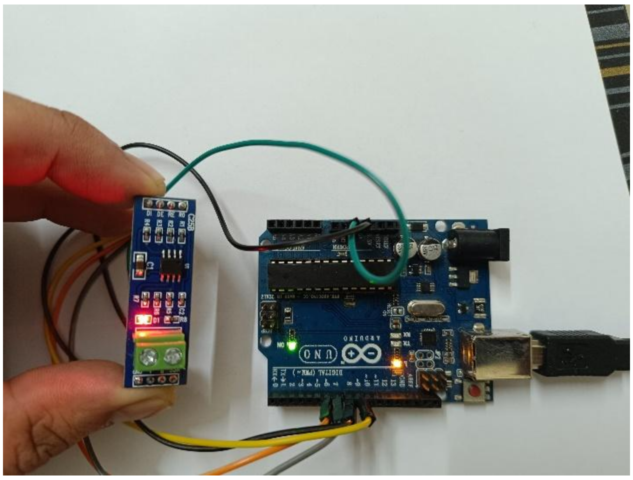

Figure 3. Arduino Uno Connections

The connection for pins are as follow:

- From ARDUINO UNO TO RS485

TABLE 1. ARDUINO UNO TO RS485

Arduino PINs	RS485 PINs
~6	RE
7	DE
8	RO
~9	DL
VCC	5V
GND	GND

- RS 485 TO SOIL 7 IN 1 SENSOR

TABLE 2. RS485 TO SOIL 7 IN 1 SENSOR CONNECTIONS

RS 485	SOIL 7 IN 1
A	A
B	B

For powering the soil sensor we have used a 12V adapter and connected the pins of the sensors with the positive and negative wires respectively.

From this site we have taken the code and modified it like made changes such as added the max number of readings and added a option such that it the output should be saved in a excel file, here is the base version of the code [3] feel free to modify the code.

We have created a sample environment in which we have tested on a small scale such that we can get an idea about what challenges we will be facing if we used this model for a large field.

C. Data Analysis

Due to the sophisticated capabilities of the 7-in-1 soil sensor [4], which includes built-in pre-processing features for noise reduction and data normalization, no additional pre-processing steps were required. The raw data collected were directly fed into the analysis pipeline. We utilized Python as the primary programming language for data manipulation and analysis, employing libraries such as Pandas for data handling and Scikit-learn for machine learning model implementation.

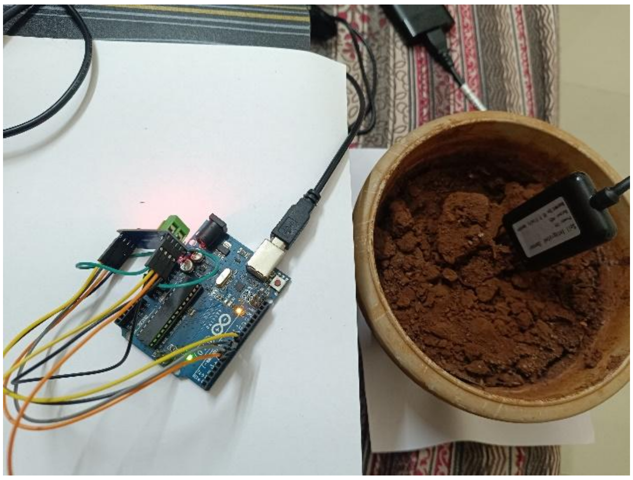

Figure 4. Sample Environment

D. Model Development

We implemented four different machine learning algorithms to predict the most suitable crops for the sampled soils:

- Support Vector Machine (SVM): Chosen for its effectiveness in high-dimensional spaces and its capacity to model non-linear decision boundaries.[5]
- Decision Tree Classifier: Selected for its interpretability and ease of use in understanding how decisions are made.
- Random Forest Classifier: Utilized for its robustness and effectiveness in handling overfitting, making it ideal for complex datasets.[6]
- Gradient Boosting Classifier: Employed for its ability to optimize on different loss functions and to provide superior predictive accuracy.[7]

Each model was trained using a split of 80% training data and 20% testing data. The training process involved parameter tuning through grid search to optimize each model's performance. Model evaluation was primarily based on accuracy, recall, and precision metrics, calculated using cross-validation techniques [8] to ensure the reliability of the predictions.

