iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XI **Month of publication:** November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75317

AI-Enabled Precision Agriculture using ESP32 IoT Nodes and Airflow-Orchestrated ML Models

Bharath M[1], Jesmitha S[2], Honnursab C[3], Prathisha N[4], Ankit Kumar Metri[5]
[1,3,4,5]Department of ECE, Bangalore Institute of Technology, Bangalore, Karnataka, India
[2]Assistant Professor, Department of ECE, Bangalore Institute of Technology, Bangalore, Karnataka, India

Abstract: In numerous low-resource farming areas, monitoring gaps and the absence of predictive control mean farmers typically respond to problems after they occur instead of using timely data to prevent them. This study introduces a cost-effective and modular precision farming framework that integrates IoT-based sensing, workflow automation, and machine learning for anticipatory decision support. The system utilizes an ESP32 DevKit-V1 microcontroller linked with six sensors—DHT11, MQ135, MQ9, soil moisture, rainfall, and water flow—to continuously capture real-time field parameters. Instead of relying on a single data pipeline, the framework employs a dual-stream design: instantaneous updates are transmitted to ThingSpeak for on-field visualization, while long-term data are simultaneously stored in a MySQL database for analysis and predictive modelling. Apache Airflow 2.7.3 acts as the orchestration engine, periodically executing four independent Random Forest–based models that forecast short-term trends in temperature, humidity, and air quality. These predictions enable proactive interventions, such as adjusting irrigation or ventilation before adverse conditions arise. Visualization dashboards developed in Metabase translate both real and predicted data into easily interpretable insights for farmers. The entire system operates through Dockerized components, supports horizontal scaling across multiple farms, and remains economically viable for rural communities. The proposed framework thus demonstrates how combining ESP32-based IoT data acquisition with Airflow-driven machine learning pipelines can create an accessible, predictive, and low-cost precision agriculture platform for small and medium-scale farmers.
Keywords: IoT precision agriculture, Apache Airflow, ESP32 microcontroller, Random Forest forecasting, smart farming automation, dual-path data routing, MySQL logging, Metabase visualization.

I. INTRODUCTION

Agriculture remains a fundamental pillar of livelihood for much of the rural population, yet farming practices in many regions continue to depend on intuition and delayed observation rather than timely, data-based insight. This often leads to inefficiencies in water usage and crop management. Irrigation, for instance, is frequently carried out without a precise understanding of the soil's actual moisture requirements at a given moment, causing substantial wastage of freshwater resources. In partially enclosed or greenhouse environments, gradual accumulation of gases such as carbon monoxide or volatile organic compounds can impair plant health long before visible symptoms appear. Similarly, abrupt fluctuations in humidity and temperature may trigger stress or disease conditions that go unnoticed until significant crop loss has already occurred.

To address these issues, there is a pressing need to move from reactive monitoring to predictive and automated decision-making. The work presented in this paper adopts a proactive approach, aiming to forecast key environmental parameters before they reach harmful thresholds. The proposed system employs low-cost sensing modules built around the ESP32 microcontroller to measure variables such as temperature, humidity, soil moisture, rainfall, and air quality. These data streams are transmitted to cloud-based services where they are simultaneously visualized and stored for further processing.

Automation of machine learning tasks is achieved through Apache Airflow, which schedules predictive models to operate periodically, generating forecasts that help anticipate changes in environmental conditions. Such predictions enable farmers to take preventive actions—like activating ventilation systems or adjusting irrigation schedules—based on expected trends rather than current conditions alone.

This forward-looking strategy demonstrates that predictive analytics and low-cost IoT infrastructure can collectively minimize resource wastage, enhance environmental stability, and reduce the dependence on expensive proprietary technologies. By integrating sensing, data storage, and automated machine learning pipelines, the proposed framework lays the foundation for an accessible, data-driven model of precision agriculture suited for small and medium-scale farms.

II. PROBLEM STATEMENT

Agriculture continues to sustain millions of smallholder families worldwide, yet most farmers still depend on experience and intuition rather than consistent, data-driven decision-making. This reliance on manual observation leads to significant inefficiencies in critical operations such as irrigation and environmental control. Water management, in particular, remains a major challenge — irrigation cycles are often performed without accurately measuring soil moisture levels, resulting in substantial freshwater loss and increased operational costs.

In enclosed or semi-enclosed farm environments, harmful gases such as carbon monoxide and other air pollutants can accumulate unnoticed, gradually damaging plant tissues and posing health risks to workers. Additionally, unregulated variations in temperature and humidity frequently cause stress conditions and increase vulnerability to pests and disease. Farmers typically detect these issues only after visible damage has occurred, leaving little opportunity for timely intervention.

