2nd INNOVEX 2026 - International Conference
Website:
ISBN: **-**

AI-DRIVEN AGRICULTURE FRAMEWORK USING ENSEMBLE ML WEBGIS AND RAG SYSTEM

Macalisang, DIG.[1]*, Magdadaro, JJM.[2], Say, RAL.[3], Ilustrisimo, AJB.[4]

[123]Bachelor of Science in Information Technology, Southwestern University PHINMA (Philippines)
[4]College in Information Technology, Southwestern University PHINMA (Philippines)

*Corresponding email: daga.macalisang.swu@phinmaed.com

ABSTRACT

The agricultural advisory process in the Philippines is constrained by manual soil analysis, static regulatory mapping, and systemic administrative latency. This study evaluates the operational bottlenecks experienced by the Department of Agriculture, Region 7, and proposes an Internet of Things (IoT) and Machine Learning (ML) framework to advance precision agriculture and modernize extension services. Utilizing a mixed-methods approach across 41 stakeholders in Dalaguete, Cebu, the baseline assessment revealed critical limitations, including a lack of in-situ diagnostic hardware and reliance on error-prone visual slope estimations. These manual workflows contribute to a 30-day turnaround time for official advisories. To resolve these inefficiencies, this study introduces an integrated architecture consisting of three conceptual domains: IoT perception and tiered data acquisition, a cloud-based web system engine, and a multi-tiered Artificial Intelligence (AI) synthesis pipeline. The framework leverages ensemble machine learning for crop suitability forecasting, Web Geographic Information System (WebGIS)-based regulatory compliance enforcement, and Retrieval-Augmented Generation (RAG) for localized advisory synthesis. By centralizing disparate data streams into a unified processing pipeline, the architecture effectively bypasses administrative bottlenecks, providing a scalable, data-driven, and legally compliant solution for agricultural extension efforts.

Keywords: *Precision Agriculture, Internet of Things (IoT), Ensemble Machine Learning, Regulatory Compliance, Retrieval-Augmented Generation (RAG)*

1. INTRODUCTION

In an era defined by rapid digital transformation, the agricultural sector is transitioning toward precision agriculture and Agriculture 4.0 to optimize resource allocation and enhance yield predictability [1]. However, in regions such as Dalaguete, Cebu, this transition is heavily constrained. The critical frontier lies in accurately assessing soil capability while strictly adhering to legal land-use regulations, factors that significantly impact local farming economies, sustainable land management, and long-term market viability.

Traditionally, agricultural health and soil assessments have relied on in-situ manual methods, such as rapid Soil Test Kits (STKs) and subjective visual estimations for land topography. These methods are labor-intensive, highly subjective, and time-consuming. With advancements in technology, the Internet of Things (IoT), machine learning (ML), and geospatial mapping are emerging as powerful tools to enhance the speed, precision, and efficiency of agricultural advisory processes [1], [2], [3]. The primary problem addressed in this research is the inefficiency and lack of scalability of conventional agricultural advisory methods, especially in detecting precise soil nutrient deficits and enforcing regulatory compliance promptly and accurately. Constraints in time, hardware access, and fragmented administrative silos make traditional methods highly susceptible to a 20- to 30-day reporting delay, rendering them less suitable for responsive agricultural planning.

The existing agricultural advisory process is fundamentally hampered by inaccuracies in data acquisition and nutrient matching. Sabado [4] identified a systemic mismatch where manual extension services fail to address site-specific farm requirements. The complexity of traditional soil testing frequently discourages farmers from adopting scientific methods, forcing a reliance on intuition [5]. Beyond agronomic limitations, the manual workflow is inefficient for regulatory

verification. Conflicting static map data makes manual validation of land classification prone to errors, failing to accurately enforce spatial restrictions like the 18% slope limit of Presidential Decree No. 705 [6]. Furthermore, traditional extension services are paralyzed by operational latency. Outdated processing methods result in up to 30-day turnaround times, rendering manual recommendations obsolete by the time they reach the stakeholder [7].

To overcome the data latency inherent in manual advisory, contemporary precision agriculture increasingly relies on real-time Internet of Things (IoT) architectures. Edge computing nodes, such as the ESP32, are highly effective for integrating multi-parameter sensors for pH and NPK values [1]. This transition to real-time, in-situ sensing is the only viable method to capture the temporal variability of soil nutrients that static manual testing frequently misses [2]. By shifting toward a hardware-agnostic model, systems can utilize various ingestion streams, including RS485 Modbus digital probes, to mitigate signal noise and ensure reliable offline buffering in environments with poor network connectivity [8].