III. RESULTS

In this study, we evaluated the performance of four different machine learning algorithms to predict the suitability of various crops based on soil characteristics. The accuracy of each model was calculated to determine its effectiveness in making precise predictions [9]. Here are the summarized results:

- Support Vector Machine (SVM): Achieved an accuracy of 90%.
- Decision Tree Classifier: Reached an accuracy of 93%.
- Random Forest Classifier: Emerged as the most effective, with an accuracy of 95%.
- Gradient Boosting Classifier: Recorded an accuracy of 94%.

International Journal of Scientific Engineering and Science
Volume 9, Issue 3, pp. 81-84, 2025.
ISSN (Online): 2456-7361

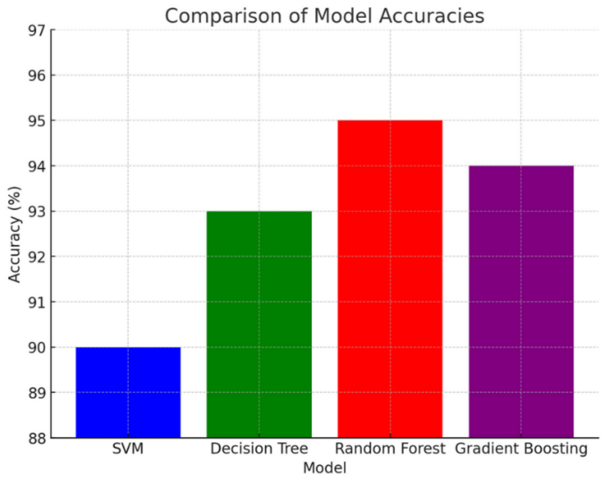

Figure 3. Comparison of Model Accuracies

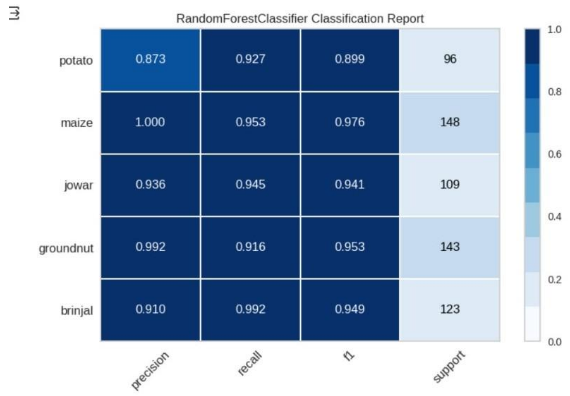

Figure 4. Random Forest Classification Report

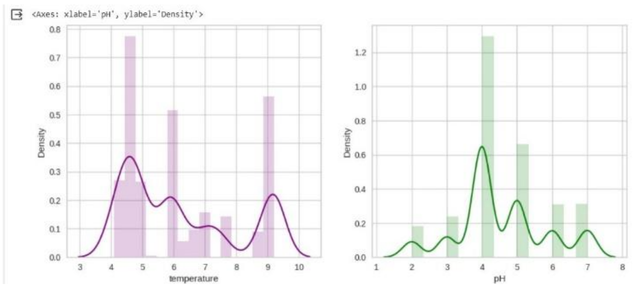

Figure 5. Temperature V/S pH

Discussion of Results

The Random Forest Classifier's higher performance implies that ensemble approaches, which combine the results of several learning algorithms, produce more trustworthy predictions by lowering the danger of overfitting associated with individual decision trees. This is especially useful in agricultural applications, where inaccurate forecasts may be costly, hurting crop output and economic return.