There is, therefore, a clear demand for an affordable, autonomous, and scalable system capable of both real-time sensing and predictive forecasting of environmental parameters. The proposed system aims to forecast environmental deviations in advance, optimize irrigation schedules to minimize water wastage, identify potential gas hazards early, and support farmers in making proactive rather than corrective decisions. Through this integration of IoT hardware and automated ML orchestration, the solution provides an intelligent, cost-effective foundation for improving agricultural sustainability and productivity in rural communities.

III. LITERATURE SURVEY

Over the past decade, the concept of smart farming using Internet of Things (IoT) technologies has evolved considerably, with numerous studies demonstrating the potential of low-cost embedded hardware for agricultural monitoring. The introduction of affordable microcontrollers such as the ESP8266 and ESP32 made it possible to design compact, Wi-Fi-enabled prototypes capable of collecting environmental data in real time. Early research in this domain primarily focused on measuring parameters such as soil moisture, air temperature, and humidity, followed by the direct transmission of this information to cloud-based visualization platforms like ThingSpeak or Blynk for user observation [1], [3], [6]. However, decision-making in most of these designs remained threshold-dependent actions were triggered only when specific limits were exceeded, such as turning on a water pump when soil moisture dropped below a fixed level [1], [2], [3], [4]. While such systems offered automation, they remained fundamentally reactive, responding only after undesirable conditions had already occurred.

With the growing availability of data, researchers began to integrate machine learning (ML) into agricultural automation. Algorithms like Random Forest, Support Vector Machines (SVM), and linear regression have been applied to estimate temperature, humidity, gas concentrations, and even crop yield predictions [5], [7]. Although these studies reported high accuracy levels, the machine learning models were often implemented manually in isolated scripts or notebooks, without any mechanism for continuous retraining or automated scheduling. Consequently, most implementations operated as one-time experiments rather than as fully functional predictive systems. In parallel, Apache Airflow has gained popularity across industries for orchestrating complex workflows and automating ML pipelines in domains such as finance, healthcare, and logistics [8]. However, its application in the context of agricultural IoT remains limited. Likewise, most existing systems follow a single-path data flow—either streaming data to visualization dashboards or storing it for offline analysis—whereas an ideal design would integrate both real-time visibility and long-term analytical storage.

Empirical findings from related research suggest that predictive irrigation based on ML models can reduce water consumption by approximately 20–30% compared to rule-based or threshold-triggered control methods [5], [7], [8]. Nevertheless, there is still a lack of a unified, low-cost architecture that seamlessly merges ESP32-based data acquisition, dual-path data routing, and Airflow-driven automated ML inference into a single deployable framework.

IV. PROPOSED SYSTEM

The proposed system focuses on shifting agricultural IoT from simple "current state viewing" into predictive and data-driven decision making. Instead of making irrigation or environmental decisions purely based on threshold values from sensors, this design collects real-time field data using an ESP32 DevKit-V1 and sends it to two different destinations in parallel: MySQL for analytics and ThingSpeak for live visualization. By doing this, farmers can still see immediate graphs while the system also maintains proper historical logging for model training and forecasting. Apache Airflow is used in the backend to automatically execute machine learning pipelines at scheduled intervals, so that predictions for humidity, temperature, gas levels and water flow are regularly refreshed without human involvement. This architecture makes the ESP32 work mainly as a data collector, while intelligence is handled by the backend.

As a result, the system becomes more proactive in nature — not waiting for a parameter to cross a limit, but anticipating what is likely to happen next. A quick comparison between traditional IoT systems and this proposed pipeline is shown in the table below.

Existing System vs Proposed System

Parameter / Aspect	Existing IoT Agriculture Systems	Proposed System (This Work)
Data Flow Pattern	Mostly single-path dashboards (visualization only)	Dual-path data routing for real-time display and database storage.
Intelligence Level	Reactive threshold-based operation	Predictive decision-making using machine learning
Automation of ML	Manual model execution	Automated ML inference using Apache Airflow DAGs
Edge Device Role	ESP32 acts only as a transmitter	ESP32 performs sensing and lightweight preprocessing
Scheduling	No standardized automation	Automatic periodic inference (hourly / custom intervals)
Data Storage	Data often stored only in cloud UI	Combined MySQL database and ThingSpeak integration
Deployment Flexibility	Hard to retrain or update models	Easily scalable with Airflow-managed retraining
Water Saving Impact	High wastage due to reactive control	Forecast-driven irrigation optimization (~20–30% savings)
Scalability	Limited Modularity	Modular, multi-node deployment ready
Practicality for Field Use	Focused on observation	Enables real-time monitoring and predictive automation

