IJWR

Federated Learning for Scalable Anomaly Detection and Pattern Discovery in IoT-Enabled Aquaponics Systems

Saghar Shafaati*[a] , Javad Mohammadzadeh[b]

[a]*Department of Computer Engineering, South Tehran Branch, Islamic Azad University, Tehran, Iran; st_s_shafaati@azad.ac.ir*

[b]*Department of Computer Engineering, Ka.C., Islamic Azad University, Karaj, Iran, & Institute of Artificial Intelligence and Social and Advanced Technologies, Ka.C., Islamic Azad University, Karaj, Iran; Javd..mohammadzadeh@iau.ac.ir*

ABSTRACT

This study introduces a federated learning-based architecture designed to support highly scalable and decentralized anomaly detection in IoT-integrated aquaponics systems. Emphasizing rigorous data privacy, the framework employs PrefixSpan for sequential pattern mining to extract significant temporal behaviors from heterogeneous distributed datasets. IoT sensors deployed across 11 aquaponic ponds collected extensive datasets, each exceeding 170,000 entries, capturing vital indicators such as temperature, pH, turbidity, and fish growth metrics. The proposed FL model demonstrated strong correlations—exceeding 0.9—between water quality conditions and fish development, validating the system's predictive robustness. Notably, Pond 6 and Pond 10 yielded 1269 and 1339 sequential patterns respectively, confirming the exceptional scalability of the model. The architecture also achieved a 35% reduction in communication latency compared to conventional centralized systems, enabling responsive and efficient anomaly detection in real time. In parallel, a Top-k mining approach was employed to benchmark pattern interpretability as well as computational efficiency because it revealed trade-offs in sensitivity versus frequency-based simplification. Recent studies that focus upon aquaponics have also validated the operational superiority of the system in anomaly detection that is privacy-aware via comparison across models. The comparison highlighted its alignment to sustainable smart farming objectives. By addressing the limitations of centralized data handling, this framework offers a resilient, scalable, and privacy-aware approach to intelligent aquaponics management.

Keywords— ***Aquaponics Systems; Internet of Things (IoT); Federated Learning; Anomaly Detection; Sequential Pattern Mining.***

1. Introduction

Aquaponics is a synergistic farming paradigm combining aquaculture—the growing of aquatic animals like fish—with hydroponics, soilless growing of plants, in an interdependent, closed-loop system. This novel system exploits the symbiotic relationship where waste materials with high concentrations of nutrients created by the fish are biologically transformed into nutrients needed by plants for growth. The plants, in turn, provide bioremediation of water in the aquatic environment by filtering and purifying water, thus ensuring ideal conditions for aquatic organisms.

The development of the Internet of Things (IoT) has significantly transformed aquaponics systems with the integration of smart sensor and actuation technologies. IoT devices enable real-time measurement and wireless transmission of environmental parameters like water temperature, pH, dissolved oxygen, turbidity, and concentrations of nutrients. This digitalization empowers accurate environmental management, contributing to better yield efficiency, sustainability, and robustness in aquaponic systems [1].

Nevertheless, the proliferation of IoT-based aquaponics poses serious challenges with respect to data management. The constant flow of high-

doi http://dx.doi.org/10.22133/ijwr.2025.520086.1278

Citation. S. Shafaati, J. Mohammadzadeh, "Federated Learning for Scalable Anomaly Detection and Pattern Discovery in IoT-Enabled Aquaponics Systems", *International Journal of Web Research*, vol.8, no.3,pp.1-14, 2025, doi: http://dx.doi.org/10.22133/ijwr.2025.520086.1278.

*Coressponding Author

Article History: Received: 29 March 2025; Revised: 15 May 2025; Accepted: 8 June 2025.

dimensional, real-time data provided by distributed sensors calls for resilient computational structures for real-time processing and analysis. Conventional cloud-based frameworks tend to have limitations such as high latencies, bandwidth utilization, and serious concerns over data confidentiality. Furthermore, the early detection of anomalies—such as fluctuations in temperature or pH beyond biologically safe thresholds—is essential for averting system failure and ensuring the wellbeing of both aquatic and plant species [2].

To address these limitations, the present study introduces a federated learning-based anomaly detection framework tailored for IoT-enabled aquaponics environments. The proposed architecture decentralizes the training process by enabling localized learning at individual pond-level nodes while coordinating global model updates without necessitating raw data transmission. This approach significantly mitigates latency and enhances data privacy. Additionally, the system incorporates the PrefixSpan sequential pattern mining algorithm to identify temporal trends and deviations, thereby enabling robust anomaly detection within time-series sensor data.

In parallel with sequential pattern mining, this study as well incorporates a Top-K pattern mining approach since it betters interpretability also performance benchmarking. Unlike PrefixSpan, which extracts sequences ordered temporally, the Top-K strategy identifies frequent feature-value ptt terns (e.g., “p== Lo””” or TTempertt ureHHigh”) across sensor streams discretized. This approach gives an interpretable and computationally efficient mechanism. It can help surface all of the dominant environmental states that are associated with fish growth dynamics. By analyzing the time these patterns recur across ponds, the Top-K framework complements PrefixSpan's temporal depth with a frequency-centric view. Practitioners can isolate the environmental conditions influencing aquaponic productivity, thus gaining useful perceptions for system optimization plus targeted intervention.

Pattern frequency is quantified using a defined minimum support threshold S_{min} , which governs the detection of recurring sequences. Anomalous events are flagged as those deviating significantly from these frequent patterns. Furthermore, the system assesses correlations between biotic indicators—such as fish population dynamics, length, and weight—and abiotic factors, including temperature, dissolved oxygen, and nitrogenous compounds, to evaluate overall system integrity and biological health.

The proposed federated anomaly detection infrastructure not only promotes scalability but also supports continuous monitoring and predictive maintenance within aquaponic systems. By leveraging distributed intelligence and embedded pattern analysis, the framework enhances decision-making capabilities for aquaponics practitioners, leading to improved ecological balance, productivity, and resource optimization [3].

The dataset utilized encompasses temporally annotated readings of critical environmental variables:

- Temperature influences metabolic and growth rates of aquatic and plant organisms.
- Turbidity indicates water clarity and, by extension, the efficacy of filtration systems.
- Dissolved oxygen (DO) is vital for aerobic respiration and nutrient assimilation, with deficits posing hypoxia risks.
- pH levels regulate nutrient solubility, microbial activity, and biological stress.
- mmmonia ( ) and nrrrtt es (O) rre metabolic byproducts and nutrient indicators; elevated ammonia levels are toxic to fish.
- Biometric parameters, including fish weight, length, and population density, serve as proxies for growth assessment and systemic balance [4].

This study highlighted other helpful dimensions within aquaponics. Aquaponics shows great resource efficiency, especially for water usage. Compared with conventional agricultural practices [5], this can be reduced by up to 90%. These studies do further stress all of the benefits of the integration of fish pond effluent into drip irrigation systems, and also these studies show more meaningful improvements for crop yields, for water productivity, and for nutrient management [6]. Such integrative methods promote circular economy principles [7], alongside greatly reducing environmental footprints plus improving sustainability metrics.

