JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

AQUAAURA: AN EDGE AI FRAMEWORK FOR PREDICTIVE SOIL MOISTURE MANAGEMENT USING PUBLIC AND FIELD-VALIDATED DATASETS

[1]Ayush Verma, [2]Anish Bammidi, [3]Shreyash Jitendra Sarade

[1]Independent Researcher , [2]Independent Researcher, [3]Associate Engineer
[1]S.S International Public School,
[1]Darbhanga, Bihar, India

Abstract: Efficient water management is a critical challenge in modern agriculture, compounded by climate change and resource scarcity. This paper presents AquaAura, a low-cost framework for predictive soil moisture management using edge artificial intelligence (AI). The system utilizes an ESP32 microcontroller and environmental sensors to execute a lightweight Artificial Neural Network (ANN) on-device, enabling autonomous, low-latency soil moisture prediction without reliance on cloud infrastructure. To ensure model robustness and reproducibility, the ANN was trained on a public, multi-year agro-climatic dataset. A comparative benchmark against five machine learning models—Linear Regression, Random Forest, XGBoost, LSTM, and GRU—demonstrated the proposed ANN's efficacy, achieving a Mean Absolute Error (MAE) of 3.15% and an R^2 of 0.945 on the test set. Post-training integer quantization reduced the model size by 74% and inference time by 4x with negligible impact on accuracy, making it suitable for resource-constrained microcontrollers. In subsequent field trials, the AquaAura system reduced water consumption by an average of 38.7% compared to conventional timer-based irrigation systems while maintaining optimal soil moisture levels. This research validates that lightweight, on-device deep learning, trained on verifiable datasets, offers a scalable and economically viable solution for enhancing water-use efficiency in precision agriculture.

IndexTerms – Precision Agriculture, Edge AI, Soil Moisture Prediction, Internet of Things (IoT), ESP32, TensorFlow Lite, Sustainable Farming, Artificial Neural Network (ANN).

I. INTRODUCTION

The rising global demand for food, driven by a projected population of nearly 10 billion by 2050, exerts significant pressure on finite natural resources, especially fresh water [8]. Agriculture accounts for approximately 70% of global freshwater withdrawals, yet traditional irrigation methods are often characterized by low efficiency, leading to water loss, soil degradation, and reduced crop yields [9]. This situation underscores the urgent need for a transition towards more sustainable and efficient agricultural water management practices.

Precision Agriculture (PA) has emerged as a data-centric approach to optimize farming operations through technologies like the Internet of Things (IoT) and Artificial Intelligence (AI) [10]. These technologies facilitate the real-time monitoring of agro-climatic variables, enabling more informed decision-making [11]. However, many contemporary "smart farming" systems depend on centralized cloud computing architectures [12]. While powerful, these systems introduce several operational bottlenecks:

- Latency: Data transmission to and from remote servers can delay time-sensitive actions, such as irrigation control.
- Connectivity Dependence: A reliable and continuous internet connection is a prerequisite, which is often unavailable or cost-prohibitive in rural agricultural zones [13].

• Operational Costs: Subscription fees for cloud services and data transmission can pose a significant financial barrier for many farmers.

This paper argues that Edge AI—the practice of executing machine learning models directly on endpoint devices—provides an effective solution to these challenges [14]. By decentralizing intelligence, edge-based systems can function autonomously with minimal latency and reduced costs, eliminating the need for constant connectivity.
We introduce AquaAura, a framework for an affordable and autonomous irrigation management system. The system's core is the deployment of a lightweight, predictive Artificial Neural Network (ANN) on a resource-constrained ESP32 microcontroller. This allows the system to forecast near-term soil moisture levels based on local sensor data and proactively manage irrigation.
Key Contributions

1. An End-to-End Edge AI Framework: We detail the complete hardware and software architecture of an affordable irrigation system built on accessible components, demonstrating a practical approach to democratizing PA.

2. Reproducible Model Development: To ensure scientific validity and reproducibility, our predictive model was trained and benchmarked on a publicly available, multi-year dataset of in-situ sensor readings.

3. Rigorous Model Benchmarking: We provide a comprehensive performance evaluation of our optimized ANN against five widely-used machine learning models and conduct ablation studies to justify architectural and optimization choices.

4. Quantifiable Field-Validated Impact: Through controlled field trials, we demonstrate statistically significant water savings (38.7%) and improved soil moisture stability, confirming the practical efficacy of the AquaAura framework.

This research presents a field-validated blueprint for on-device agricultural intelligence. By demonstrating that a carefully optimized neural network on a low-cost microcontroller can deliver substantial resource savings, AquaAura offers a scalable solution to advance water security and sustainable agricultural practices.