The Gradient Boosting Classifier was also highly effective, most likely because to its capacity to focus on fixing earlier trees' flaws, so gradually refining the model. This model's performance demonstrates its ability to handle complicated agricultural datasets with various interacting elements that impact the outcome.

These findings suggest that machine learning might considerably improve agricultural decision-making by giving accurate, data-driven insights about crop suitability based on individual soil conditions [9]. This has the potential to transform traditional farming processes by increasing accuracy and efficiency, resulting in better resource use and improved agricultural yield.

IV. Discussion

A. Interpretation of Model Performance

The findings of this study demonstrate the possibilities of modern machine learning approaches in predicting viable crops depending on soil conditions. The Random Forest Classifier achieved the best accuracy (95%), owing to its ensemble learning strategy [11}, which mixes many decision trees to decrease variance and prevent overfitting. The performance of this model implies that it is especially useful when dealing with complicated datasets with nonlinear variable correlations and multidimensional interactions.

- The Gradient Boosting Classifier followed closely with 94% accuracy. This model's performance is likely due to its sequential approach, where each new tree is built to correct errors made by previously built trees. This method is effective in optimizing performance across a range of conditions, gradually improving prediction accuracy through iterations.
- The Decision Tree Classifier, with an accuracy of 93%, offers a more interpretable model compared to the others. Despite its simplicity, it performed admirably, which underscores the importance of feature selection and the inherent power of the model to handle agricultural data.
- The Support Vector Machine (SVM), although slightly less accurate at 90%, is known for its effectiveness in high-dimensional spaces, which implies it could still be very useful in scenarios where the number of soil parameters is significantly high.

B. Discussion on Outcomes

The practical ramifications of these discoveries are significant. Using these prediction models, agricultural stakeholders may make better judgments about which crops to plant in which places based on real-time soil data. This strategy can result in various benefits: The practical ramifications of these discoveries are significant. Using these prediction models, agricultural stakeholders may make better judgments about which crops to plant in which places based on real-time soil data. This strategy can result in various benefits:

- Increased Crop Yields: By matching crops to the most suitable soil conditions, farmers can maximize the potential output of their lands.
- Resource Efficiency: Optimal crop selection helps in the efficient use of resources such as water and fertilizers, which is crucial for sustainable agricultural practices.
- Risk Reduction: Advanced prediction models can reduce the risk of crop failure due to unsuitable soil conditions, thereby increasing the predictability and stability of agricultural output.

Furthermore, incorporating such models into agricultural decision-making processes represents a shift toward precision agriculture, in which each farming choice is data-driven and adapted to unique environmental circumstances. This move not only helps to optimize agricultural productivity but also to preserve the environment by eliminating wasteful resource consumption and the ecological imprint of farming operations.

V. CHALLENGES

Our study's execution faced various initial hurdles, the most significant of which were linked to code and hardware configuration. Integrating the 7-in-1 soil sensor with the Arduino UNO and maintaining steady data transfer were key challenges. The complexity of the setup necessary to correctly record real-time soil data caused the initial sensor connection challenges. The code originally written for data acquisition occasionally fails to handle errors generated by sensor disconnections or data corruption during transmission.

These technological concerns were solved in a methodical manner by improving the connection protocols and the resilience of our data gathering programs. We included error-handling methods into our program to control disruptions and assure uninterrupted data flow. In addition, firmware upgrades were deployed to the sensors, and the physical connections were improved with stronger shielding and grounding to reduce data loss and interference.

Moreover, the sensors says that it need a power source of 5V but in reality it will be needing a power supply of 12V otherwise it won't work and won't give the desired output.

VI. LIMITATIONS AND FUTURE RESEARCH

While our research has produced useful insights, it contains limits that open the way for future research. One key drawback is the reliance on specialized hardware, which may not be easily accessible or cost-effective for all agricultural stakeholders. The models' performance, particularly in different and less controlled conditions, may differ, which was not completely investigated due to the homogeneity of our test fields.