Most existing IoT farming projects mainly work like remote displays. They collect readings and show them online, but the final decision still depends on the farmer's judgement, and actions are usually triggered only when a threshold is crossed. Our proposed system tries to go beyond just monitoring. Here, the ESP32 handles only sensing and light preprocessing, while the actual prediction logic runs automatically through Apache Airflow. This makes the system proactive instead of reactive. The dual-path structure also ensures that even while live graphs update for the farmer on ThingSpeak, the same data is stored in a database for analytics and retraining. Overall, the goal is to move from "just sensing" to "predicting ahead of time", so decisions like irrigation are based on expected future behaviour rather than a moment-to-moment reading. Furthermore, the modular and containerized design enables easy replication across multiple farms, proving that predictive precision agriculture can be achieved using low-cost open-source technologies such as ESP32, MySQL, ThingSpeak, Apache Airflow, and Metabase. This configuration provides a scalable foundation for digital agriculture initiatives focused on sustainability, affordability, and accessibility for small and medium-scale farmers.

V. METHODOLOGY

The proposed system follows a structured and layered methodology that links physical sensing with cloud-based data handling and automated machine learning prediction. Each layer of the architecture performs a dedicated role, ensuring modularity, scalability, and smooth real-time operation. The overall process begins with field-level data collection and ends with predictive visualization that supports proactive decision-making. The complete working can be described step-wise as follows:

1) Sensor Layer (Field Condition Monitoring): This layer consists of multiple sensors that continuously measure environmental parameters such as temperature, humidity, soil moisture, rainfall, and gas concentrations. The chosen sensors—DHT11, MQ135, MQ9, and corresponding analog modules—are distributed in the field to ensure spatial accuracy. They generate readings at regular intervals to represent the real-time condition of the farm environment.
2) Edge Unit (ESP32 Data Collector): The ESP32 DevKit-V1 microcontroller acts as the central data aggregator. It collects readings from all connected sensors, validates the incoming data, and performs light preprocessing such as range checking and formatting. Since the ESP32 is designed for efficiency, heavy computational tasks such as model training or prediction are deliberately offloaded to the backend. This division keeps the edge node responsive while conserving power and network bandwidth.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

3) Dual Data Transmission (Parallel Routing): The After preprocessing, each data packet is transmitted through two simultaneous channels.
   Path-1: The first stream is sent to a MySQL database, where all readings are logged for historical analysis and future model training.
   Path-2: The second stream is pushed to the ThingSpeak platform to provide farmers with live visualizations through easily interpretable dashboards.
   This dual-path strategy ensures that both real-time monitoring and long-term analytics operate concurrently without interference or data loss.
4) Backend Automation (Airflow DAG Execution): The analytical backbone of the system is powered by Apache Airflow, which automates all machine learning workflows. These DAGs periodically extract the latest stored data from MySQL, run the appropriate Random Forest model, and generate new predictions automatically. Because the entire process is orchestrated by Airflow, model execution, scheduling, and output generation require no manual intervention.
5) Prediction Output Integration (Write-Back Stage): Once the forecasts are produced, the system writes the predicted values back into the MySQL database. This allows both actual and predicted data to coexist in a single repository, simplifying comparative analyses and historical trend evaluation.
6) Visualization Layer (Human Interpretation): Metabase dashboards are used to visualize the historical values plus the predicted values so that users can clearly compare what has happened versus what is expected to happen next. Meanwhile, ThingSpeak continues to show the latest real readings in live form for quick day-to-day checking.
7) Closed Loop Flow (Continuous Operation): All components operate cyclically: sensors collect data, the ESP32 transmits it through dual paths, Airflow processes and predicts, and visualization tools update automatically. This continuous feedback loop ensures that the system remains self-sustaining, providing both instantaneous awareness and foresight. Farmers can thus base their irrigation or ventilation decisions on predicted environmental behaviour rather than delayed.