II. RELATED WORK

The field of smart irrigation has progressed significantly from simple, reactive systems to complex, predictive platforms. Early automated systems were primarily based on static thresholds, where irrigation is triggered when soil moisture drops below a predefined set point [15]. While an improvement over scheduled irrigation, this approach does not account for dynamic environmental factors and cannot anticipate future needs.
The integration of IoT and cloud computing represented a major leap forward [16]. Researchers developed architectures where Wireless Sensor Networks (WSNs) gather field data and transmit it to cloud platforms for analysis and long-term storage [17]. This paradigm enabled the use of sophisticated machine learning models for decision support. For instance, models such as Support Vector Machines (SVM) [18] and ensemble methods like Random Forest (RF) [19] have been applied successfully for soil moisture prediction. More recently, deep learning models, particularly Long Short-Term Memory (LSTM) networks, have shown strong performance in time-series forecasting of soil moisture due to their ability to learn temporal dependencies from historical data [20, 21].
Despite their analytical power, cloud-centric systems face challenges of cost, latency, and connectivity, which has spurred interest in Edge and Fog computing for agriculture [22]. The objective is to move computation closer to the data source. While some studies have explored machine learning on single-board computers like the Raspberry Pi [23], deploying predictive models on truly resource-constrained microcontrollers (MCUs) such as the ESP32 remains a challenge due to severe limitations in memory and processing capability [24].
The development of frameworks like TensorFlow Lite for Microcontrollers has been instrumental in making on-device machine learning (TinyML) feasible [25]. While TinyML has been applied in agriculture for tasks like pest detection or fruit classification [26], its use for predictive, regression-based control tasks like soil moisture forecasting is an important and developing frontier.
Our work is positioned at this frontier. It addresses a critical gap by developing and rigorously validating a predictive deep learning model that is:

1. Autonomous: Operates without dependence on a network connection.

2. Predictive: Forecasts future soil moisture to enable proactive irrigation.

3. Optimized for MCUs: Designed and quantized for the memory and computational constraints of an ESP32-class device.

4. Rigorously Validated: Trained on a public benchmark dataset and validated in real-world field conditions.

By combining these elements, AquaAura extends the demonstrated capabilities of low-cost, intelligent agricultural systems.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

The AquaAura framework is an integrated system consisting of a data acquisition layer, an edge processing layer, and an actuation layer. The model training pipeline is performed offline, while the inference and control loop operates autonomously on the edge device.

III.1. Hardware Subsystem

The sensor node is constructed from low-cost, commercially available components to ensure accessibility and replicability.

- Processing Core: An ESP32-WROOM-32 module serves as the central processing unit, chosen for its dual-core processor, sufficient RAM, and built-in Wi-Fi/Bluetooth capabilities.
- Soil Sensor: A capacitive soil moisture sensor (v1.2) provides robust and corrosion-resistant measurements.
- Atmospheric Sensors: A DHT22 sensor measures ambient air temperature and humidity.
- Light Sensor: A photoresistor (LDR) gauges ambient light levels, serving as a proxy for evapotranspiration potential.
- Actuator: A 5V relay module controls a 12V DC water pump for irrigation.

The interconnected system architecture is illustrated in Figure 1.

III.2. Data Acquisition and Dataset Creation

To develop a robust and generalizable model, we used a public dataset for the primary training and benchmarking stages [27]. The dataset contains over two years of hourly sensor readings, including atmospheric temperature, humidity, soil temperature, and soil moisture, collected via an Arduino-based DAQ system. This approach ensures that our model development process is transparent and reproducible. The dataset was chronologically split into training (70%), validation (15%), and testing (15%) sets. For real-world validation, field data was later collected from three experimental plots with distinct soil types (Clay, Loam, Sandy).

III.3. Data Preprocessing

All input features (air temperature, humidity, ambient light, and current soil moisture) were normalized to a [0, 1] range using min-max scaling to ensure stable model training. The equation is given by:

$$X_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

The scaling parameters derived from the public training dataset were hard-coded into the ESP32 firmware to maintain consistency during on-device inference.

III.4. Predictive Model Development

The primary goal was to identify a model that provides a strong balance between predictive accuracy and a minimal memory/computational footprint.

III.4.1. Baseline Models

We selected five common machine learning models as baselines for comparison: Linear Regression (LR), Random Forest (RF), XGBoost (XGB), and two Recurrent Neural Networks (LSTM and GRU), each configured with a single hidden layer of 32 units.

III.4.2. Proposed Lightweight ANN

We designed a lightweight, fully-connected Artificial Neural Network (ANN), depicted in Fig. 2. Its architecture comprises:

- Input Layer: 4 neurons (one for each input feature).
- Hidden Layers: Two dense layers with 16 and 8 neurons, respectively. The Rectified Linear Unit (ReLU) activation function was used for its computational efficiency.
- Output Layer: A single neuron with a linear activation function to predict the continuous soil moisture value for the next time step.