Further study might involve testing the models in a range of agricultural situations to assess their resilience and adaptability. Furthermore, using satellite images and other remote sensing data may improve the models' forecasting potential by giving more extensive environmental information. Future research should also look at including more different crop varieties and soil conditions to widen the applicability of the results.

Finally, creating a user-friendly interface for non-technical users might greatly improve the practical use of these models in common agricultural activities. This would assist to bridge the gap between cutting-edge agricultural research and practical, on-the-ground farming operations, making advanced analytics available to a wider variety of farmers.

VII. CONCLUSION

This study highlighted the great potential of combining real-time IoT sensor networks with powerful machine learning algorithms to improve precision agriculture. We were able to precisely identify the most suited crops for certain soil conditions by deploying a 7-in-1 soil sensor and then analyzing the data using sophisticated machine learning algorithms. This technique not only improves agricultural production projections, but it also promotes sustainable farming practices by allowing for better educated crop selection using real-time data.

Our findings show that technologies like the Random Forest Classifier and Gradient Boosting Classifier may significantly enhance agricultural decision-making by offering accurate, data-driven insights that conventional approaches cannot.

However, this study has limitations, including the necessity for specific hardware and the inherent difficulties of integrating high-tech solutions in varied agricultural contexts. Future research should focus on improving these technologies, making them more accessible and cost-effective for farmers throughout the world, and broadening their applicability to encompass a wider range of crops and soil types.

Finally, the integration of IoT and machine learning in agriculture presents a viable road to more efficient and sustainable farming techniques. Precision agriculture, by using the power of real-time data and sophisticated analytics, can result in more resilient agricultural systems that are better suited to the challenges of a changing global environment.

REFERENCES

[1]: James Sumberg, Ken E. Giller,What is "conventional' agriculture?,Global Food Security,Volume 32,2022, 100617,ISSN-2211-9124, https://doi.org/10.1016/j.gfs.2022.100617

[2]: Sarma, Pratiksha, Hidam Kumarjit Singh, and Tulshi Bezboruah. "A real-time data acquisition system for monitoring sensor data." *Int. J. Comput. Sci. Eng* 6.6 (2018): 539-542.

[3] Pemasiri, P. R. B., Rathnayaka, R. A. A. S., Jayaneththi, J. P. H. U., Sirisena, U. G. A. I., & Sooriyaarachchi,.Suitability of the 7 in 1 Integrated Sensor for Measuring N,P, K Levels in ALF Soils in S. J. 2022, SRI LANKA.

[4] Pisner, Derek A., and David M. Schnyer. "Support vector machine." *Machine learning*. Academic Press, 2020. 101-121.

[5] Pal, Mahesh. "Random Forest Classifier for Remote Sensing Classification.", *International journal of remote sensing* 26.1 (2005): 217-222.

[6] Bentéjac, Candice, Anna Csörgő, and Gonzalo Martínez-Muñoz. "A Comparative Analysis of Gradient Boosting Algorithms." ,*Artificial Intelligence Review* 54 (2021): 1937-1967.

[7] Tougui, Ilias, Abdelilah Jilbab, and Jamal El Mhamdi. "Impact of the Choice of Cross-Validation Techniques on the Results of Machine Learning-based Diagnostic Applications." ,*Healthcare informatics research* 27.3 (2021): 189.

[8] Liu, Zhijian, et al. "Accuracy Analyses and Model Comparison of Machine Learning Adopted in Building Energy Consumption Prediction." ,*Energy Exploration & Exploitation* 37.4 (2019): 1426-1451.

[9] Beriya, Abhishek, and V. N. Saroja. *Data-Driven Decision Making for Smart Agriculture*. No. 8. ICT India Working Paper, 2019.

[10] Zhou, Zhi-Hua, and Zhi-Hua Zhou. *Ensemble Learning*. Springer Singapore, 2021.