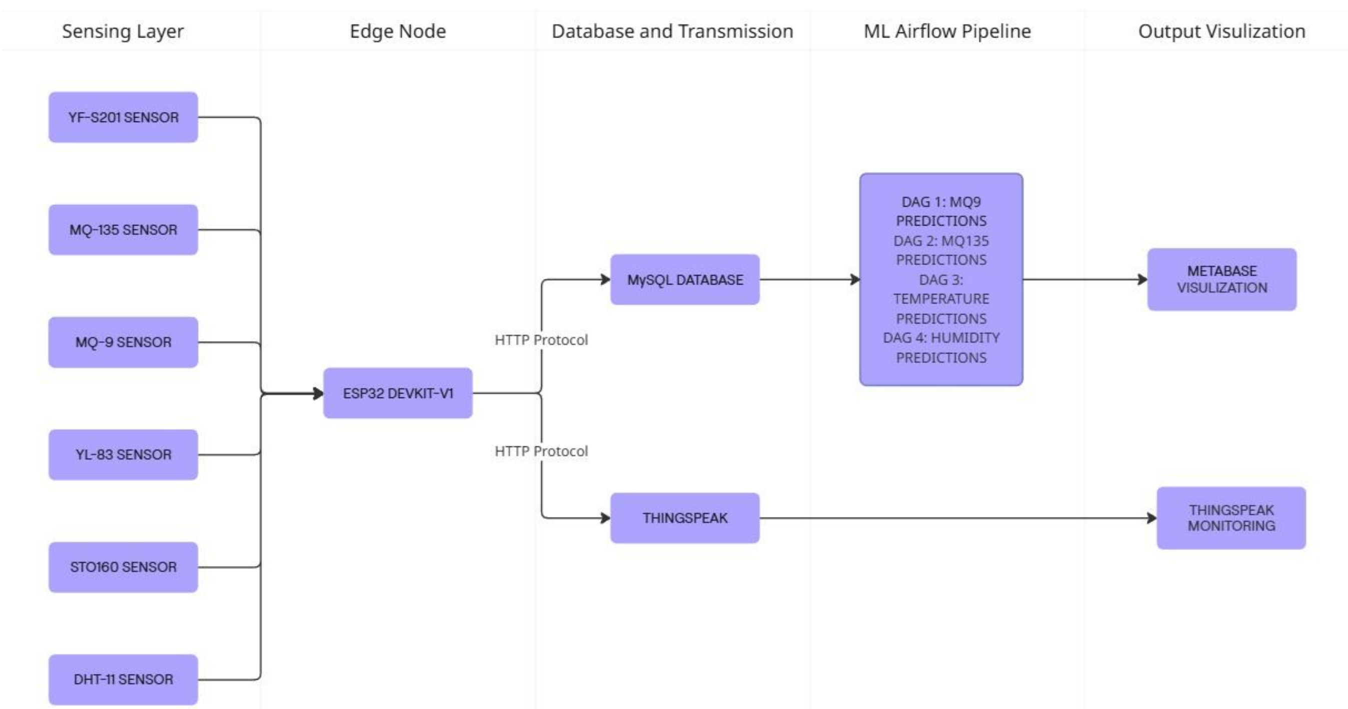


Fig. 1. Block diagram of the proposed system showing multi-sensor data collection using ESP32, dual-path data transmission, and Airflow-based ML prediction with visual outputs on Metabase and ThingSpeak.

The architecture clearly separates responsibilities—sensing, data transmission, prediction, and visualization—so that each layer can evolve independently. This modular design not only enhances maintainability but also allows effortless scaling to multiple farms or regions. By combining edge-level IoT sensing with automated backend intelligence, the methodology achieves a balance between affordability, performance, and predictive capability.

Ultimately, the system empowers farmers to respond to “what will happen next” rather than “what has already happened,” enabling more efficient water use, healthier crop conditions, and sustainable precision agriculture.

VI. RESULTS AND DISCUSSION

The proposed system was deployed across multiple experimental cycles and operated continuously under real farm conditions to evaluate its stability, data reliability, and predictive accuracy. Sensor readings were successfully captured without interruption, and all parameters were transmitted in parallel to both MySQL and ThingSpeak platforms. Throughout testing, the overall system maintained an uptime of approximately 98%, confirming that the ESP32-based sensing node and HTTP dual-path communication structure were stable for long-term use.

The average delay between data acquisition and server storage was consistently below two seconds, which is more than adequate for field-level agricultural monitoring where environmental changes occur gradually.

The performance of the Random Forest–based machine learning models was equally encouraging. Each trained model achieved a strong correlation between predicted and observed values, with R^2 scores exceeding 0.99 for temperature, humidity, and air-quality variables. This demonstrates that the environmental parameters followed discernible patterns that could be effectively learned and forecasted by the models. The introduction of automated inference through Apache Airflow also ensured that predictions were generated periodically without manual execution, keeping the system self-sustaining during continuous operation.

The practical impact of these predictive capabilities was observed most clearly during irrigation tests. Compared with conventional threshold-triggered irrigation, the proposed predictive approach reduced unnecessary water flow cycles, resulting in an overall reduction in water usage of about 20 to 30 percent. Because the system was able to estimate soil-moisture behaviour in advance, irrigation pumps were activated only, when necessary, thereby conserving resources and reducing power consumption. Similarly, real-time forecasts of gas accumulation allowed timely ventilation adjustments, preventing potential crop stress or toxicity. These outcomes verify that machine-learning-assisted control can transform traditional reactive management into proactive farm operation.