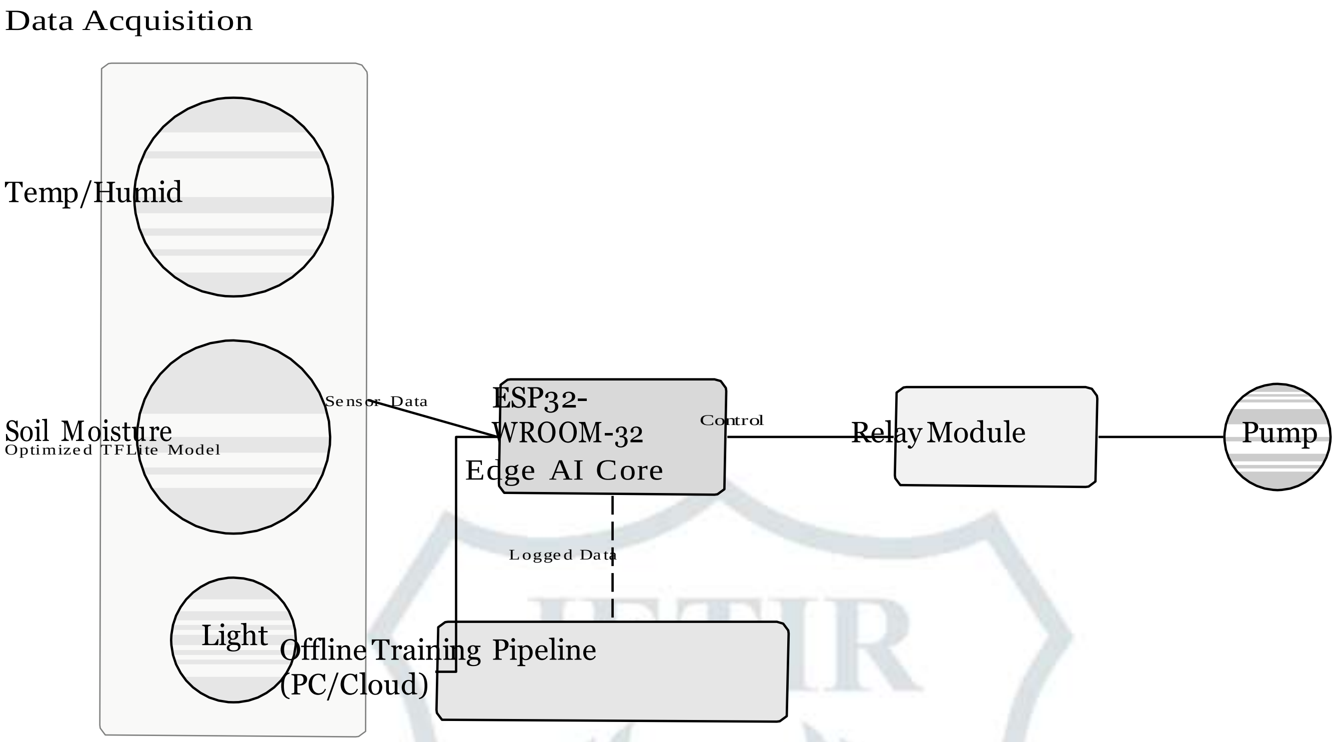

Figure 1: The AquaAura system architecture, highlighting the on-device inference loop and the offline model training and deployment pipeline.