Transforming raw environmental telemetry into actionable foresight requires sophisticated, non-linear modeling. While traditional systems rely on static charts, advanced ensemble Machine Learning (ML) techniques, specifically the combination of XGBoost and Ridge Regression (KRR) stacking, drastically reduce prediction variance and bias when forecasting crop suitability and market viability [3], [9]. To bridge the gap between numerical ML outputs and user accessibility, recent frameworks integrate Large Language Models (LLMs) protected by Retrieval-Augmented Generation (RAG). Constraining the LLM strictly to a vetted corpus of agricultural science ensures the system generates human-readable advisories without dangerous agronomic hallucinations [10]. Finally, to enforce regulatory compliance, modern geospatial architectures rely on cloud-based WebGIS Digital Elevation Models (DEMs). These tools precisely compute land topography gradients, seamlessly enforcing slope restrictions without the errors associated with physical field manipulation or static maps [11], [12].

As a solution, this study proposes the development of an integrated IoT and Machine Learning framework capable of analyzing exact crop suitability through multidimensional data processing [3]. This model will be integrated into a hardware-agnostic, cloud-based Software-as-a-Service (SaaS) platform accessible to various stakeholders, including Agricultural Extension Workers (AEWs), laboratory personnel, and municipal administrators. The main objective of this research is to design and develop an architectural framework that provides automated, rapid, and precise advisory assessments. By leveraging this approach, the study specifically aims to: (a) evaluate the current methodology used by the Department of Agriculture in assessing soil health to identify its operational limitations; (b) systematically compare available technological models to formulate a new conceptual architecture for efficiently assessing crop suitability and regulatory compliance; and (c) design an IoT and Machine Learning framework capable of bypassing traditional administrative latency. The obtained result will serve as the structural foundation to develop an IT solution that supports more responsive and data-driven agricultural extension efforts.

2. METHOD

This study utilized a descriptive mixed-methods research design within the Dalaguete, Cebu pilot zone to evaluate existing agricultural advisory frameworks. The site was purposefully selected due to its high-density agricultural activity and the significant topographical constraints that necessitate both precise nutrient matching and strict adherence to the slope restrictions mandated by Presidential Decree No. 705. The research employed non-probability purposive sampling for farmers and Agricultural Extension Workers (AEWs) to capture direct operational field experience, alongside total enumeration for Department of Agriculture (DA) Regional Soils Laboratory Personnel to integrate critical technical expertise. This resulted in a final dataset of 41 validated respondents 21 farmers, 16 AEWs, and 4 DA laboratory personnel providing a representative cross-section of both service providers and end-users.

Data collection utilized a Technical Baseline Audit via structured Google Forms surveys, configured with a 5-point Likert scale

(1 = "Strongly Disagree" to 5 = "Strongly Agree"). To ensure instrument reliability and technical accuracy, all survey instruments and interview guides underwent rigorous content validation by a panel of information technology and agricultural science experts. Feedback from this panel was integrated to refine phrasing and eliminate ambiguities prior to deployment. Quantitative data was processed using descriptive statistics, including weighted mean and standard deviation, to identify the central tendency and variability of operational bottlenecks. Following the baseline assessment, a system analysis and literature synthesis utilized a Comparative Decision Matrix to conceptually construct the proposed hardware-agnostic, cloud-based SaaS architecture, ensuring the design remained directly aligned with the identified practical constraints of the current agricultural extension system.

3. RESULT AND DISCUSSION

The baseline audit established that the current field-level data acquisition framework is highly constrained, yielding a sub-mean score of 4.00. Agricultural Extension Workers (AEWs) identified the absence of real-time, in-situ hardware as the primary operational limitation (Mean = 4.06). Compelled to rely on rapid Soil Test Kits (STKs) instead, personnel reported inconsistent nutrient readings driven by subjective manual color interpretation (Mean = 3.94), a vulnerability further corroborated by recent technical evaluations [5]. Geospacially, AEWs confirmed that verifying farm slopes against the Presidential Decree (PD) 705 18% restriction using static 2D maps is disproportionately time-consuming and error-prone (Mean = 3.81). Such manual mapping methods frequently fail to capture dynamic land cover changes, increasing the risk of inaccurate assessments in critical upland zones [6]. At the administrative backend, fragmented data silos and manual logging (Mean = 4.00) push official laboratory turnaround times to an unsustainable 20–30 working days (Mean = 3.75).