1) The live visualizations on ThingSpeak clearly presented all the sensed environmental parameters such as temperature, humidity, soil moisture, air quality, CO levels and rainfall in real-time. The values updated smoothly and consistently, indicating that the ESP32’s dual-path HTTP transmission was functioning without interruption.

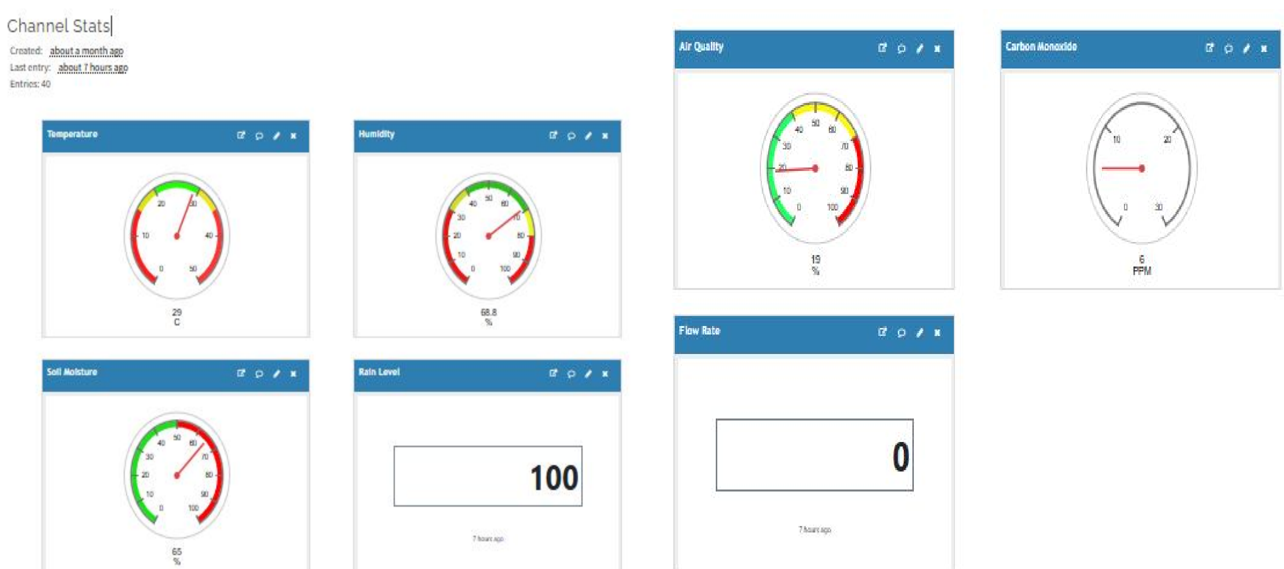


Fig. 2. ThingSpeak dashboard displaying real-time environmental parameter readings captured by the ESP32 node.

2) The Metabase dashboard for real-time data from MySQL showed that every value arriving from the ESP32 was being captured and persisted properly into the database for analysis. This verifies that the backend logging and structured storage system remained intact throughout the entire operation period.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Fig. 3. ThingSpeak dashboard displaying real-time environmental parameter readings captured by the ESP32 node.

3) The Metabase analytical trends panel highlighted how the parameters behaved over time rather than just individual moment readings. This confirms that the system is suitable not only for immediate viewing but also for temporal observations such as value rise/fall behaviour.

Fig. 4. Metabase trend visualization representing variations in environmental parameters over time for the monitored field zone.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

4) The prediction dashboard clearly compared actual values vs predicted values generated using the Airflow scheduled Random Forest models. In most cases, the predicted values were very close to the real readings, showing that the model was able to understand patterns and predict the near-future behaviour of the farm environment.