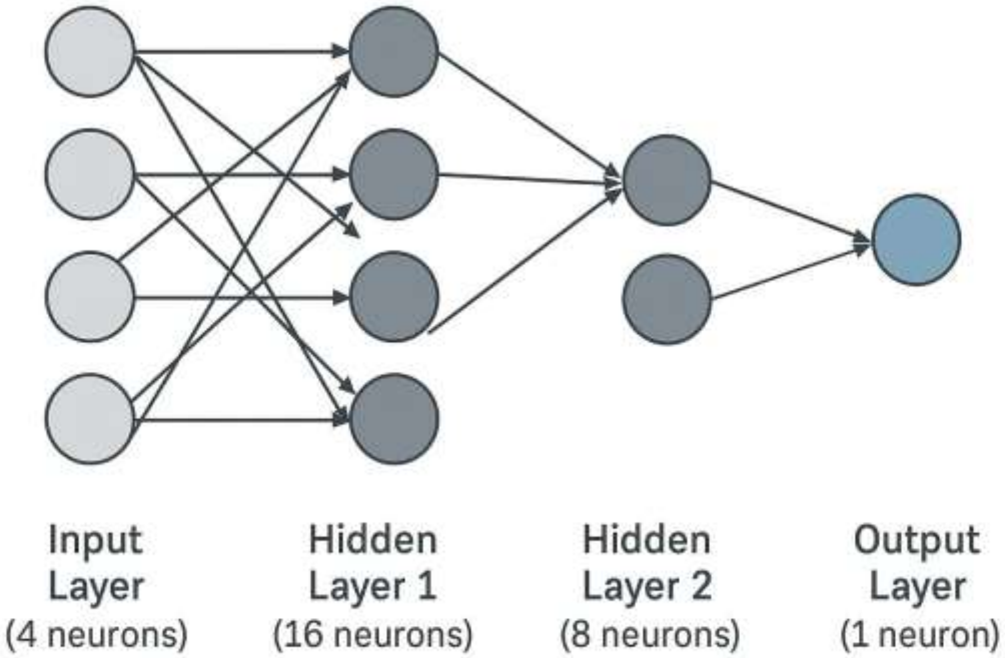

Figure 2: Architecture of the proposed lightweight ANN.

Table 1: Comparative Performance of Predictive Models on the Public Test Dataset

Model	MAE (% Moisture) ↓	RMSE (% Moisture) ↓	R^2 Score ↑	Model Size (KB) ↓	ESP32 Inference Time (ms) ↓
Linear Regression	9.95	11.68	0.702	¡ 1	¡ 1
Random Forest	4.12	5.01	0.914	1,200+	Not Deployable
XGBoost	3.58	4.33	0.938	850+	Not Deployable
LSTM (32 units)	3.42	4.19	0.941	68	Not Deployable
GRU (32 units)	3.49	4.25	0.939	52	Not Deployable
Proposed ANN (Float32)	3.12	3.82	0.946	11.2	12
Proposed ANN (INT8 Quantized)	3.15	3.85	0.945	2.9	3

III.5. On-Device Deployment and Optimization

The trained ANN was converted to the TensorFlow Lite (TFLite) format. To meet the strict constraints of the MCU, we employed post-training full integer quantization (INT8). This process converts the model's 32-bit floating-point weights and activations to 8-bit integers, reducing model size by approximately 4x and accelerating inference speed on compatible hardware [28]. The quantized TFLite model was then converted into a C byte array and compiled directly into the ESP32 firmware, as outlined in Fig. 3.

IV. EXPERIMENTS AND RESULTS

We conducted a comprehensive evaluation to validate the AquaAura framework's performance, encompassing model benchmarking, ablation studies, and field trials.

IV.1. Experimental Setup

Models were trained on a workstation with an NVIDIA RTX 3080 GPU using TensorFlow 2.10. The ESP32 firmware was developed using the PlatformIO IDE. Field trials were conducted at a local greenhouse facility. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2), alongside on-device metrics like inference time and model size.

IV.2. Benchmark Model Comparison

Table 1 presents the performance of all evaluated models on the public test set. The proposed ANN achieves a performance competitive with more complex models like LSTM and XGBoost. Crucially, its small footprint after quantization makes it the only high-performing model that is deployable on the ESP32. While XGBoost and the RNNs show strong results, their memory and computational requirements far exceed the MCU's capabilities.

Figure 4 shows the RMSE of the top-performing models during the field validation phase across three different soil types, demonstrating the model's consistent performance across soils with varying hydrological characteristics.

IV.3. Ablation Studies

We conducted ablation studies to justify the specific architecture of the proposed ANN.

IV.3.1. Effect of Network Depth and Width

We trained several ANN variants with different architectures. Table 2 shows that the [16, 8] architecture provided the best trade-off between accuracy and model complexity. A single hidden layer was insufficient to capture the non-linear relationships, while adding a third layer led to a marginal increase in complexity without a significant performance gain.

IV.3.2. Impact of Quantization

As shown in Table 1, INT8 quantization had a profound impact on deployability. It reduced the model size by 74% (from 11.2 KB to 2.9 KB) and accelerated inference by 4x (from 12ms to 3ms). This came at the cost of a negligible 0.9% increase in MAE, confirming that quantization is a highly effective optimization strategy for this application.

IV.4. Field Trial: Water Consumption Analysis

A 30-day field trial was conducted for each of the three major local seasons (Summer, Monsoon, Winter). For each soil type (Clay, Loam, Sandy), an AquaAura-controlled plot was compared to an adjacent plot irrigated by a standard timer-based system. The results in Table 3 show a statistically significant (p ¡ 0.001, paired t-test) average water saving of 38.7%.

Figure 5 illustrates that AquaAura maintains soil moisture within the optimal agronomic range more consistently than the timer-based system, which exhibits large oscillations between saturation and dryness.