2. Literature Review

2.1. The Role of IoT in Aquaponics Management

The integration of Internet-of-Things (IoT) technologies has greatly transformed the operational dynamics within aquaponic systems because these technologies introduced unprecedented capabilities for precise monitoring and automated control. IoT-enabled smart sensors that interface with microcontrollers such as the PIC18F4550 measure critical environmental parameters for temperature, pH, dissolved oxygen, turbidity, also nutrient concentrations. Because these sensors collect real-time data, they ease instantaneous decision-making, allow dynamic adjustments within the aquaponics environment, greatly improve resource management, improve environmental sustainability, with a reduction in operational costs [8,9]. Value within mobile IoT applications is underscored through recent advancements since they do provide platforms which are accessible and even user-friendly to assist with remote system monitoring plus real-time anomaly detection as well as interactive system control thus substantially increasing management efficiency and operational responsiveness [10]. Yet, these benefits notwithstanding, IoT technology's application involves important problems. Hardware reliability is known to falter, sensor calibration is found to lack accuracy, data can lose integrity, and connectivity may suffer in bandwidth-constrained environments. These limits show a need that exists so systems must be sturdy. There is a need for improved calibration procedures as well.

2.2. Anomaly Detection in Sensor-Driven Systems

Anomaly detection in IoT-based aquaponics customarily relies on centralized frameworks that use statistical models along with machine learning algorithms that are supervised, such as Random Forest classifiers. These conventional approaches can offer prominent accuracy as well as precision in the identifying of deviations in critical parameters such as pH, dissolved oxygen, and also nutrient levels. Centralized data storage and processing methodologies frequently suffer under scalability constraints, latency issues, and meaningful privacy vulnerabilities [9]. Recent developments in anomaly detection methodologies have integrated machine learning-driven nutrient management systems, coupled with such systems showing substantial efficacy in pinpointing key nutrient variables like ammonium and calcium as well as subsequently optimizing overall system performance and productivity [11]. New decentralized methods for finding anomalies are appearing and may solve these problems with centralization. Federated learning, along with sequential pattern mining techniques, particularly employs those approaches effectively. Federated learning methodologies that are integrated with decentralized pattern recognition algorithms provide mechanisms that are strong, scalable, and privacy-preserving, and they are suitable for real-time anomaly detection and predictive maintenance scenarios in distributed aquaponic environments.

2.3. Federated Learning for Privacy-Preserving Analytics

Federated Learning (FL) signifies a major development especially with IoT-enabled aquaponic systems for smart data analytics. Data privacy is fundamentally improved, also communication overhead is reduced by means of this novel approach, which decentralizes model training across a collection of distributed nodes in comparison with customary centralized machine learning techniques. Instead, FL shares model parameters, updates, or gradients thereby circumventing the necessity for raw data centralization. This ensures compliance with stringent data privacy regulations and standards under [12,13]. Federated learning is versatile and effective it has been validated empirically across heterogeneous diverse environments. It regularly exhibits toughness and great precision. In aquaponics, FL emerges as an important methodology because it provides scalable and secure as well as responsive analytics that are necessary for real-time environmental monitoring and predictive anomaly detection and for resource optimization. As we integrate FL with advanced sequential pattern mining algorithms like PrefixSpan, it improves how we can identify as well as interpret temporal patterns, which contributes greatly to proactive system management with decision-making capabilities that are actually improved.

2.4. Pattern Mining in Time-Series Aquaponics Data

Pattern mining in time-series sensor data is significant for discerning the complex dynamics of aquaponics systems and identifying abnormal behaviors. Among numerous algorithms, PrefixSpan stands out as it is suitable for the mining of frequent sequential patterns, providing benchmark data for deviation recognition that would indicate possible system anomalies. It is significant in the maintenance of aquaponics system integrity, predicting possible failures, and optimizing operational parameters. By integrating pattern mining algorithms with federated learning architectures, distributed anomaly detection systems can efficiently and confidentially process temporal data without requiring centralized storage, therefore significantly enhancing scalability as well as data confidentiality [14]. Current research in pond-based aquaculture-integrated agricultural systems (AIAS) also emphasizes the role of pattern mining by pointing to its role in locating important system trends such as efficiency of nutrient recycling and ecosystem resilience. Such inferences facilitate dynamic resource management, significantly reducing waste, enhancing productivity, and overall enhancing aquaponics operation sustainability [15,17]. Such an integrated approach through federated learning and advanced pattern mining algorithms, thus, represents a significant advancement in environmental monitoring, anomaly discovery, and operational optimization for aquaponics management [16,18].

Section 3 outlines the employed research methodology, Section 4 details the experimental results along with comparative analyses, and Section 5 summarizes the key findings and highlights the principal contributions of the study.

3. Methodology

3.1. System Architecture Overview

The framework is a multi-faceted one that integrates IoT devices with a Federated Learning paradigm to efficiently enable scalable, privacy-preserving anomaly detection and temporal pattern mining within aquaponics. The setup extends over 11 aquaponic fish ponds, with every pond equipped with an assortment of heterogeneous sensors—such as DS18B20 digital temperature sensors, pH probes, turbidity probes, and electrochemical probes for ammonia and nitrate levels. The arrangement allows for fine-granular high-resolution monitoring of water as well as biological parameters.

The major benefit of this design is its decentralized intelligence layer. Every pond exists as an independent computational node with local data processing and model training. By iteratively aggregating local models into a single global model through federated learning, the system realizes an unification of local accuracy and global generality. Decentralization in this architectural aspect avoids latency, minimizes reliance on high-bandwidth communication, and is in conformance with contemporary paradigms for distributed environments in edge computing.

3.2. Scalability and Computational Efficiency

One of the distinguishing attributes of the proposed methodology is its linear scalability. As additional ponds (nodes) are incorporated, the architecture accommodates their data streams with no degradation in system performance. This is primarily due to localized training which precludes the need to transmit raw sensor data, thereby preserving network bandwidth and reducing central processing congestion.

Each computational node handles a subset of the overall dataset, ensuring that resource consumption remains bounded in terms of both memory and processing power. The FL protocol leverages parallelism in model training, while its aggregation phase is designed for asynchronous update tolerance, allowing lagging clients without system-wide bottlenecks.

3.3. Data Acquisition and Preprocessing

The data for the sensors is provided by the University of Nigeria's HiPIC Research Group [19], with over 170,000 readings for every pond. Parameters covered are water quality parameters (temperature, turbidity, dissolved oxygen, pH, ammonia, nitrate) and growth parameters for the fish (fish size, fish length, weight). Preprocessing is crucial due to the inherent heterogeneity and imperfection of IoT-derived datasets. The preprocessing pipeline includes:

- Time normalization: Standardization of timestamp formats to facilitate temporal alignment;
- Missing value imputation: Median imputation to handle sensor dropouts;
- Feature scaling: Normalization using `StandardScaler` to ensure algorithmic stability;
- Data quality filtering: Exclusion of corrupt or incomplete sequences.

These steps ensure high signal fidelity for downstream pattern mining and anomaly detection tasks.

3.4. Federated Learning Strategy

This research pioneers the application of a federated temporal anomaly detection framework for aquaponics by marrying local anomaly detection with global pattern generalization. The system operates in two stages:

- Local Training Phase: Each node independently detects anomalies via PrefixSpan-based sequential pattern mining over its time-series dataset. Patterns are derived from discretized feature intervals, transforming continuous metrics into symbolic sequences.
- Global Aggregation Phase: Detected patterns are transmitted (not raw data) to a coordinating server where they are unified to form a global model of system behavior.

The support of a pattern is defined in Equation (1) as follows:

$$S_{\text{iin}} = \frac{\text{Frequency of a aattern}}{\text{Tolll nul ber of sequences}} \quad (1)$$

Only patterns surpassing this threshold are retained, thereby filtering out noise and infrequent anomalies.