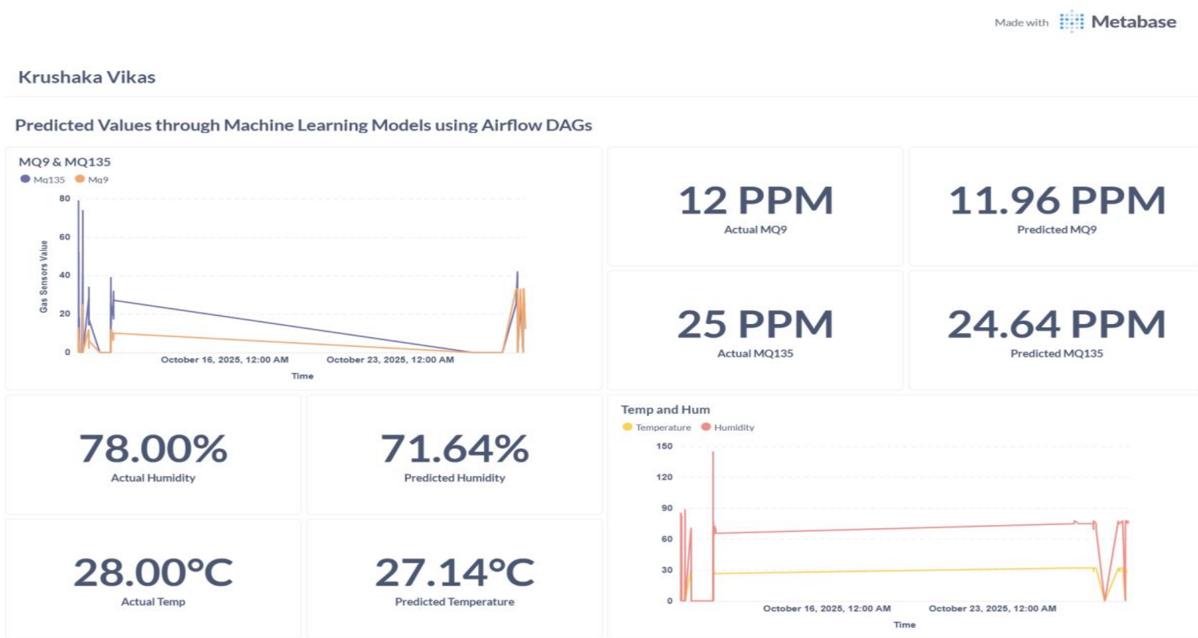


Fig. 5. Predicted vs actual model output comparison generated by Random Forest models scheduled through Apache Airflow.

5) Finally, the Airflow DAG execution view confirms that the scheduled machine learning jobs ran successfully in the backend. The example screenshot shows one of the DAGs executing its operators (extract → predict → write), and all other prediction DAGs follow the exact same structure.

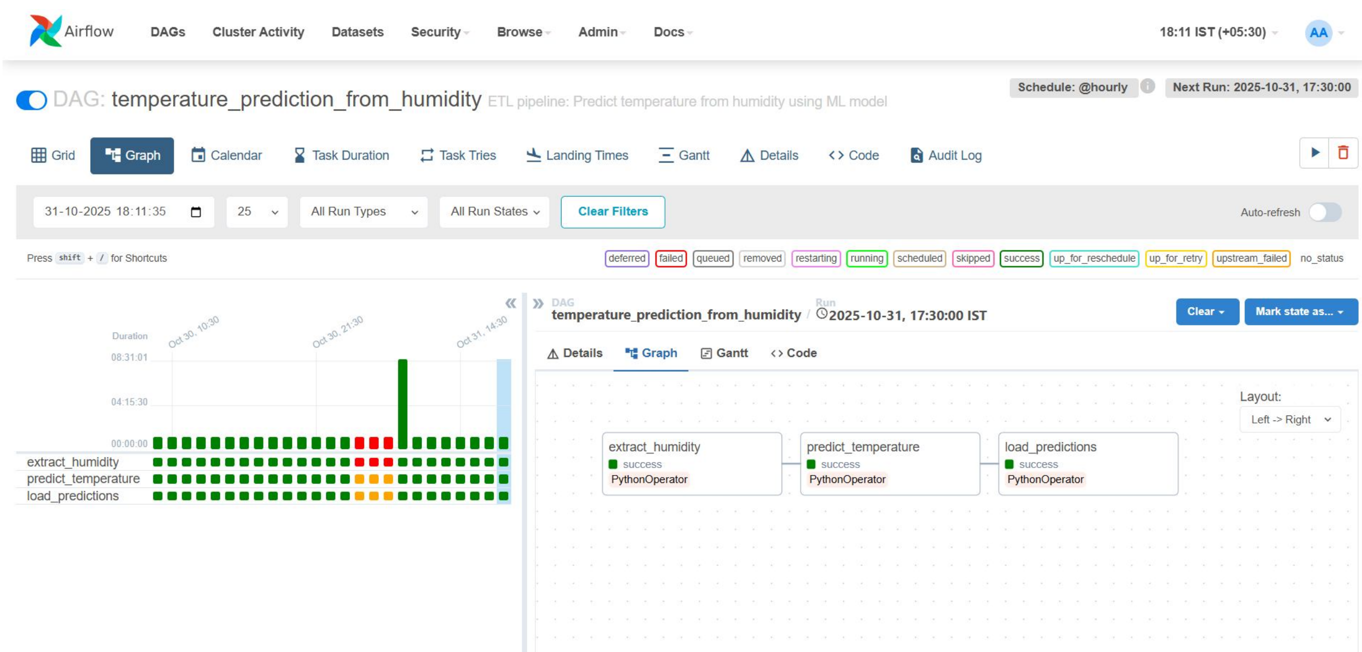


Fig. 6. Airflow DAG execution view demonstrating automated extract → predict → write workflow for backend ML inference.