Table 1. Operational Bottlenecks in the Manual Advisory Workflow.

Operational Indicator	Mean	Interpretation
Lack of real-time, in-situ hardware sensors (AEWs)	4.06	Agree
Inconsistent nutrient readings from manual STKs (AEWs)	3.94	Agree
Manual verification of PD 705 slope restrictions (AEWs)	3.81	Agree
Fragmented data silos & manual cross-referencing (DA Lab)	4.00	Agree
Official advisory turnaround time (20–30 working days)	3.75	Agree

A complementary qualitative thematic analysis of interviews with Department of Agriculture personnel and Agricultural Extension Workers (AEWs) revealed three cascading failures underpinning the quantitative data. First, subjectivity-driven hardware constraints impede precision. The absence of digital sensors forces reliance on test kits yielding only generalized categorical outputs rather than precise numerical concentrations, lacking the granularity necessary for predictive accuracy. Furthermore, physical data acquisition is labor-intensive, requiring significant exertion across rugged terrain. This is compounded by static geospatial vulnerabilities, where persistent cloud cover in upland areas obstructs optical satellite imagery [2], preventing reliable topographical detection through traditional 2D mapping.

Second, the data exposed the impracticality of Presidential Decree (PD) 705 regulatory enforcement. Lacking dynamic geospatial tools, AEWs rely on unscientific visual estimations. Interviews revealed that personnel often bypass regulations to avoid socio-economic conflict, as farmers prioritize planting cycles over land-use restrictions. This confirms that the regulatory compliance gap is a social enforcement challenge, not merely a technical one, that requires an objective, automated resolution.

Third, systemic administrative latency creates a pattern of "dead-end data," where identified soil chemical issues fail to translate into actionable recommendations. Current manual color-matching provides only basic acidity readings, lacking specific, data-driven application guidelines for farmers. Combined with fragmented logging workflows that rely on isolated spreadsheets requiring manual

transcription, this bottleneck drives a 30-day advisory delay, forcing farmers to miss critical planting windows and rely on generational intuition.

To resolve the identified operational failures, a Comparative Decision Matrix (Table 2) was employed to systematically evaluate technological alternatives for each architectural layer. The selected frameworks were chosen to ensure operational stability, scalability, and scientific accuracy. For data acquisition, the ESP32 node with an RS485 Modbus digital probe was selected over analog alternatives to prevent signal noise and ensure superior memory buffering in low-connectivity rural environments [1]. In the predictive intelligence layer, Ensemble Machine Learning, specifically the combination of XGBoost and Ridge Regression stacking, was prioritized over Deep Neural Networks to mathematically reduce prediction variance without the massive dataset overhead required by deep learning architectures [3]. To guarantee safety in advisory generation, a Retrieval-Augmented Generation architecture was implemented over standard Large Language Models. This prevents agronomic hallucinations by constraining advisory generation strictly to verified Department of Agriculture literature [10]. Finally, cloud-based WebGIS Digital Elevation Models were utilized over onboard hardware sensors to eliminate micro-topography biases and compute precise slope gradients for automated Presidential Decree 705 enforcement [11].

Table 2. Comparative Decision Matrix for System Architecture Selection.

Architectural Parameter	Evaluated Alternatives	Selected Framework
Data Acquisition	Analog Probes (I2C); Arduino Uno	**ESP32 Edge Node with Digital RS485**
Predictive Engine	Deep Neural Networks; Decision Trees	**Ensemble ML (XGBoost + KRR)**
Advisory Synthesis	Rule-Based Generation; Standard LLMs	**Retrieval-Augmented Generation (RAG)**
Geospatial Analysis	Edge Hardware (Accelerometers)	**WebGIS API via Digital Elevation Models**

The selected frameworks were systematically chosen to ensure operational stability, scalability, and scientific accuracy. For data acquisition, the ESP32 node with RS485 Modbus digital probe was selected over analog I2C alternatives to prevent signal noise and ensure superior SRAM for offline buffering in low-connectivity rural environments [1]. In the predictive intelligence layer, Ensemble ML (XGBoost + KRR Stacking) was prioritized over Deep Neural Networks to mathematically reduce prediction variance without the massive dataset overhead required by deep learning architectures [5]. To guarantee safety in advisory generation, a Retrieval-Augmented Generation (RAG) architecture was implemented over standard LLMs; this completely prevents agronomic hallucinations by constraining advisory generation strictly to verified DA literature [4]. Finally, cloud-based WebGIS Digital Elevation Models (DEMs) were utilized over onboard hardware sensors to eliminate micro-topography biases and compute precise slope gradients for automated PD 705 enforcement [6].

The validated architectural solutions from the Comparative Decision Matrix were synthesized into three conceptual frameworks that map the data flow from the field to the final advisory output. Engineered as a direct response to identified operational failures, the Hybrid AI architecture executes through three distinct domains deployed on scalable cloud infrastructure.

The first domain, IoT perception and data acquisition, resolves the limitations of the manual baseline. The process initiates when an Agricultural Extension Worker designates a site via a mobile-responsive interface. To guarantee spatial tracking without exhausting edge hardware power, the system employs a dual-redundant positioning architecture, utilizing mobile geolocation as the primary source with a physical hardware module serving as an autonomous fail-safe. The physical IoT node captures raw physicochemical telemetry and transmits it to an embedded microcontroller, which buffers data locally during network dropouts before syncing via secure protocols to the cloud.

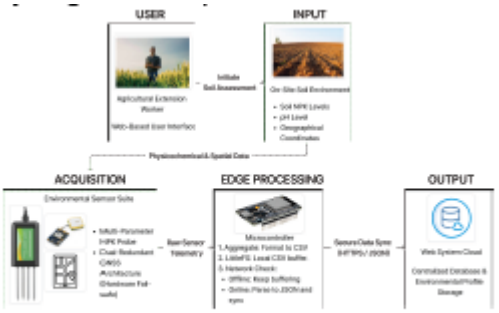


Figure 1. IoT Perception & Data Acquisition Architecture.

The second domain comprises the web architecture and cloud processing engine. The system aggregates environmental profiles through automated IoT telemetry, manual web-based encoding, or bulk laboratory uploads. These payloads converge into a centralized database for multi-tiered processing. The engine first performs spatial validation against elevation models to programmatically enforce regulatory slope restrictions. The pipeline then advances to the ensemble learning engine for crop suitability forecasting, concluding with an intelligent integration layer that synthesizes numerical outputs into localized, human-readable advisory reports.

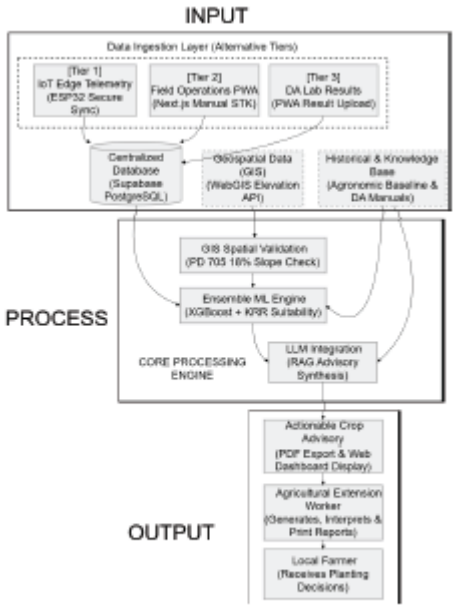


Figure 2. IPO Web System Architecture.

Finally, the multi-tiered AI synthesis layer acts as the decision architecture. The spatial validation layer evaluates boundary coordinates against elevation models as a legal gatekeeper. The predictive intelligence layer funnels input payloads through a normalization engine, which standardizes data into a feature vector processed by the ensemble learning model. The framework converges at the AI synthesis layer, where the system dynamically consolidates numerical machine learning metrics and retrieves semantic context from a secure knowledge base. By embedding policy constraints directly into the synthesis layer, the framework mathematically forces the generative model to pivot recommendations toward sustainable agroforestry if a slope violation is detected.

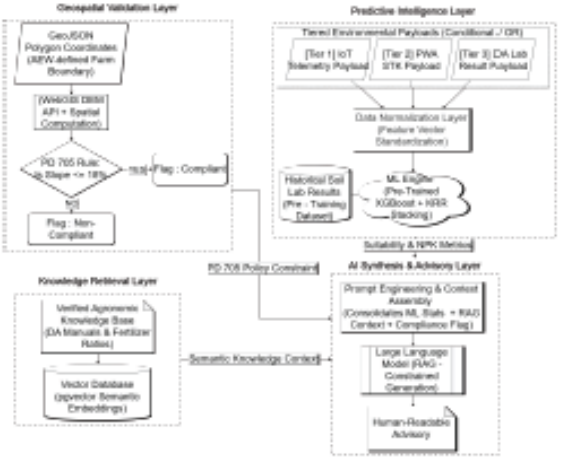


Figure 3. Multi-Tiered AI Synthesis Architecture.

4. CONCLUSION

The baseline evaluation definitively confirms that the traditional agricultural advisory process is paralyzed by subjective heuristic tools, static mapping vulnerabilities, and severe administrative latency. To bridge the gap between empirical data and on-the-ground action, the implementation of an integrated IoT and Machine Learning framework is structurally justified. By combining RS485 IoT edge telemetry, ensemble machine learning for precise crop suitability, automated geospatial regulatory enforcement, and RAG-constrained LLM reporting, the proposed architecture successfully conceptualizes a scalable solution to bypass the 30-day turnaround delay. Ultimately, this system transitions extension services toward true precision agriculture, providing a fully engineered architecture that is structurally ready for comparative experimental acceptability testing.

Acknowledgements

The researchers extend their deepest gratitude to the Department of Agriculture - Region 7, the Municipal Agriculture Office of Dalaguete, and the College of Information Technology and Engineering at Southwestern University PHINMA for their operational and academic support.

References

[1] A. Abdullah, E. I. Raharjo, M. Iwan, R. Faizal, and Maryogi, "Development of an IoT-based soil nutrient monitoring and GIS mapping system for precision agriculture," Advance Sustainable Science, Engineering and Technology, vol. 7, no. 4, 2025.

[2] L. Burton, K. Jayachandran, and S. Bhansali, "The 'real-time' revolution for in situ soil nutrient sensing," Journal of The Electrochemical Society, vol. 167, no. 3, 2020.

[3] K. N. Vhatkar, S. A. Koparde, S. Kothari, J. Sarwade, and K. Sakur, "Enhancing prediction of crop yield and soil health assessment for sustainable agriculture using machine learning approach," MethodsX, vol. 14, p. 103418, 2025.

[4] W. Sabado, "The current status of precision farming in the Philippines: A Delphi method approach," Universitas: The Official Research Journal of University of Makati, vol. 12, no. 2, pp. 85-97, 2024.

[5] A. L. Samaniego and M. Gallego Jr., "Farm Soil Testing Techniques, Soil Health, and Potential Impact of Artificial Intelligence as Factors of Small-Scale Farming Soil Management," E3S Web of Conferences, vol. 595, p. 03001, 2024.

[6] M. J. C. Gabriel, "Dynamics and drivers of deforestation in the Philippines," Ecosystems and Development Journal, vol. 13, no. 1, pp. 18-32, 2023.

[7] L. T. H. Sen, L. T. H. Phuong, P. Chou, F. B. Dacuyan, Y. Nyberg, and J. Wetterlind, "The Opportunities and Barriers in Developing Interactive Digital Extension Services for Smallholder Farmers as a Pathway to Sustainable Agriculture: A Systematic Review," Sustainability, vol. 17, no. 7, p. 3007, 2025.

[8] P. Musa, H. Sugeru, and E. P. Wibowo, "Wireless sensor networks for precision agriculture: A review of NPK sensor implementations," Sensors, vol. 24, no. 1, p. 51, 2024.

[9]O. Folorunso et al., "Exploring machine learning models for soil nutrient properties prediction: A systematic review," Big Data and Cognitive Computing, vol. 7, no. 2, p. 113, 2023.

[10] D. J. Samuel Ravindran, I. Skarga-Bandurova, V. Sivakumar, M. Awais, and S. Mithra, "AgroLLM: Connecting Farmers and Agricultural Practices Through Large Language Models," AgriEngineering, vol. 8, no. 1, p. 38, 2026.

[11] L. Polidori and M. El Hage, "Digital elevation model quality assessment methods: A critical review," Remote Sensing, vol. 12, no. 21, p. 3522, 2020.

[12] N. R. Almarines, S. Hashimoto, J. M. Pulhin, C. L. Tiburan Jr, A. T. Magpantay, and O. Saito, "Influence of image compositing and multisource data fusion on multitemporal land cover mapping of two Philippine watersheds," Remote Sensing, vol. 16, no. 12, p. 2167, 2024.