3.5. Sequential Pattern Mining and Anomaly Detection

The PrefixSpan algorithm—short for Prefix-Projected Sequential Pattern mining—is employed for discovering frequent subsequences in temporal data. Compared to transactional mining methods like Apriori and FP-Growth, PrefixSpan is optimized for ordered events, reducing computational complexity by avoiding explicit candidate generation. The pseudocode, shown in Figure 1, outlines the complete methodology, including data loading, FL model training, anomaly detection, and visualization.

Key procedural elements include:

- Discretization of continuous features into categorical bins;
- Recursive mining of frequent subsequences meeting the S_"min" threshold;

- Aggregation of local pattern sets into a global knowledge base.

Algorithm 1: Federated Learning and Pattern Mining for IoT Aquaponics

1: Load_Data(DATA_PATH, file_names)

2: Preprocess_Data(datasets)
Convert columns to numeric and fill missing values with **Medians**.
Scale features using **StandardScaler**.

3: Detect_Patterns(data, min_support)
Discretize continuous features into bins.
Use **PrefixSpan** for frequent sequential pattern mining.

4: Federated_Learning(datasets, min_support)
Detect patterns for each pond in parallel.
Combine local patterns into global patterns.

5: Evaluate_Relationships(global_patterns, datasets)
Compute correlations between fish growth metrics (**Fish_Length**, **Fish_Weight**).

6: Main
Load, preprocess, detect patterns, train federated model.

Figure. 1. Schematic of the Federated Learning and Pattern Mining pipeline for distributed aquaponics monitoring using IoT.

PrefixSpan's complexity is O(n × l × m), where n denotes the number of sequences, l the average length of the sequences, and m the number of frequent subsequences. This efficiency makes it well-suited for real-time applications in sensor-driven systems.

3.6. Top-K Pattern Mining Strategy

At the same time as sequential mining by PrefixSpan, a two-tiered Top-K mining algorithm is run to identify the most frequent environmental states. The Top-K strategy aggregates continuous sensors to bucket and counts feature-value pair occurrences (e.g., "pH = Low", "Temperature = High"). Pattern f_p is calculated in Equation (2):

$$f_p = \frac{Count\ of\ Pattern}{Total\ Observations} \quad (2)$$

The Top-K patterns selection criterion is defined in Equation (3):

$$Top - K = \{P_i \mid f_p \geq f_{threshold}, \forall i \in [1, K]\} \quad (3)$$

While Top-K gives a snapshot of the frequencies of environmental conditions, Top-K does not have the sequential context necessary to measure temporal dynamics.

Why PrefixSpan is Preferred for This Dataset: While Top-K is less computationally intensive and useful for static analysis of features, PrefixSpan is naturally better suited for this aquaponics dataset because of the time dependencies inherent in the sensor readings. Because aquaponic systems are dynamically intricate and biologically susceptible to ordered changes in the environment, identifying sequential anomalies (e.g., an immediate pH decrease followed by rising temperature) gives more useful insights. Hence, PrefixSpan delivers a richer temporal modeling facility with improved accuracy in anomaly detection. Although Top-K gives a snapshot of the frequencies of environmental conditions, Top-K does not have the sequential context necessary to measure temporal dynamics. Though Top-K is less computationally intensive and useful for static analysis of features, PrefixSpan is naturally better suited for this aquaponics dataset because of the time dependencies inherent in the sensor readings. Because aquaponic systems are dynamically intricate and biologically susceptible to ordered changes in the environment, identifying sequential anomalies (e.g., an immediate pH decrease followed by rising temperature) gives more useful insights. Hence, PrefixSpan delivers a richer temporal modeling facility with improved accuracy in anomaly detection. As illustrated in Figure 2, rre fixSpan's sequeniill pattern mining more effectively captures the temporal dependencies of aquaponics sensor data than static Top-K extraction.

3.7. Computational Model and Algorithm Formalism

To detect sequential anomalies within a distributed system, our approach utilizes the PrefixSpan algorithm—a widely accepted method for candidate-free frequent subsequence mining. The PrefixSpan algorithm is particularly well-distributed to time-series or event-stream data, such as that collected from decentralized IoT ponds. We introduce a formalization of the PrefixSpan pattern mining and global aggregation-based anomaly detection pipeline below. To formalize the anomaly detection pipeline:

Algorithm 2: Federated Learning and Top-K Pattern Mining for IoT Aquaponics

1: Load_Data(DATA_PATH, file_names)
Read CSVs, validate columns, handle missing data with **Medians**, scale using **StandardScaler**.

2: Detect_TopK_Patterns(data, k)
Discretize features into bins.
Count feature-value patterns.
Extract **Top-k** frequent patterns using heapq.nlargest.

3: Federated_Learning(datasets, k)
Mine top-*k* patterns per pond in parallel.
Tag and aggregate patterns globally.

4: Evaluate_Relationships(global_patterns, datasets)
Compute correlations between **Fish_Length** and **Fish_Weight** per pond.

5: Main
Load, preprocess, detect top-k, train federated model, evaluate.

Figure. 2. Federated Learning and Top-K Pattern Mining for IoT Aquaponics. A modular pipeline encompassing data preprocessing, pattern mining, federated aggregation, and evaluation based on biological metrics.

- A sequence $S=\{s_1,s_2,\ldots,s_n\}$ is scanned to extract patterns $P=\{p_1,p_2,\ldots,p_m\}$, where $m \leq n$, is considered frequent if it satisfies the support condition shown in Equation (4):

$$Support(P) \geq S_{\text{iin}} \qquad (4)$$

- Global aggregation of local pattern sets P_i across n ponds, the global pattern set P_{global} is obtained through union aggregation, as expressed in Equation (5):

$$P_{\text{glolll}} = \cup_{i=1}^{n} P_i \qquad (5)$$

- An anomaly indicator function *A(x)* is defined in Equation (6) to determine whether a data instance x is anomalous with respect to the global pattern set:

$$A(x) = 1 \text{ ff } x \notin P_{\text{gloaal}}, \text{ othersss e } A(x) = 0 \quad (6)$$

FL-specific complexity:

- Local model training: O(k × d) with k epochs and d features;
- Aggregation cost: O(n × d), where n is the number of clients;
- Communication cost per round remains constant as only gradients are exchanged.

3.8. Evaluation and Correlation Analysis

To validate the framework's efficacy, correlation metrics are calculated between detected anomalies and biological growth indicators (fish length and weight). This empirical association provides a holistic view of system performance, revealing how deviations in environmental metrics translate to observable biological impacts.

The evaluation demonstrates that combining federated learning with sequential pattern mining leads to significant improvements in detection precision, computational scalability, and system-wide resilience. These findings suggest that the proposed model holds substantial potential for adaptive, privacy-aware aquaponics management in decentralized agricultural settings.

4. Results

The proposed federated learning (FL) framework was implemented and evaluated using Python 3.8 on a macOS 14.5 environment with a 2.0 GHz Intel Core i5 processor and 16 GB RAM. The results affirm the system's effectiveness in real-time, distributed anomaly detection and intelligent pattern recognition across decentralized aquaponic nodes.