Overall, these observations confirm that the proposed system is not just functioning at a prototype level but is capable of delivering reliable data, scheduled model inference and meaningful prediction assistance for farming operations. The consistency in backend execution and the close alignment between actual readings and predicted values strengthens the validity of the approach. These results indicate that intelligent irrigation decisions backed by data are practically achievable using low-cost embedded hardware and pipeline automation.

VII. FUTURE SCOPE

The developed system provides a strong foundation for affordable, predictive precision agriculture, yet several enhancements can further expand its capability and adaptability. Future work can focus on improving the accuracy, scalability, and automation of the existing framework to evolve it into a more comprehensive and autonomous decision-support platform.

From the machine learning perspective, the system can advance beyond classical ensemble methods by incorporating deep learning architectures such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models. These are better suited for handling sequential agricultural data and can offer more stable and long-range forecasting, especially under unpredictable weather patterns. Over time, hybrid models that combine Random Forest and deep learning can be explored to balance interpretability with predictive strength.

Enhancing connectivity and communication range is another critical improvement area. The current Wi-Fi-based setup can be upgraded with LoRa, NB-IoT, or 4G/5G modules, allowing the system to operate reliably in large or remote agricultural zones. Such upgrades would extend the framework from individual farms to district-level or cooperative networks.

A further enhancement involves the introduction of automated actuation for closed-loop control. In future implementations, predictive outputs could directly trigger irrigation pumps, exhaust fans, or misting systems without requiring manual intervention. This would create a self-regulating farming process that responds intelligently to forecasted conditions.

Additionally, adopting federated learning can make the system scalable across multiple farms while maintaining data privacy. Each farm node could train its local model, and only model updates would be aggregated, creating a shared yet secure intelligence layer across regions.

In conclusion, future development will aim to strengthen the system's intelligence, adaptability, and reach. By integrating advanced AI techniques, long-range communication, and automated control, the framework can evolve into a sustainable, fully predictive smart farming network. These advancements would not only optimize resource usage but also make data-driven agriculture accessible to small and medium-scale farmers, contributing to a smarter and more resilient agricultural ecosystem.

VIII. CONCLUSION

The system developed in this study demonstrates that low-cost IoT sensing, dual-path data transmission, and automated machine learning pipelines can be effectively combined to create a practical and scalable foundation for data-driven agriculture. By distributing responsibilities between edge-level sensing and backend prediction, the ESP32 functions as a lightweight data acquisition unit while Apache Airflow manages continuous and dependable model orchestration.

The application of Random Forest algorithms yielded a strong correlation between forecasted and observed values, confirming that accurate environmental prediction is achievable without the need for costly or proprietary hardware. The findings highlight that prediction-aware irrigation and environmental monitoring can substantially minimize resource wastage and provide farmers with clearer, more informed decision-making capabilities.

Ultimately, the success of this system reinforces the idea that open-source toolchains and low-cost IoT architectures can serve as a sustainable foundation for the digital transformation of agriculture. By enabling accessible, predictive, and automated control, the framework strengthens the potential for machine-learning-driven precision farming in developing regions—making smart agriculture both practical and attainable for smallholder farmers worldwide.

REFERENCES

[1] M. R. Islam, T. K. Roy and S. Hossain, "IoT Based Smart Agriculture Monitoring System," International Journal of Advanced Research in Computer Science, vol. 8, no. 5, 2017.
[2] A. A. Bhosale and P. D. Pawar, "Smart Farming Using IoT," International Journal of Electrical and Computer Engineering, vol. 9, no. 3, 2019.
[3] P. Gupta and R. K. Singh, "Soil Moisture Monitoring using ESP32 and Cloud Platforms," International Journal of Engineering Research & Technology, vol. 9, no. 11, 2020.
[4] S. J. Patel, K. Shah and H. Naik, "IoT Based Agricultural Decision Support System," International Journal of Innovative Technology and Exploring Engineering, vol. 8, no. 9, 2019.

[5] V. Sharma, R. Kumar and S. Goyal, “Machine Learning Techniques for Prediction of Environmental Parameters in Agriculture,” International Journal of Computer Applications, vol. 182, no. 32, 2018.

[6] J. L. José, A. N. Regis and V. H. Rodrigues, “ESP32 Based Remote Sensor Node for Smart Farming,” International Journal of Research in Engineering and Technology, vol. 7, no. 4, 2021.

[7] K. R. Khandelwal and P. K. Jain, “Data Driven Irrigation Control using Random Forest Models,” International Journal of Scientific Research in Computer Science, Engineering and Information Technology, vol. 6, no. 5, 2020.