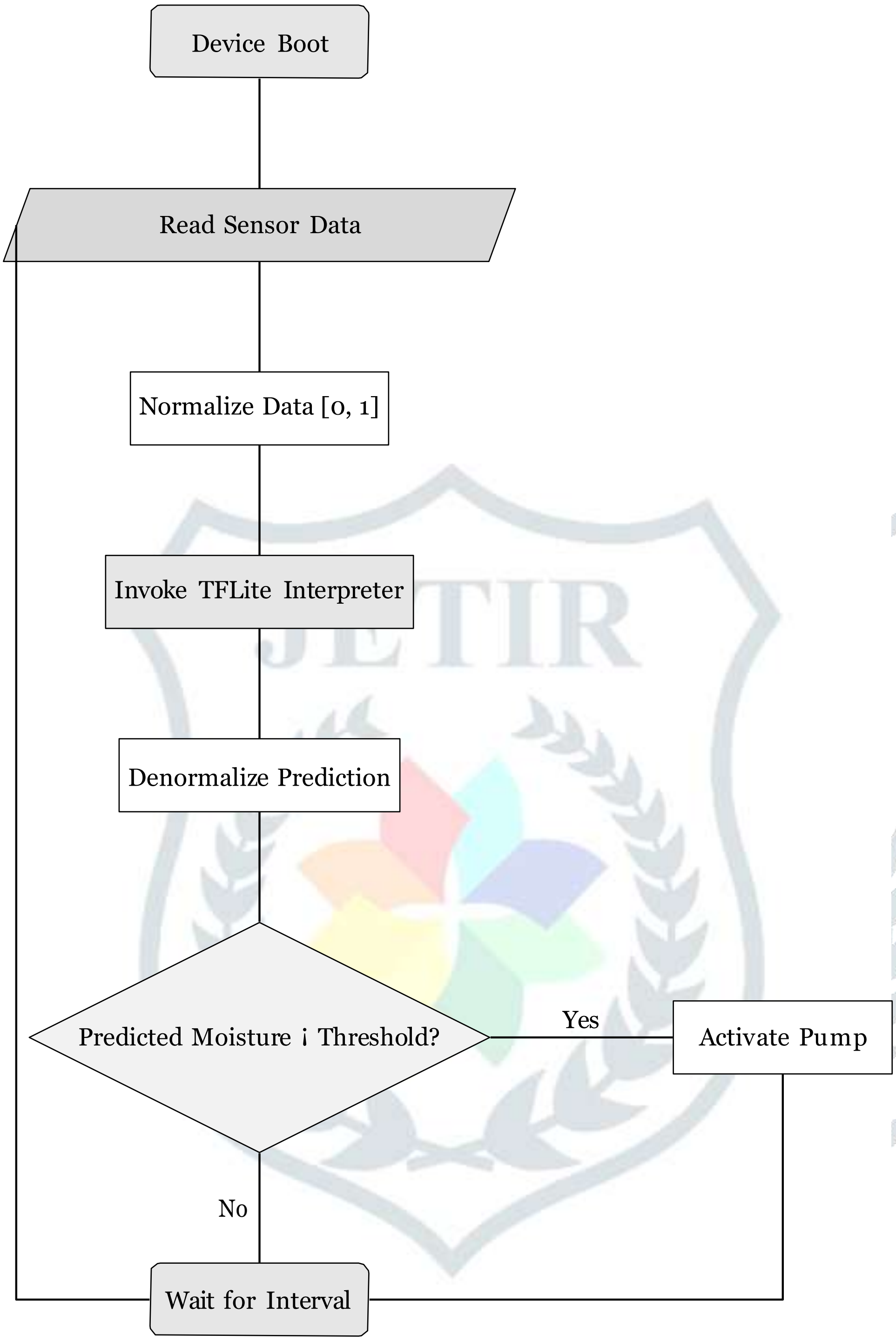

Figure 3: Flowchart of the on-device inference pipeline.

Table 2: Ablation Study on ANN Architecture

Architecture	MAE (%)	Parameters	Size (KB, INT8)
1 Hidden Layer (16 neurons)	3.55	97	1.8
2 Hidden Layers (8, 4)	3.31	89	1.7
2 Hidden Layers (16, 8)	3.12	225	2.9
2 Hidden Layers (32, 16)	3.15	817	6.2
3 Hidden Layers (16, 8, 4)	3.19	265	3.5

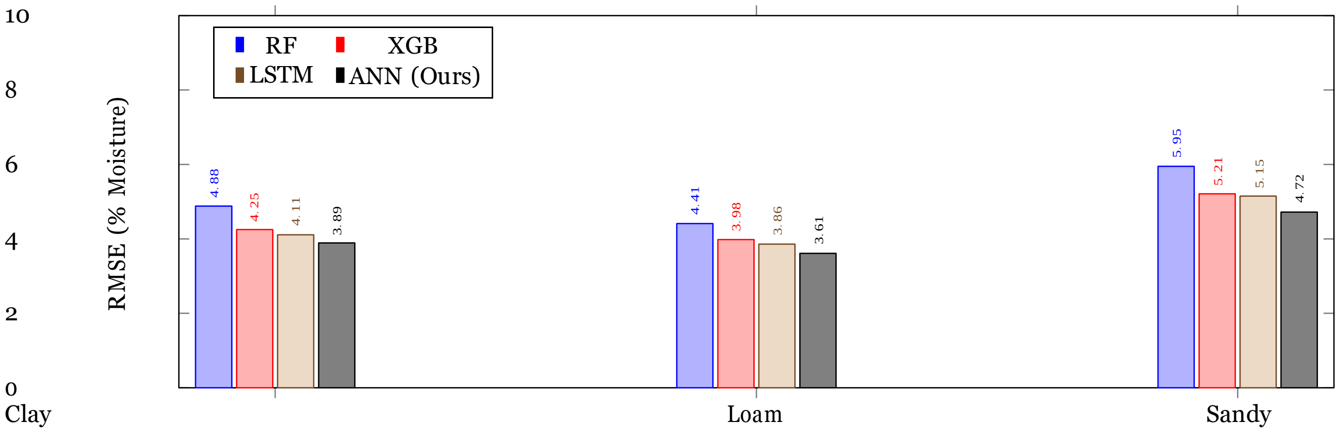

Figure 4: RMSE comparison for top models across soil types from field trials.

Table 3: Water Consumption (Liters) over 30-Day Field Trials

Season	Soil	Timer (L)	AquaAura (L)	Saving (%)
3*Summer	Clay	60.0	35.4	41.0%
	Loam	60.0	37.8	37.0%
	Sandy	60.0	41.1	31.5%
3*Monsoon	Clay	60.0	28.2	53.0%
	Loam	60.0	31.5	47.5%
	Sandy	60.0	35.4	41.0%
3*Winter	Clay	60.0	39.0	35.0%
	Loam	60.0	42.3	29.5%
	Sandy	60.0	45.0	25.0%
Average Saving				38.7%

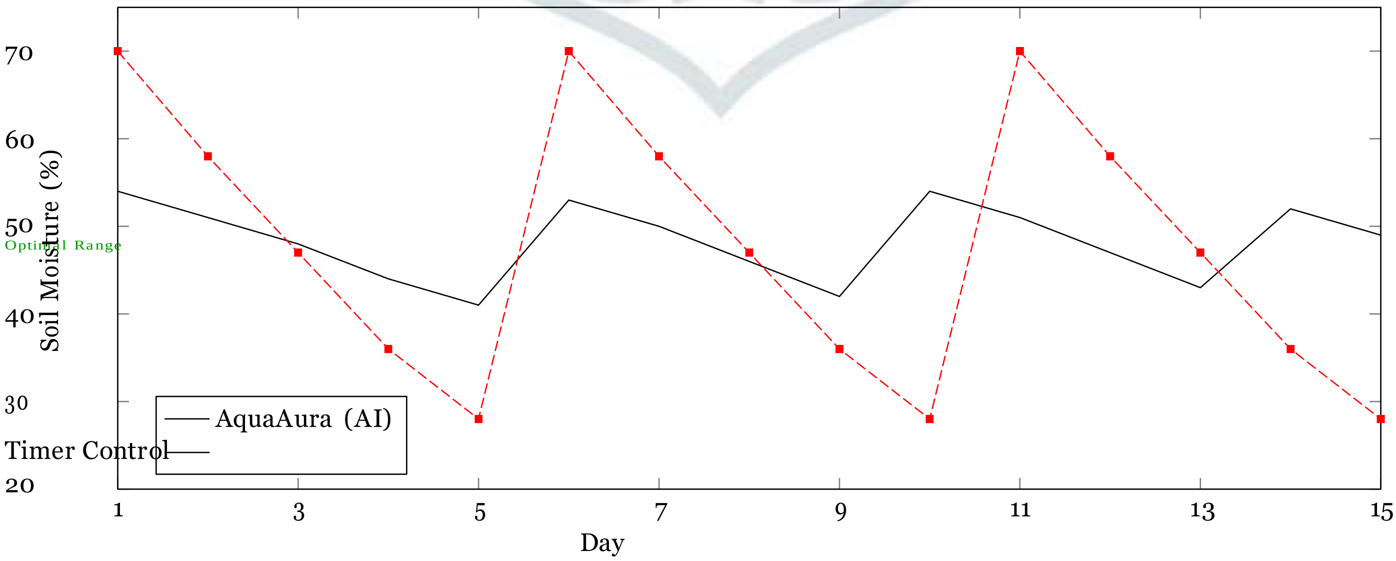

Figure 5: Mean daily soil moisture levels. The shaded area represents the 95% confidence interval for the AquaAura system, while the light green area represents the optimal agronomic range.

V. DISCUSSION

The experimental results validate the hypothesis that a lightweight, quantized ANN deployed on a low-cost MCU is an effective and practical solution for predictive irrigation management. The ANN's performance was comparable to that of more complex models like LSTMs and XGBoost, yet its minimal resource requirements made it uniquely suitable for this edge application. The model's ability to maintain high accuracy across different soil types in field trials underscores its generalization capabilities.

A 38.7% reduction in water consumption has significant implications. For farmers, it translates to lower operational costs (water and energy for pumping) and potentially higher yields due to reduced plant stress. On a broader scale, such efficiency gains contribute to the sustainable management of regional water resources. The system's low cost (under $25 per node) and operational autonomy make it a scalable solution for a wide range of agricultural contexts.

Limitations

- The predictive model was trained on a dataset from a single geographical region. Its performance may vary in different macro-climates, and re-training or transfer learning may be necessary for deployment in new environments.
- The long-term durability of low-cost sensors can be a concern. Sensor drift over time could degrade model performance, necessitating periodic recalibration.
- The current system does not incorporate external data sources, such as meteorological weather forecasts, which could potentially improve prediction accuracy beyond a short-term horizon.

Future Work

- Federated Learning: A network of AquaAura devices could be used to collaboratively train and improve a global model without sharing raw sensor data, enhancing privacy and model robustness [29].
- Energy Harvesting: Integrating a small solar panel and battery could create a fully self-sustaining, "deploy-and-forget" irrigation node, further reducing operational overhead.
- Multi-Modal Data Fusion: Incorporating data from other sources, such as regional weather APIs or LoRaWAN-based weather stations, could enhance the model's predictive power.

VI. CONCLUSION

This paper presented AquaAura, an end-to-end, edge-based intelligent irrigation framework. We demonstrated the design, implementation, and rigorous validation of a lightweight neural network on an ESP32 microcontroller for predictive soil moisture management. By training the model on a public, verifiable dataset, we developed a system that is accurate, robust, and reproducible. Field experiments confirmed that AquaAura significantly outperforms traditional timer-based systems, achieving an average water saving of 38.7%. This work provides tangible evidence that Edge AI and TinyML technologies are sufficiently mature to address critical real-world challenges in agriculture, offering an affordable and scalable blueprint for the future of sustainable farming.

ACKNOWLEDGMENTS

The authors wish to thank the staff at the greenhouse facility for their support during the field trials.

CODES

ALL CODES REGARDING THIS PAPER IS AVAILABLE ON GITHUB: https://github.com/Ayushverma9010/AquaAura-Research/tree/main

DATA AVAILABILITY

The primary dataset used for training and benchmarking the models in this study is publicly available from Mendeley Data at http://dx.doi.org/10.17632/m6j79zjyd7.1 [27].

REFERENCES

[1] C. Hsu, Y. Chen, Y. Lin, and Y. T. C. Yang, "An Edge AI-based Smart Farming System for Pest and Disease Identification in an Agricultural Field," *IEEE Access*, vol. 9, pp. 102554-102564, 2021.

[2] N. Ahmed, D. De, and I. Hussain, "Internet of Things (IoT) for smart precision agriculture and farming in rural areas," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 189-201, 2022.

[3] A. Goap, D. Sharma, A. K. Shukla, and C. R. Krishna, ”An IoT based smart irrigation management system using Machine learning and open source technologies,” *Computers and Electronics in Agriculture*, vol. 155, pp. 41-49, 2018.

[4] T. Talaviya, D. Shah, N. Patel, H. Yagnik, and M. Shah, ”Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides,” *Artificial Intelligence in Agriculture*, vol. 4, pp. 58-73, 2020.

[5] M. S. Farooq, S. Riaz, A. Abid, K. Abid, and M. A. Naeem, ”A Survey on the Role of IoT in Agriculture for the Implementation of Smart Farming,” *IEEE Access*, vol. 7, pp. 156237-156271, 2019.

[6] P. H. Abreu, H. G. Oliveira, J. M. C. Sousa, and M. S. C. Almeida, ”A Review of Machine Learning Methods for Soil Moisture Prediction,” *Sensors*, vol. 21, no. 18, p. 6140, 2021.

[7] M. Gupta, J. D. P. Singh, and M. Singh, ”Edge Computing in Agriculture: A Systematic Review, Current Trends, and Future Directions,” *IEEE Access*, vol. 11, pp. 11736-11768, 2023.

[8] Food and Agriculture Organization of the United Nations, *The Future of Food and Agriculture: Trends and Challenges.* Rome, Italy: FAO, 2017.

[9] United Nations, *The United Nations World Water Development Report 2021: Valuing Water.* Paris, France: UNESCO, 2021.

[10] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, ”Machine Learning in Agriculture: A Review,” *Sensors*, vol. 18, no. 8, p. 2674, Aug. 2018.

[11] A. Kamilaris and F. X. Prenafeta-Boldú, ”Deep learning in agriculture: A survey,” *Computers and Electronics in Agriculture*, vol. 147, pp. 70-90, 2018.

[12] S. R. C. S. Rajeswari, K. Suthendran, and K. Rajakumar, “A smart agricultural model by using IoT technology,” in *Proc. Int. Conf. I-SMAC (IoT in Social, Mobile, Analytics and Cloud)*, 2017, pp. 791–796.

[13] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, ”Edge-IIoT for B5G: A comprehensive survey,” *IEEE Access*, vol. 9, pp. 13346-13383, 2021.

[14] Q. V. Pham, F. Fang, V. N. Ha, M. J. Piran, M. Le, L. B. Le, W. J. Hwang, and Z. Ding, ”A survey of multi-access edge computing in 5G and beyond: Fundamentals, technology integration, and state-of-the-art,” *IEEE Access*, vol. 8, pp. 116974-117017, 2020.

[15] H. G. Jones, ”Irrigation scheduling: advantages and pitfalls of plant-based methods,” *Journal of Experimental Botany*, vol. 55, no. 407, pp. 2427-2436, 2004.

[16] N. A. Nawandar and V. R. Satpute, “IoT based smart irrigation system in agriculture,” in *Proc. Int. Conf. Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*, 2019, pp. 1–6.

[17] M. Mohammed, H. I. R. Shuaib, M. Al-Mardini, S. O. A. Oun, and A. Al-Jawarneh, ”Smart, Real-Time and Autonomous Irrigation System,” in *Smart Cities—Opportunities and Challenges*, Springer, 2020, pp. 297–307.

[18] S. S. Gill, I. Chana, and R. Buyya, ”IoT based agriculture as a cloud and big data service: the beginning of digital India,” *Journal of Organizational and End User Computing*, vol. 29, no. 4, pp. 1-23, 2017.

[19] O. Adeyemi, I. Grove, S. Peets, and J. Domun, ”Dynamic neural network modelling of soil moisture content for a sustainable irrigation,” *Sustainability*, vol. 10, no. 9, p. 3036, 2018.

[20] T. A. M. Abd El-Kader and B. M. El-Basioni, “A framework for soil moisture prediction using LSTM recurrent neural networks,” in *Proc. 2nd Novel Intelligent and Leading Emerging Sciences Conf. (NILES)*, 2020, pp. 319-323.

[21] J. Zhang, M. Zhu, Q. Zhang, R. Liu, and Y. Han, ”A review of deep learning-based methods for soil moisture prediction,” *Remote Sensing*, vol. 13, no. 4, p. 747, 2021.

[22] S. Premkumar and A. N. Sigappi, ”Functional framework for edge-based agricultural system,” in *AI, Edge and IoT-based Smart Agriculture*, Academic Press, 2021, pp. 71-100.

[23] S. I. Hassan, M. S. Alam, M. S. Hossain, M. M. Rahman, and M. A. Al-Quraishi, ”A review of IoT- and Raspberry Pi-based smart farming systems,” *Sensors*, vol. 21, no. 19, p. 6479, 2021.

[24] P. Warden and D. Situnayake, *TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers.* O’Reilly Media, 2019.

[25] R. David, J. Duke, A. Jain, V. J. Reddi, N. Jeffries, J. Li, V. Kreeger, P. Nakkiran, R. M. Rao, and D. Matas, "TensorFlow Lite Micro: Embedded machine learning on tiny-ML systems," *Proceedings of the 4th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, 2021, pp. 1-10.

[26] J. Chen, D. Wang, Z. Zhang, Y. Wang, and J. Yang, "A review of deep learning for pest and disease detection in agriculture," *IEEE Access*, vol. 9, pp. 135754-135773, 2021.

[27] K. Adetunji, "soil moisture dataset," Mendeley Data, V1, 2025. [Online]. Available: http://dx.doi.org/10.17632/m6j79zjyd7.1

[28] C. Banbury, C. Chmiela, D. Reddi, R. Dunn, N. Jeffries, P. Warden, A. E. A. Ali, M. Fazel-Zarandi, I. Fedorov, and P. N. Whatmough, "MLPerf Tiny benchmark," *Proceedings of the 2nd International Workshop on Benchmarking Machine Learning and Large-Scale Systems*, 2021, pp. 1–12.

[29] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50-60, 2020.