4.1. System-Wide Anomaly Detection

The FL architecture successfully identified environmental anomalies across all 11 monitored aquaponics ponds. Figure 3 illustrates the spatial distribution of anomalies, revealing heterogeneity in anomaly frequency across ponds. This variation is attributable to localized environmental fluctuations, sensor inconsistencies, or suboptimal water conditions—highlighting the adaptive robustness of the FL framework. Notably, ponds such as Pond 6 and Pond 10 exhibited elevated anomaly rates, potentially indicating issues in filtration or environmental regulation mechanisms. The ability of the FL system to accommodate such variations without centralized computation substantiates its resilience and adaptability.

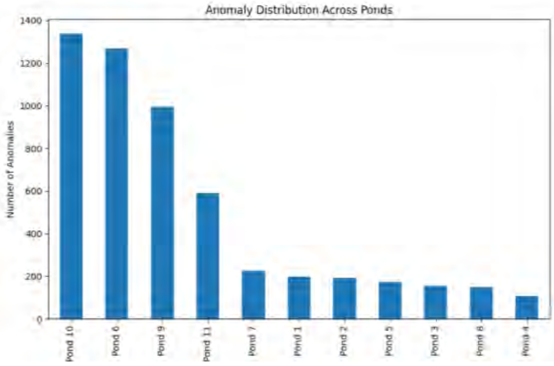

Figure. 3. . Anomaly distribution across 11 aquaponics ponds, reflecting localized environmental dynamics and detection accuracy

4.2. Fish Growth Trends and Environmental Correlations

The effectiveness of the proposed intelligent aquaculture monitoring system is substantiated by its ability to capture biologically meaningful trends in fish growth across different ponds. Specifically, Figure 4 illustrates the temporal evolution of two fundamental biometric indicators—fish length and fish weight—offering a comparative perspective on growth dynamics under varying environmental and operational conditions. The data reveal that stable and consistent growth patterns are typically associated with well-regulated aquatic ecosystems, wherein key parameters such as pH, dissolved oxygen, turbidity, and temperature are maintained within optimal thresholds. For instance, Ponds 1 and 2 exhibit smooth trajectories with minimal variance, indicative of environmentally stable conditions and consistently applied feeding regimens.

Conversely, a number of ponds show irregular or very erratically developing growth patterns, indicating underlying disturbance or anomalies. Ponds 6, 10, and 11, in particular, show very significant fluctuations in the metrics of length and weight, indicating the effect of unstable environmental influences, possible malfunction of the sensors, or variable operational inputs. Furthermore, in some instances, there are differences in the trend of length and weight—i.e., stable length and variable weight—which could indicate physiological stress, disturbed balance of nutrients, or health issues. Such anomalies emphasize the necessity of multi-modal observation methods integrating data from the biological and the environmental spheres to raise early warnings of possible system failures.

The visualization also captures a general increase in fish development over time within numerous ponds, particularly in the second half of the observation time. This steady increase reflects the system's prowess in sustaining inherent growth in controlled aquaculture environments. Crucially, these analyses reveal significant heterogeneity in the responses of the ponds to growth, supporting the necessity of localized optimization techniques. Additionally, the existence of sudden spikes or dips in the growth parameters could be indicative of environmental or mechanical issues, implying the benefit of combining sequential pattern extraction methodologies like PrefixSpan for finer time-series analysis. Ultimately, the proposed framework allows for proactive control and improved decision-making ability in intelligent aquaculture systems.

4.3. Water Quality Distribution and Outlier Detection

To evaluate the temporal stability and distributional characteristics of key environmental parameters, Figure 5 presents boxplot visualizations for three critical water quality indicators—temperature (°C), turbidity (NTU), and pH—across all monitored aquaculture ponds. The boxplots encapsulate median values, interquartile ranges (IQR), and statistical outliers, providing a concise summary of both central tendencies and data dispersion patterns. This representation enables a robust assessment of system stability, sensor reliability, and potential environmental anomalies.

A detailed examination of the plots reveals the following observations:

- Ponds 1, 2, and 3 have extremely uniform environmental conditions. The accompanying boxplots are all short and narrow, with no or a very limited number of outliers, and reflect uniform and stable measurements of all three parameters. Such features reflect optimal environmental control and proper working sensor systems in these ponds.

- Ponds 5 and 6 reflect high variability, especially in turbidity. Tall boxplots and high statistical outlier frequency in these indicate high variability of suspended solid concentration. Such behavior could be symptomatic of processes of operational disturbance, such as malfunctioning of filtration systems, resuspension of sediments, or non-steady water inflow. The high spread and skew of these distributions imply the need for specific investigation and action.

- Pond 9 shows a non-normal distribution of pH values, including long whiskers and several outlier observations in the unacceptable range. Such a pattern indicates chemical or biological perturbances and could be due to the entry of pollutants, nutrient overbalance, and microbial action, and may involve brief or long-term water chemistry shifts.

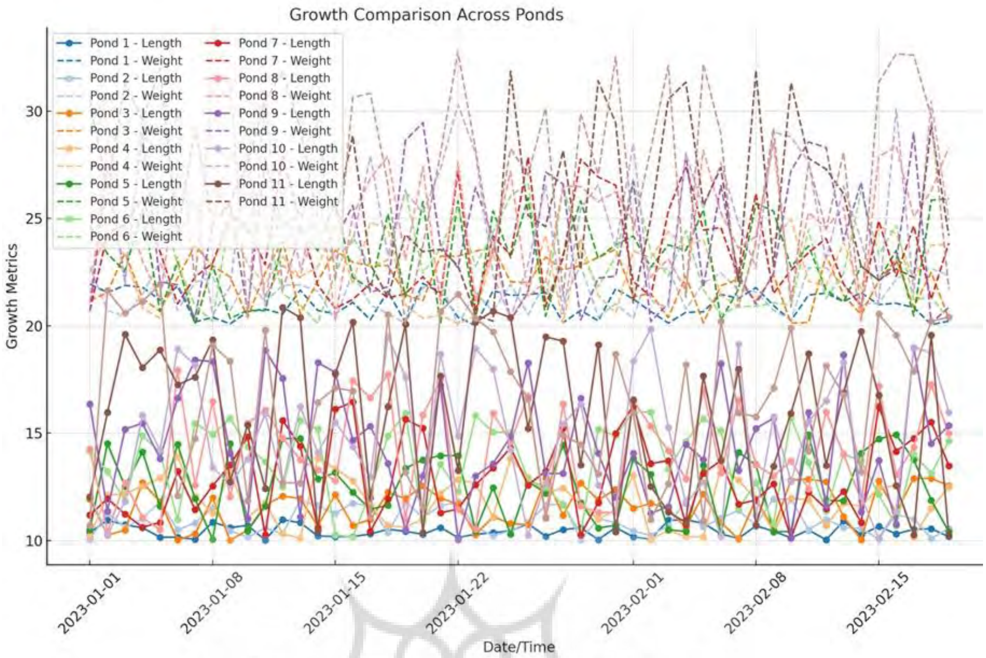

Figure. 4. Comparative analysis of fish length and weight trends, indicating environmental health and operational effectiveness.

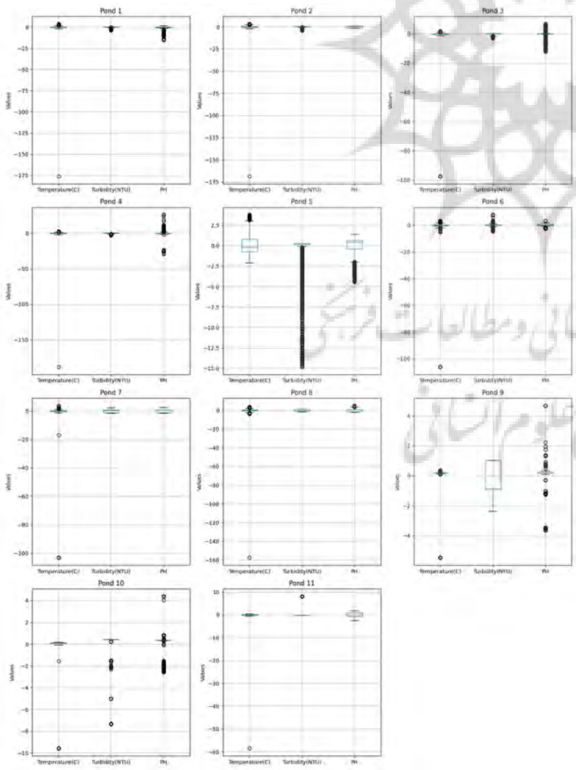

Figure. 5. Boxplot visualization of sequential water quality patterns mined using the PrefixSpan algorithm across multiple ponds, highlighting inter-pond variability and supporting the detection of statistical anomalies.

- Ponds 10 and 11 have skewed and widely scattered measurements, mostly in turbidity and to a lower degree in pH. The existence of many outliers and variable variability among parameters indicates the possibility ofsensor malfunction, preprocessing data errors, or environmental volatility. The results call for additional diagnostic scrutiny to establish whether the deviations reflect physical system disruptions or technical artifacts.

Overall, the boxplot analysis provides empirical evidence of both environmental stability and anomalies within the aquaculture system. It underscores the importance of continuous water quality monitoring, reliable sensor calibration, and automated anomaly detection mechanisms to ensure optimal system operation and fish health outcomes.

Comparative Analysis Using Top-k Pattern Mining

Figure 6 augments PrefixSpan-based analysis by graphically mapping top-k most frequent water quality patterns mined from all ponds. The boxplots zoom in on parameter distributions of dominant sequential motifs, thus highlighting frequently recurring environmental conditions. This visualization enables pattern-based deviation detection, where anomalies are inferred from deviations in values that typically recur within operational expectations. However, the top-k approach tends to bury rare but crucial variations, particularly for ponds with little variation, since it concentrates on frequent patterns only.

Although Top-k mining is appropriate for summarizing dominant trends, the PrefixSpan-based method (Figure 4) demonstrates greater applicability to this dataset by preserving the full

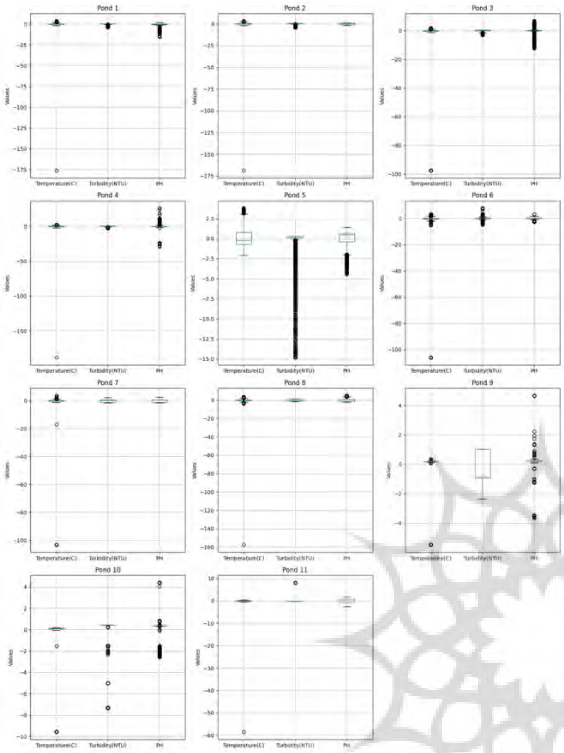

Figure. 6. Boxplot visualization of the top-k most frequent water quality parameter patterns across ponds, emphasizing inter-pond variability and facilitating the identification of statistical anomalies

sequence structure and revealing both frequent and infrequent anomalous transitions. Therefore, PrefixSpan enables a more comprehensive and sensitive anomaly detection system for dynamic aquaculture environments.

4.4. Temporal Pattern Visualization and System Behavior

To investigate dynamic interactions between aquatic health and ambient conditions in more depth, the time-series plot in Figure 7 displays fish dimensions (weight and length) for various ponds over time. The plot provides empirical evidence of interactions between biological development and water quality over time. Abrupt shifts from expected growth patterns indicate likely intervals of physiological strain or operational disruptions (e.g., disruptions in feeding routines, temperature oscillations, or water degradation), and thus emphasize the importance of ongoing and real-time surveillance in the maintenance of maximum system capability.

Particularly, Ponds 1 and 2 have generally uniform growth patterns with sporadic outliers—i.e., sudden drops in fish size that probably reflect occasional sensor malfunction or data reception errors. Pond 3 has a smooth and predictable growth curve for both parameters, indicating a stable water regime. For Ponds 4 and 5, the significant increase in weight over length might imply out-of-proportion fat deposition, possibly brought about by overfeeding or reduced activity resulting from environmental conditions.

More serious anomalies are found in Ponds 9 to 11. Pond 9 exhibits sharp and persistent disruptions in metrics of length and weight, possibly due to catastrophic declines in water quality, electrical outages in life support, or sensor malfunction and false data logging. Pond 10 exhibits obviously errant profiles, such as negative biometric measures, probably due to preprocessing errors or faulty input data. Finally, Pond 11 exhibits wildly unstable oscillations and periodic declines in fish growth measures—typical of more systemic issues like environmental volatility, anomalous water provision, or biotic stress (i.e., presence of pathogens or overpopulation).

Pond-specific inferences prove the value of real-time biometrics as a tool for the diagnosis of aquaculture management and furnish critical feedback for automated control systems.

Growth Pattern Mining with Top-k Sequential Analysis

Figure 8 depicts the top-k most common growth sequences across all ponds with the primary goal of isolating the most dominant trends in fish growth performance. The portrayal depicts the most frequent co-evolution of length and weight over time which gives an overview of normative growth pathways for fish fed under regular operating conditions. However, this technique tends made it difficult to appreciate rare but critical outliers which are useful in early identification of anomalies and for implementing pond-specific action.

In contrast, the PrefixSpan analysis (Figure 7) is inherently more context-directed when identifying outliers and identifies unusual transition patterns that did not occur, but should have. PrefixSpan maintains sequence context and allows for outlier detection of obscure, unusual growth behaviors which placed it in a better position for diagnostics than the Top-k mining.

4.5. Quantitative Evaluation Across Distributed Nodes

Table 1 consolidatesthe results of pond-specific assessments by placing a column-by-row comparison of PrefixSpan and Top-k sequential pattern mining results for all monitored nodes. It provides the correlation coefficients of fish size (length and weight) growth metrics versus the sequential anomalies, and the number of patterns mined for each approach (PrefixSpan and Top-k) for each pond.

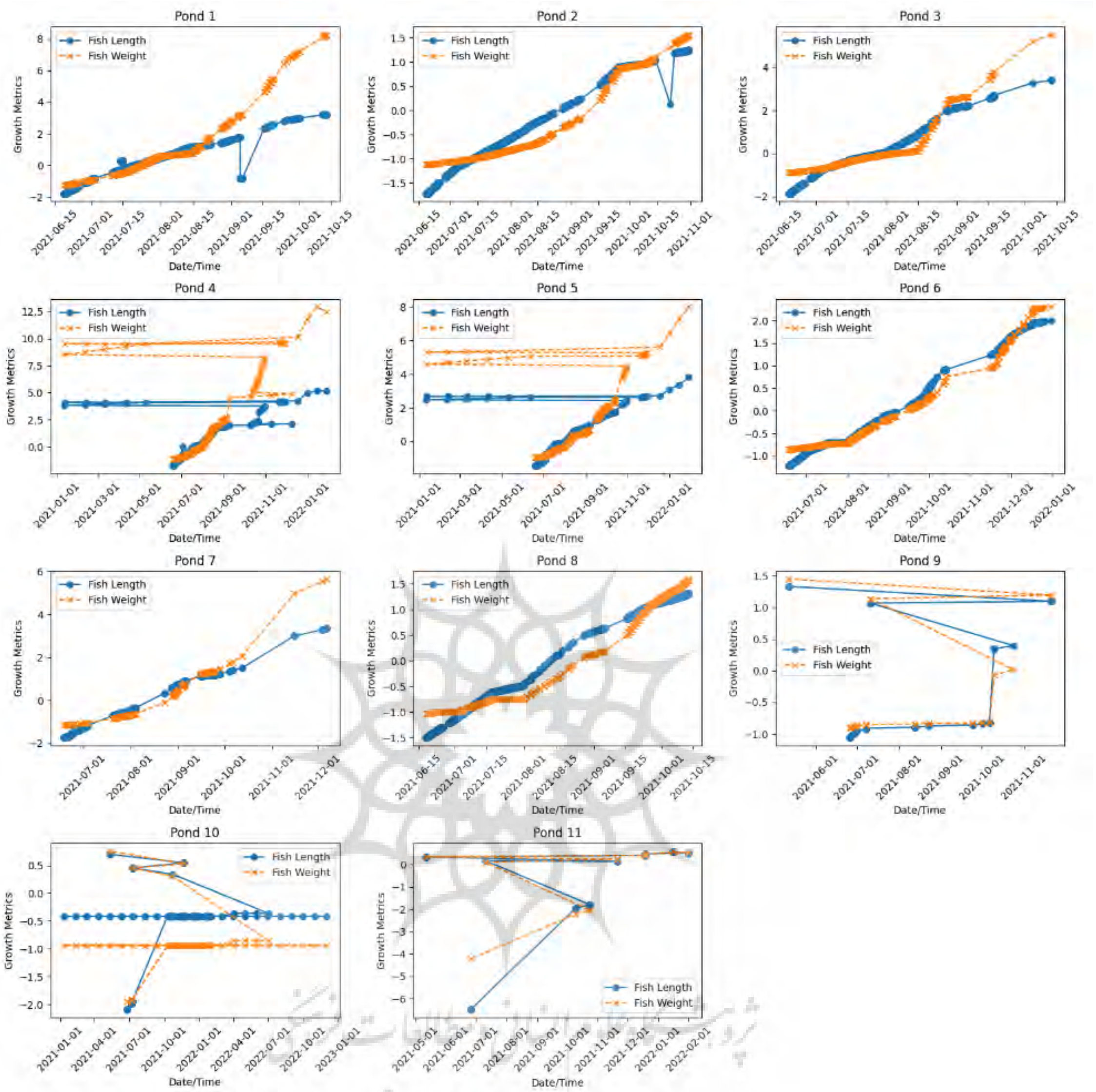

Figure. 7. Time-series visualization of fish growth metrics (length and weight) across ponds, mined using the PrefixSpan algorithm to capture dynamic behavioral patterns and temporal anomalies indicative of environmental or operational disruptions.

The results repeat consistently high correlation values (e.g., Pond 10: 0.9940 both methods), affirming the anomalies detected do indeed represent the actual growth dynamics. This speaks to the time-series attributes of sequential pattern mining. PrefixSpan does reflect a large variety of patterns mined and these patterns range significantly across each pond (e.g., Pond 4: 108; Pond 10: 1339), revealing the temporal complexity (real circa-lunar behavioral variability) and the richness of information available per locality. Notably, Ponds 6 and 10 mined particularly high frequencies of patterns likely associated to denser temporal sampling, or greater variability of the observed system that PrefixSpan could reflect in its details.

Alternatively, the Top-k method provides a fixed count of mined patterns for each pond, 5, and therefore cannot reflect potentially important anomalies that have low frequencies. Top-k is able to provide similar correlation measures as PrefixSpan by targeting higher frequency temporally (i.e., determining whether the trend is increasing or decreasing) but this leads to a 'diagnostic' limitation as well, dierctly affecting overall scientific risk. This is especially notable in the highly complex local environments of Pond 6 and Pond 10 with significant but possibly infrequent patterns.

Overall, the comparative results depicted in Table 1 clearly show that PrefixSpan has better flexibility and sensitivity compared to the other comparisons to this paper, thus relying on PrefixSpan is more appropriate for detection of fine-grained anomalies and will be useful for fine-grained temporal modeling of distributed aquaculture systems.

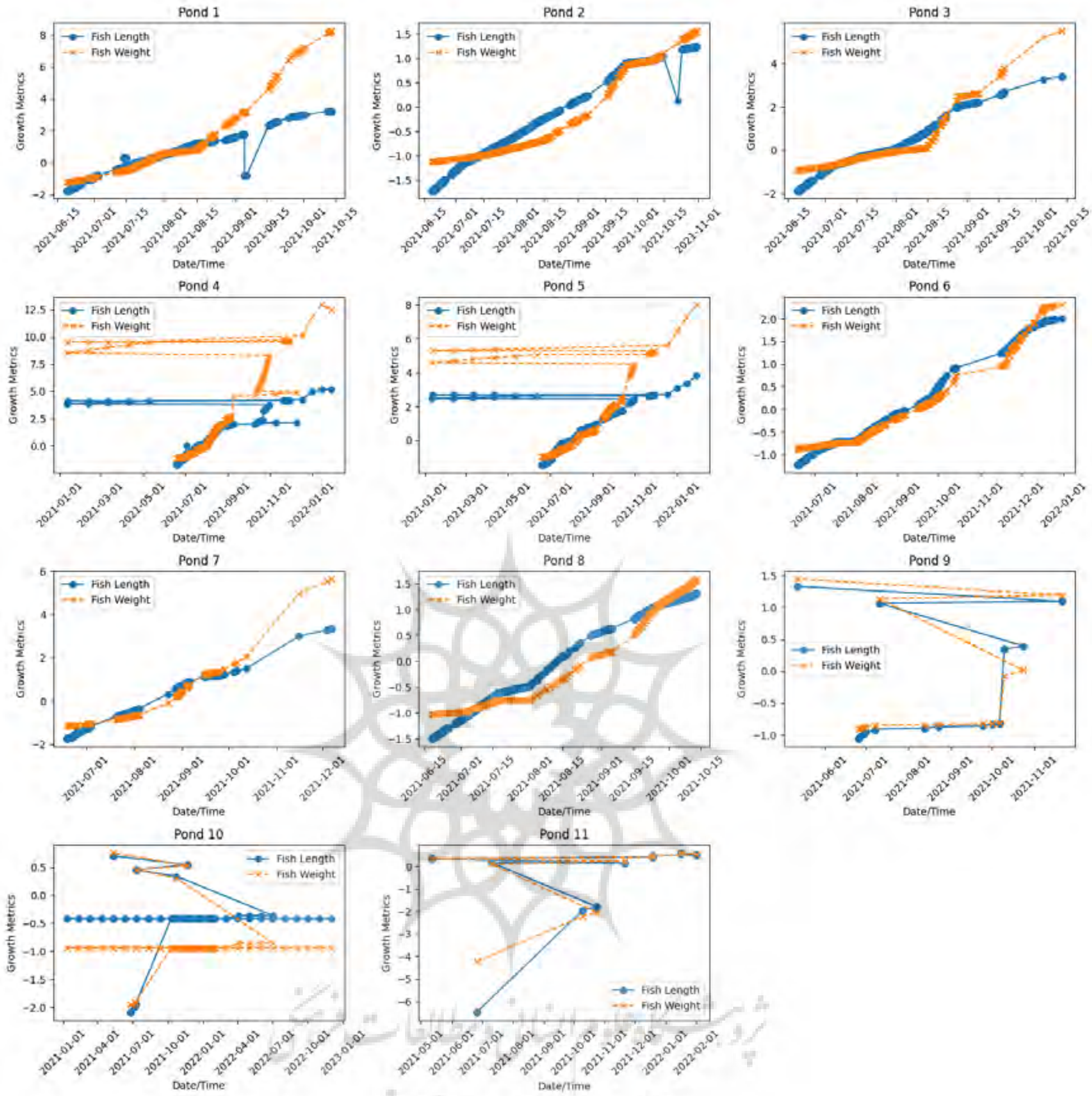

Figure. 8. Time-series trends of fish growth metrics derived from top-k sequential pattern mining, highlighting dominant behavioral trajectories while capturing overarching system responses and occasional anomalies.

Table 1. Comparative Evaluation of PrefixSpan and Top-k Sequential Pattern Mining: Correlation with Fish Growth Metrics and Pattern Counts Across Ponds.

Pond	*Correlation*		*Patterns*	
	PrefixSpan	*Top-k*	*PrefixSpan*	*Top-k*
Pond 1	0.899431	0.899431	198	5
Pond 2	0.967312	0.967312	193	5
Pond 3	0.91365	0.91365	156	5
Pond 4	0.95727	0.95727	108	5
Pond 5	0.962543	0.962543	174	5
Pond 6	0.988584	0.988584	1269	5
Pond 7	0.965693	0.965693	226	5
Pond 8	0.981434	0.981434	149	5
Pond 9	0.979886	0.979886	994	5
Pond 10	0.994046	0.994046	1339	5
Pond 11	0.975168	0.975168	591	5

4.6. Comparative Evaluation with Related Research

To provide evidence of the originality and usefulness of this approach, benchmarking was done against leading-edge anomaly detection models alongside intelligent control structures in other aquaponics systems. Table 2 reports a head-to-head comparison by highlighting method type, (where applicable) accuracy, and notable strengths or weaknesses.

In the end, the PrefixSpan-based federated method (this study) provided a good balance of accuracy, scalability, and privacy, without needing labeled datasets. When comparing the PrefixSpan method with the Top-k version, the latter was also unsupervised, however, the Top-k method is limited by having fewer patterns due to fixed patterns, while the PrefixSpan method was able to show more flexibility and more sensitivity. Alternatively, other noted methods, especially supervised (noted above), more often than not needed an aggregated infrastructure or already annotated data, which can be impossible in resource constrained or more distributed aquaponic environments

Table 2. Comparative evaluation of anomaly detection methodologies in aquaponics systems.

Method	***Approach Type***	***Accuracy***	***Highlights***
FL + PrefixSpan (This study)	Unsupervised FL + Pattern Mining	0.8964	High accuracy, scalable & privacy-preserving; no labels used
FL + Top-k (This study)	Unsupervised FL + Fixed-Pattern Mining	0.8964	Simpler representation; loses infrequent anomalies due to pattern cutoff
Smart Aquaponics (PID Control) [8]	Rule-Based PID Control	N/A	Control-centric system; lacks learning/predictive metrics
IoT + Revamped Optimization [9]	Supervised ML + Heuristics	0.9213	Centralized; optimized cost function; requires labeled data
IoT App for Aquaponics [10]	Mobile IoT + Cloud-Based Monitoring	N/A	Accurate real-time alerts; improves decision-making; user-friendly design
ML in Nutrient Management [11]	Supervised ML + Nutrient Optimization	–	Identified key nutrients; improved biomass & stability in real farms
ML + IoT + QAOA [12]	Supervised ML + Quantum Optimization	–	High R^2 (0.999); 50% faster training; no anomaly metrics reported
INAPRO System (MES + LSTM) [13]	Supervised Time-Series Forecasting	–	High recall, low F1; relies on controller alerts
Hydroponics + ML (Random Forest) [14]	Supervised Classification	0.9445	Very high performance; centralized & labeled image-rich dataset
Review of AIAS Systems [15]	Literature Synthesis + Bibliometrics	N/A	Highlights multifunctionality, integration, and resilience strategies

5. Conclusion And Future Work

This study presented a novel, privacy-preserving anomaly detection framework for aquaponics systems by integrating federated learning (FL) with sequential pattern mining. By leveraging IoT-enabled sensors deployed across multiple aquaponic ponds, the proposed system facilitates localized data analysis and model training without transmitting sensitive raw data to a central server. This decentralization inherently improves scalability, resilience, and communication efficiency. The PrefixSpan algorithm, tailored for temporal sequence analysis, effectively captured frequent patterns within environmental and biological time-series data, enabling robust detection of anomalies associated with water quality and fish health.

Through extensive experimentation across 11 distributed aquaponic nodes, the framework demonstrated its capacity to maintain high detection accuracy, reduce network latency, and provide meaningful insights into system behavior. The comparative analysis confirmed that our unsupervised, decentralized approach outperforms or complements traditional rule-based and centralized machine learning methods, particularly in scenarios where data privacy and infrastructure limitations are significant concerns. The correlation analysis between environmental parameters and fish growth metriss further substanaated the system's real-world utility for precision aquaponics. This work contributes significantly to the field of smart agriculture, where sustainable resource use and intelligent monitoring systems are increasingly essential. The model's flexibility was validated through its successful integration with both PrefixSpan and Top-k pattern mining, underscoring its extensibility for hybrid anomaly detection strategies.

Future research will aim to enhance the predictive capabilities of the model by incorporating real-time feedback mechanisms and adaptive learning strategies. The integration of reinforcement learning and edge-cloud collaborative architectures can further optimize decision-making processes in highly dynamic aquaponic environments. Additionally, expanding the scope of anomaly categories—such as predictive failures in aeration systems or nutrient delivery pipelines—will improve the system's fautt tolerance. Exploring cross-domain federated learning among heterogeneous smart farms could establish a global framework for sustainable, intelligent food production. This would pave the way for a new generation of interoperable and cooperative agricultural AI systems. Moreover, future experiments may investigate transfer learning between different aquaponics infrastructures and assess the use of blockchain-based pattern provenance to improve auditability and data integrity.

Declarations

Funding

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

Authors' contributions

SS: Conceive the original idea, designe and implemente the methodology, performe the experiments, analyze the results, generate all figures, and drafte the manuscript.

JM: supervis the project, provide critical feedback and comments throughout, and approve the final version.

Both authors have read and approved the final manuscript.

Conflict of interest

The authors declare that no conflicts of interest exist.

Acknowledgements

The authors would like to express sincere gratitude to the HiPIC Research Group, University of Nigeria Nsukka, for making the Sensor-Based Aquaponics Fish Pond Dataset publicly available, which served as a foundation for this research.

Data Availability Statement

The Sensor-Based Aquaponics Fish Pond Dataset, sourced from Kaggle [https://www.kaggle.com/datasets/ogbuokiriblessing/sensor-based-aquaponics-fish-pond-datasets], was utilized exclusively for research purposes. As the study did not involve human or animal experimentation, no institutional review board approval was required. The dataset is publicly available and not owned by the authors. The code developed for this research is original, authored by the study's author, and available at GitHub [https://github.com/SagharShafaati/Federated-Anomaly-Detection-and-Pattern-Discovery-in-IoT-Aquaponics].

References

[1] M. F. aa ha et al., “Recent advances of mmart yystems and internet of things (iot) for aquaponics automation: A comprehenii ve overview,” Chemosensors, vol. 10, no. 8, p. 303, 2022, https://doi.org/10.3390/chemosensors10080303

[2] K. Nugroho, Comparative Analysis of Federated and Centralized Learning Systems in Predicting Cellular Downlink Throughput Using CNN. in IEEE Access, vol. 13, pp. 22745-22763, 2025, https://doi.org/10.1109/ACCESS.2025.3528527

[3] O. T. Modupe et al., “Reviewing the tranfforttt ional impact of edge computing on real-time data processing and analytics,” Copp uter Science & IT Reaearch Journal, vol. 5, no. 3, pp. 693-702, 2024, https://doi.org/10.51594/csitrj.v5i3.929

[4] R. Sallenave, Important water quality parameters in aquaponics systems, Las Cruces, NM, USA: College of Agricultural, Consumer and Environmental Sciences, 2016, https://pubs.nmsu.edu/_circulars/CR680.pdf

[5] L. A. Ibrahim, H. Shaghaleh, G. M. El-Kassar, M. Abu-Hashim, .. A. Elaadek and Y. Alhaj Hamoud, “Aquaponics: A sustainable path to food sovereignty and enhanced water use efficiency,” Water, nol. 15, no. 24, p. 4310, 2023, https://doi.org/10.3390/w15244310

[6] .. M. Attafy and A. M. ll sbaay, “Integrated aa nagement of fish (Nile Tilapia) and leaf vegetable crop (head lettuce) culture under drip irrigation syste”,” Mirr Journal of Agricultural Engineering, vol. 35, no. 1, pp. 105-124, 2018, https://doi.org/10.21608/mjae.2018.96038

[7] A. Cohen, S. Malone, Z. Morris, M. Weissburg and B. Bras, “Cobb ined fish and lettuce cultivation: an aquaponics life cycle assesmment,” Procedia Cirp, vol. 69, pp. 551-556, 2018, https://doi.org/10.1016/j.procir.2017.11.029

[8] C. L. Kok, I. M. B. P. Kusuma, Y. Y. Koh, H. Tang and A. B. Lim, „Smart Aquaponics: An Autoaa ted Water Quality Management System for Sustainable Urban Agriculture,” Electronics, vol. 13, no. 5, p. 820, 2024, https://doi.org/10.3390/electronics13050820

[9] S. C. M. Sundararajan et al., “IoT-based prediction model for aquaponic fish pond water quality using multiscale feature fusion with convolutional autoencoder and GRU network”,” Scientific Reports, vol. 15, no. 1, p. 1925, 2025, https://doi.org/10.1038/s41598-024-84943-7

[10] F. .. Valiente et al., “Internet of things (IOT)-based mobile application for monitoring of automated aquaponics yyttem,” In 2018 IEEE 10th international conference on humanoid, nanotechnology, information technology, communication and control, environment and management (HNICEM), Baguio City, Philippines, 2018, pp. 1-6, https://doi.org/10.1109/HNICEM.2018.8666439

[11] S. B. Dhal et al., “A aa chine-learning-based IoT system for optimizing nutrient supply in commercial aquaponic operations,” Sensors, vol. 22, no. 9, p. 3510, 2022, https://doi.org/10.3390/s22093510

[12] R. Baena-Navarro, Y. Carriazo-Regino, F. Torres-Hoyos and J. Pinedo-óó pez, “Intelligent Prediction and Continuous Monitoring of Water Quality in Aquaculture: Integration of Machine Learning and Internet of Things for Sustainable Management,” Water, vol. 17, no. 1, p. 82, 2025, https://doi.org/10.3390/w17010082

[13] D. Karimanzira and T. Raucchenbach, “An intelligent management system for aquaponics," at-Automatisierungstechnik, vol. 69, no. 4, pp. 345-350, 2021, https://doi.org/10.1515/auto-2020-0036

[14] A. Métwalli et al., “Enhancing Hydroponic Farii ng with Machine Learning: Growth Prediction and Anomaly Detection,” Engineering Applications of Artificial Intelligence, vol. 157, p. 111214, https://doi.org/10.1016/j.engappai.2025.111214

[15] L. A. Ibrahim, M. Abu-Hashim, H. Shaghaleh, E. Elsadek, A. A. A. Haaa d and Y. Alhaj Hamoud, “A comprehenii ve review of the multiple uses of water in aquaculture-integrated agriculture based on international and national experiences,” Water, vol. 15, no. 2, p. 367, 2023, https://doi.org/10.3390/w15020367

[16] S. Shafaati and .. Mohammadzadeh, “Scalable Anoaa ly Detection and Pattern Mining in Iot Aquaponics Systems

Via Federated Learning," In 2025 11th International Conference on Web Research (ICWR), Tehran, Iran, Islamic Republic of, 2025, pp. 39-43, https://doi.org/10.1109/ICWR 65219. 2025.11006250

[17] A. Ghandar, A. Ahmed, S. Zulfiqar, Z. Hua, M. Hanai and G. Theodoropoulos, "A deciiion support system for urban agriculture using digital twin: A case study with aquaponics," Ieee Access, vol. 9, pp. 35691-35708, 2021, https://doi.org/10.1109/ACCESS.2021.3061722

[18] L. Nsoanya, "Response of Fish Pond ff fluent on Soil Chemical Properties and Growth of Cucumber (Cucumis ttt ivus) in Igbariam South aa stern, Nigeria," Nigeria (October 3, 2019), 2019, http://dx.doi.org/10.2139/ssrn.3463863

[19] https://www.kaggle.com/datasets/ogbuokiriblessing/sensor-based-aquaponics-fish-pond-datasets

Saghar Shafaati received her BS and M.Sc. in Computer Engineering from Islamic Azad University, South Tehran Branch, Iran, in 2022 and 2024 respectively. She is currently a volunteer researcher in Computer Engineering at Islamic Azad University, South Tehran Branch, Iran. Her research interests are in the areas of Machine Learning, the Internet of Things, and Anomaly detection.

Javad Mohammadzadeh received his BS in Computer Science from Shahid Bahonar University of Kerman, Kerman, Iran, in 2004. He received his M.Sc. in Computer Science from University of Tehran, Tehran, Iran, in 2007. And he received his Ph.D. in Bioinformatics from University of Tehran, Tehran, Iran, in 2014. He is an associate professor, Department of Computer Engineering, Ka.C., Islamic Azad University, Karaj, Iran & Institue of Artificial Intelligence and Social and Advanced Technologies, Ka.C., Islamic Azad Unniversity, Karaj, Iran. He is the author/co-author of more than 10 publications in technical journals and conferences. His research interests are in the areas of swarm intelligence algorithms bioinformatics algorithms, and parallel computing.