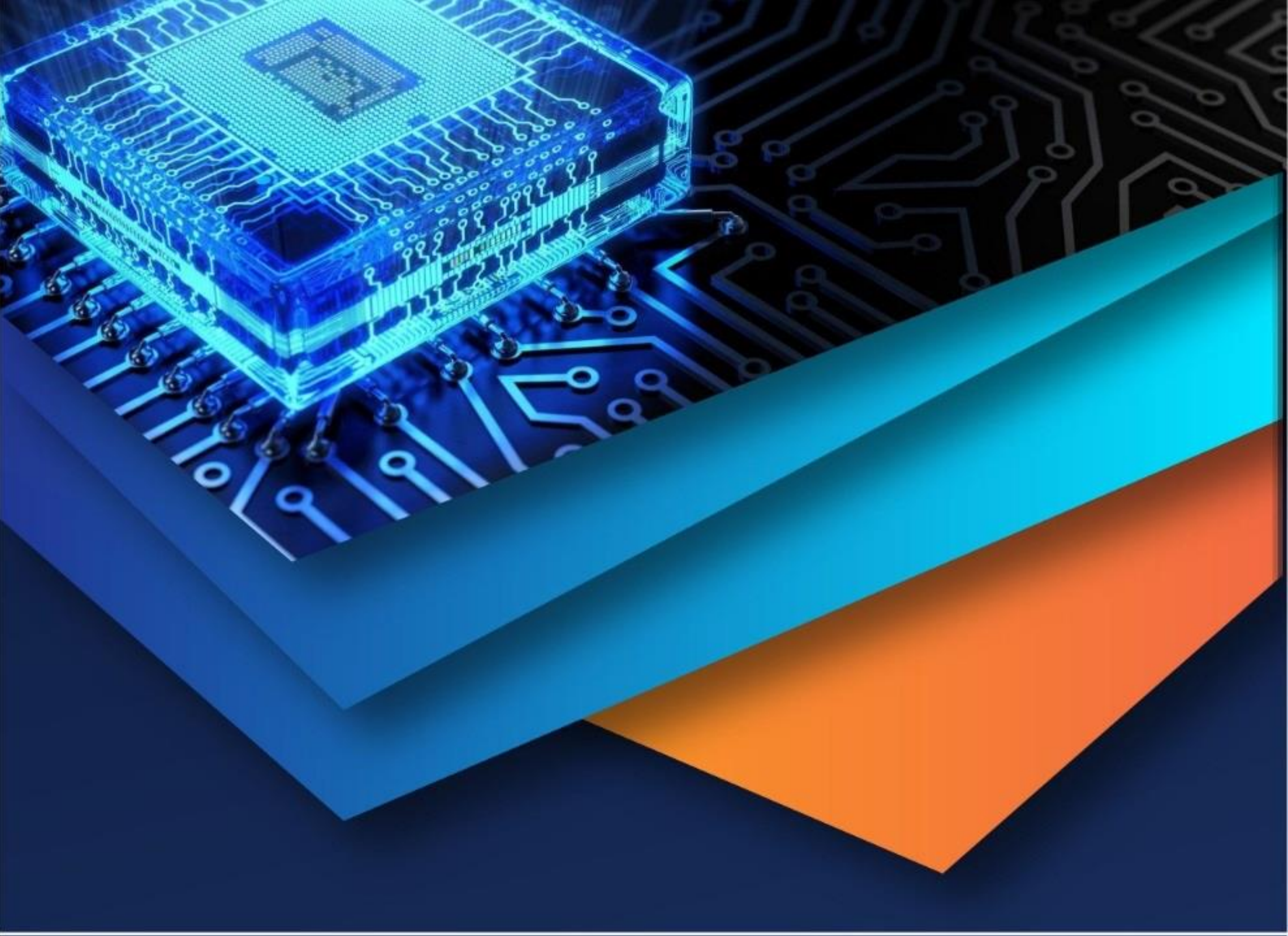
[8] A. Dash, P. Kabi and S. Sahu, “Workflow Automation using Apache Airflow for Data Engineering Pipelines,” International Journal of Engineering and Advanced Technology, vol. 8, no. 6, 2019.

[9] Apache Airflow Documentation, “Core Concepts and DAG Orchestration,” Website Link: https://airflow.apache.org/docs/ .

[10] Metabase Documentation, “Metabase Dashboards and Data Analytics,” Website Link: https://www.metabase.com/docs/latest/ .

[11] B. Karmel, “Airflow in Action,” Manning Publications, 2022. (Book on practical Airflow workflow design)

[12] A. Lahiri, “Practical Data Engineering with Apache Airflow,” Packt Publishing, 2023.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY