

Article

Early Anomaly Detection in Shrimp Pond Water Quality Using Supervised and Unsupervised Machine Learning Models

Hamilton Villamar-Barros ^{1,*}, Julián Coronel-Reyes ¹ and Alexander Haro-Sarango ²

¹ Facultad de Ciencias Agrarias, Universidad Agraria del Ecuador, Guayaquil 090107, Ecuador; jjcoronel@uagraria.edu.ec

² Instituto Superior Tecnológico España, Ambato 180101, Ecuador; alexander.haro@iste.edu.ec

* Correspondence: hvillamar@uagraria.edu.ec

Abstract

Shrimp aquaculture increasingly depends on precise water quality management, yet most farms still rely on fragmented measurements and qualitative assessments. This study aimed to evaluate whether routine physicochemical data from commercial ponds can reliably discriminate between operational categories of acceptable and residual water and thus support early warning systems. We compiled water quality records from shrimp ponds in several coastal provinces, focusing on a reduced set of variables related to salinity, alkalinity, hardness and inorganic nitrogen. Supervised and unsupervised machine learning models were trained and compared using standard classification metrics. Tree-based ensembles and margin-based models achieved high accuracy and F1 scores when predicting water status from routine variables, while clustering methods only reproduced similar patterns after an ex post mapping of clusters to classes. These results indicate that latent nitrogen loads and subtle shifts in water chemistry are systematically captured by basic monitoring data and can be translated into operational signals of risk. The study demonstrates the feasibility of integrating data-driven classification into shrimp farm monitoring and outlines a pathway toward low-cost, scalable decision support tools for aquaculture 4.0 in data-limited settings.

Keywords: shrimp aquaculture; water quality; machine learning; supervised learning; unsupervised learning; early warning systems; aquaculture 4.0

1. Introduction

Shrimp aquaculture is a strategic sector for food security and exports in many tropical countries, but its productive performance critically depends on water quality. Small variations in physicochemical parameters can trigger disease outbreaks, growth reductions, or mass mortalities, whereas traditional monitoring schemes based on periodic sampling and manual interpretation are often slow and reactive. This has driven interest in data-based early warning systems capable of anticipating anomalies and supporting timely management decisions in production contexts with limited resources.

Within this framework, machine learning has become a powerful tool for classifying and predicting water quality in diverse ecosystems, mainly through supervised models (e.g., Random Forests, Gradient Boosting, SVM) and, to a lesser extent, unsupervised approaches to clustering and anomaly detection. However, few studies systematically compare both paradigms on the same aquaculture dataset, and those specifically focused on shrimp ponds are almost non-existent. In these systems, variables follow continuous gradients and operational quality classes are defined by productive thresholds rather than



Academic Editors: Jorge Bernardino, Le Gruenwald, Elio Masciari and Laurent D'Orazio

Received: 9 December 2025

Revised: 6 January 2026

Accepted: 14 January 2026

Published: 1 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

by well-separated natural clusters. This particularity raises questions about the ability of unsupervised methods to recover categories that are genuinely useful for management, as well as about the conditions under which supervised models provide a clear advantage.

This study addresses these gaps through the analysis of 2136 water samples collected from more than 90 commercial shrimp ponds across five coastal provinces of Ecuador (2023–2025), with seven key physicochemical indicators and an operational quality label (acceptable vs. residual). On this basis, a rigorous preprocessing pipeline is implemented and supervised models (multinomial logistic regression, SVM with RBF kernel, Random Forests, and Gradient Boosting) are compared with four unsupervised clustering algorithms (K-Means, Gaussian Mixtures, Agglomerative Clustering, and Spectral Clustering). The central objective is twofold: (i) to assess the extent to which unsupervised clustering can approximate operational water quality classes in shrimp ponds and (ii) to determine under what conditions supervised models provide a decisive advantage for early anomaly detection and risk management.

The study makes three main contributions: first, a comparative and replicable benchmark between supervised and unsupervised algorithms applied to a real-world dataset from the shrimp industry; second, an integrated analysis of external classification metrics and internal clustering indices under a threshold-based labeling scheme; and third, the proposal of a hybrid framework in which supervised ensembles are combined with unsupervised clustering for pattern exploration and anomaly pre-filtering, oriented toward the design of intelligent monitoring systems in contexts with partially labeled data.

2. Theoretical Review

2.1. Importance of Water Quality in Shrimp Aquaculture

The intensification of shrimp aquaculture has made it clear that growth, feed conversion, and survival depend as much on management decisions as on the stability of water quality in increasingly closed and recirculating systems. In high-density scenarios, small changes in dissolved oxygen, nitrogen compounds, or suspended solids translate into significant variations in productive performance, making it necessary to design hydraulic configurations that buffer peaks in organic load and maintain physical and chemical gradients within optimal ranges throughout the production cycle [1]. The way feed is dosed, especially in blind feeding schemes, also reshapes the balance between growth and water stability, since uneaten feed rapidly increases biochemical oxygen demand and the accumulation of ammonium and nitrite, whereas more refined feeding strategies reduce nutrient losses to the environment and smooth the dynamics of dissolved metabolites [2]. Added to this is the architecture of recirculating systems with advanced biofilters, where parameters such as hydraulic retention time and mineral supplementation simultaneously condition nitrification efficiency, animal welfare, and system resilience to loading disturbances [3]. Even in technologically mature designs, the arrangement of filtration modules and substrates in the tanks modifies the stratification of solids and microorganisms, modulating the microenvironment around the shrimp and thereby influencing their growth potential and production uniformity [4].

In parallel, there is an ongoing transition from extensive and semi-intensive schemes toward intensive systems in covered or indoor facilities, where pressure on water quality increases due to the combination of high biomass and reduced water exchange. In these environments, the adoption of indoor biofloc systems has shown that microbial aggregates function as a biological filter, nutrient reservoir, and sanitary barrier by improving clarification, stabilizing nitrogen concentrations, and providing microbial protein, while simultaneously increasing disease resistance in high-value shrimp species [5]. The incorporation of continuous environmental monitoring schemes in semi-intensive ponds has

made it possible to link more explicitly the dynamics of variables such as temperature, oxygen, pH, and transparency with feed conversion efficiency and growth rates, demonstrating that fine control of these parameters translates into tangible improvements in productive performance [6]. Among recent innovations, ozone nanobubble technologies stand out, as they enhance oxygen transfer, reduce bacterial loads, and remodel shrimp gut microbiota, generating more stable culture environments with improved growth rates in intensive indoor systems [7]. Complementing this, the strategic use of probiotics has emerged as a tool to articulate system carrying capacity, water quality, and productivity by reinforcing beneficial microbial communities that compete with pathogens and facilitate the assimilation of organic residues in grow-out ponds [8].

Beyond technological components, the design of feeding strategies and carbon inputs shapes the trophic structure of the water column and, consequently, the degree of internal nutrient recycling. The implementation of microbubble generators is an example of how aeration can cease to be merely a source of oxygen and become a mechanism for integrally improving water quality by increasing gas transfer efficiency, promoting the suspension of solids, and fostering more homogeneous conditions around the cultured organisms [9]. In biofloc systems, the selection of carbon sources such as wheat flour makes it possible to adjust the carbon-to-nitrogen ratio of the water column, stimulating the growth of heterotrophic bacteria that capture inorganic nitrogen and convert it into microbial biomass available to juvenile shrimp [10]. The incorporation of organic inputs from adjacent ecosystems, such as mangrove leaf litter, has shown that plant detritus not only provides carbon and habitat for microorganisms, but also influences transparency, color, and biofilm structure, with direct effects on postlarval survival in transition scenarios between natural environments and culture systems [11]. Similarly, artificial salinization strategies in low-salinity systems show that the joint management of ionicity and water quality in synergistic nursery systems can optimize shrimp growth and physiological robustness, while stabilizing the functioning of microbiota associated with biofloc under freshwater or intermediate salinity conditions [12].

2.2. Technological Strategies for Water Quality Management

Understanding water quality in shrimp ponds is increasingly linked to the microbial ecology of recirculating systems, where bacterial communities associated with the water and biofilters act as mediators of biogeochemical cycles. It has been documented that bacterial composition in both culture water and bioflocs is closely related to nutrient, solid, and oxygen gradients in recirculating systems, implying that manipulating these microbial consortia can be as important as adjusting classical physicochemical variables [13]. This perspective is complemented by studies showing how different aeration intensities in biofloc systems modify the distribution of suspended sediments, the rate of oxidation of nitrogen compounds, and the system's capacity to avoid ammonium peaks, which translates into differences in shrimp growth and survival [14]. The use of probiotics in pelleted diets opens another avenue for intervention, insofar as probiotic formulations integrated into feed not only modulate gut microbiota but also reduce particulate organic loads and improve water clarity in static biofloc systems, reinforcing the connection between nutrition and environmental quality [8]. On this basis, the management of pond bottoms and sediment layers, traditionally considered a secondary aspect, is now recognized as a central component for preventing the accumulation of organic matter, hydrogen sulfide, and other toxic substances, establishing systematic soil management practices as part of water quality control in intensive ponds [15].

The development of biofloc systems has been accompanied by a diversification of strategies to enhance their benefits, where the application of specific probiotics to the flocs

has been linked to simultaneous improvements in growth, feed conversion, and stability of parameters such as ammonium, nitrite, and suspended solids in intensive cultures of white shrimp [16]. In parallel, more energy-efficient aeration solutions equipped with monitoring capabilities have emerged, such as solar aerators integrating water quality sensors that maintain adequate dissolved oxygen levels while continuously recording fluctuations in temperature, pH, and turbidity in vannamei ponds [17].

Commercial probiotic formulations used in biofloc systems also show remarkable plasticity to adapt to different regimes of organic loading and temperature, maintaining positive effects on growth and feed conversion even when water conditions change rapidly in intensive systems [18]. This approach has also been extended to early stages of the production cycle, where probiotic mixtures applied during hatching and salinity-shock phases have proven effective in protecting high-value species such as kuruma shrimp against osmotic stress and abrupt environmental fluctuations, reinforcing the idea that water quality and animal health are inseparable dimensions from the rearing stage onward [19].

More recently, combinations of probiotics, alternative carbon sources, and physical controls such as light have been explored to fine-tune biofloc system stability and their capacity to support high biomass loads. Probiotic supplementation in biofloc systems has been shown to simultaneously increase zootechnical efficiency and immunological robustness in white shrimp, reducing the dependence on water exchange and enhancing the system's ability to recycle nitrogen within the production loop [20]. The use of agro-industrial by-products such as sugarcane bagasse as a carbon source has demonstrated that it is possible to sustain active microbial communities that improve water quality and diversify biofloc structure, while fostering more sustainable use of waste and reducing the environmental footprint of culture operations [21]. Manipulating light intensity constitutes another management front, as controlled light reduction in intensive biofloc systems alters primary productivity, microbial community composition, and oxygen dynamics, thereby modulating both water quality and shrimp feeding behavior [22]. Complementarily, the development of automatic paddlewheel aerators designed to improve mixing and water column monitoring in marine farms has highlighted that aeration hardware design can integrate sensing and control functions, contributing to stabilizing water quality gradients in high-density systems [23].

2.3. Digitalization, Machine Learning, and Early Warning in Water Quality Control

Advances in monitoring technologies have made it possible to move from point-based diagnostics relying on sporadic sampling to comprehensive assessments of water quality that combine composite indicators and mathematical models. In coastal regions with intense shrimp farming activity, the Water Quality Index (WQI) has been shown to be useful for synthesizing multiple physical and chemical parameters, while allowing the environmental impact of farms on adjacent ecosystems to be assessed and mitigation measures to be guided in coastal zones [24]. On the basis of such metrics, deterministic models are built that relate water quality dynamics to controllable input variables such as aeration, feeding, or water retention time, providing tools to simulate management scenarios in closed-system farms and to estimate the behavior of critical parameters under alternative operating strategies [25]. The combination of quality indices and time series of in situ measurements also allows spatial mapping of pollution and water quality variability in intensive vannamei farming areas, offering a foundation for decisions on the location of new production units and for the design of environmental surveillance plans at regional scales [26].

In parallel, the integration of sensor networks connected to Internet of Things (IoT) platforms has given rise to real-time water quality monitoring systems in grow-out ponds,

capable of continuously transmitting information on oxygen, temperature, and pH and triggering alarms when values approach risk thresholds for shrimp [27].

The convergence between distributed sensing, cloud computing, and data analysis techniques has driven the development of more sophisticated tools for interpreting water quality and detecting anomalies at an early stage. In intensive systems, the combined use of principal component analysis and fuzzy approaches has made it possible to identify the key factors structuring water quality, group system operating states, and distinguish between normal conditions and potentially critical situations, providing a statistical basis for decision-support systems in shrimp farms [28]. Building on this analytical infrastructure, specific IoT architectures are implemented in which sensing nodes distributed across ponds communicate with central gateways that process data in real time, enable the visualization of trends, and facilitate more timely interventions, such as automatic activation of aeration equipment or adjustment of feeding rates according to oxygen demand [25]. Integration with cloud computing services has led to intelligent platforms for predicting water quality in shrimp ponds, which provide “as-a-service” functionality to end users and are capable of estimating the future behavior of critical parameters based on historical patterns and current environmental conditions [29]. This ecosystem is articulated within a broader vision of the “IoT revolution” in aquaculture, where automated feeding, continuous water quality monitoring, and advanced analytics converge in a single system, shifting management from an experience-based reactive approach to a preventive, data-driven one [30].

At the current frontier, monitoring systems incorporate low-power, high-connectivity microcontrollers that facilitate the implementation of low-cost sensor networks in resource-limited contexts. The development of platforms based on controllers such as Mappi32 has shown that it is possible to deploy continuous water quality monitoring systems in shrimp ponds, integrating multiple sensors and wirelessly transmitting data to remote servers or mobile applications—something especially relevant for small and medium-scale producers [31]. Specific predictive models are built upon these data streams, for example, for pH evolution, enabling the anticipation of changes in acidity and preventive decision-making regarding aeration, alkalinity dosing, or biofloc management, thereby reinforcing the capacity of culture systems to absorb disturbances and maintain stable conditions for shrimp [32]. Likewise, water quality assessment tools in early life stages, such as hatcheries, have highlighted that production sustainability depends on the ability to maintain suitable physicochemical conditions from the larval stage, since failures in hatchery water quality compromise the entire production chain and increase reliance on downstream corrective measures [33]. Finally, monitoring systems integrated with early warning algorithms make it possible to transform information into concrete actions by identifying patterns of water quality deterioration that precede performance declines or mortality events and communicating these risks in a timely manner to farm personnel, thereby completing the transition toward genuinely preventive water quality management in shrimp aquaculture [34].

3. Materials and Methods

This study adopted a hybrid approach combining supervised and unsupervised machine learning for the early detection of anomalies in the water quality of shrimp ponds. The methodological design was based on a systematic review of relevant scientific literature in aquaculture and Machine Learning, whose findings were integrated into a conceptual framework represented in Figure 1.

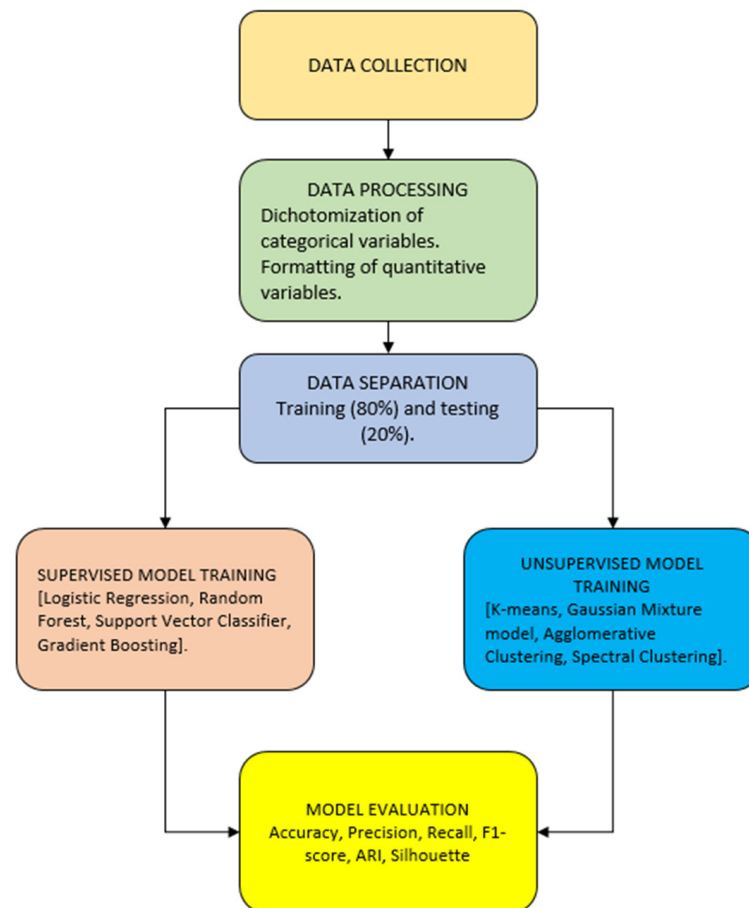


Figure 1. Diagram of stages and steps in the experimental methodology.

For the analysis of water quality in shrimp ponds, detailed information was collected from more than 90 ponds located in the provinces of Guayas, El Oro, Manabí, Esmeraldas, and Santa Elena. The study covered the period between January 2023 and January 2025, focusing on ponds managed by individuals with more than a decade of experience in shrimp farming. After collecting water samples and analyzing them in the laboratory, seven key indicators were identified to assess water quality: pH, alkalinity, ammonia, nitrite, calcium, magnesium, and salinity. These parameters are fundamental for determining the suitability of water for shrimp rearing. The dataset consists of 2136 records, which constitute a solid basis for distinguishing between good-quality water and water with residual characteristics, an essential aspect for the efficient and sustainable management of aquaculture systems.

The methodological process developed in this study aimed to compare the performance of supervised and unsupervised learning algorithms for classifying water quality based on physicochemical parameters. The dataset included variables commonly used in water monitoring studies, such as pH, alkalinity, ammonia concentration, nitrates, calcium, magnesium, and salinity levels. Before proceeding with modeling, a cleaning and preprocessing phase was carried out in which variables that did not provide informative content, such as unique identifiers or codes associated with the ponds, were removed, since their presence could introduce noise or promote overfitting. This decision is grounded in recent advances that recommend excluding attributes without a functional relationship to the biogeochemical processes under evaluation, in order to improve the predictive capacity and interpretability of the model [35]. Likewise, missing and outlier values were inspected with the purpose of ensuring the internal consistency of the dataset. For missing values, median-based imputation was used, a criterion that has proven more robust in contexts

where environmental variables may exhibit skewed distributions or the presence of extreme values [21]. Once imputation was completed, all numerical variables were standardized using a robust scaling method, which minimized the impact of outliers and ensured that scale-sensitive algorithms operated under comparable conditions.

To objectively evaluate the generalization capacity of the models, the dataset was divided into two independent partitions following a proportion of 80% for training and 20% for testing. The split was performed using stratified sampling, preserving the relative distribution of water quality categories in both subsets. This methodological strategy follows widely accepted recommendations in the machine learning literature, which emphasize that a sufficiently large training proportion allows the underlying structure of the phenomenon to be adequately captured, while the percentage reserved for testing guarantees an unbiased assessment of predictive performance [36,37]. Stratification also reduces the risk of biases arising from underrepresented classes and ensures that the final evaluation of the algorithms reflects realistic prediction conditions.

For the supervised analysis, four algorithms widely documented in recent research on environmental classification and water quality were selected: multinomial logistic regression, Random Forest, support vector machines with RBF kernel, and Gradient Boosting. Multinomial logistic regression was used as the baseline model, as it is an interpretable linear technique that serves as a reference point for assessing improvements derived from more complex methods [37,38]. Random Forest and Gradient Boosting represent ensemble approaches based on multiple trees, recognized for their ability to capture nonlinear relationships, interactions among variables, and complex patterns in environmental data, while maintaining high stability in the presence of noise and natural variability [39–41]. In turn, the SVM model with RBF kernel was included due to its robustness in modeling curved and high-dimensional decision boundaries, a particularly useful feature in ecological and environmental systems where relationships among variables are often non-linear [42–44]. Each of these algorithms was trained exclusively on the 80% training set, applying five-fold stratified cross-validation to ensure stable fitting and avoid overfitting. Subsequently, the final fitted model was evaluated on the 20% test set using accuracy, macro-precision, macro-recall, and macro-F1 metrics, which allow for a balanced analysis of performance even in the presence of imbalanced classes.

Complementarily, an unsupervised analysis was included to explore the intrinsic structure of the data and determine the extent to which spontaneous groupings coincide with the actual water quality categories. To this end, four clustering algorithms were used: K-means, Gaussian Mixture Models, agglomerative hierarchical clustering, and DBSCAN. The choice of these methods responds to their different structural assumptions and their widespread use in studies of ecological characterization and environmental pattern detection [45,46]. In the case of K-means, GMM, and the hierarchical model, the number of clusters was set according to the actual number of water quality categories, as suggested in the literature when the aim is to assess the alignment between generated partitions and known classifications [47,48]. By contrast, DBSCAN made it possible to identify groups based on local densities, facilitating the detection of outliers and non-linear structures without the need to pre-specify the number of clusters [49].

Once the clusterings were obtained, their correspondence with the actual categories was evaluated using the Adjusted Rand Index (ARI) and the Normalized Mutual Information (NMI), metrics widely accepted for validating coherence between clusters and true classes in scenarios where unsupervised algorithms are used for comparison against predictive models [47,48]. Additionally, each cluster was assigned the majority category present in the training set, thereby generating a “cluster-derived classification” mechanism

that allowed the computation of measures comparable to those used in the supervised analysis, such as accuracy and macro-F1. This methodological strategy made it possible to quantify how close unsupervised algorithms can come to the performance of models trained with explicit information on water quality.

Taken together, the adopted methodology enables a rigorous contrast between the predictive potential of supervised algorithms and the structural capacity of unsupervised methods to identify patterns in water quality. This dual approach is particularly relevant in environmental studies, where the availability of labels may be limited and where the latent structure of the data can provide complementary information to traditional monitoring systems [21,50]. To ensure the transparency and reproducibility of this study, a GitHub repository (version 3.5.4) has been provided containing the code used for data processing, preprocessing, supervised/unsupervised modeling, and analysis of the 2.136 shrimp pond water quality records, along with a sample of synthetic data that reproduces the statistical characteristics of the original dataset. These materials are available in the Supplementary Materials. All numerical data employed in this work are described in detail in the manuscript and were used exclusively for the analyses presented.

4. Results

The results show in Table 1 that the aquatic system under evaluation exhibits relatively stable physicochemical conditions, with a predominantly alkaline pH (mean 7.76) and high alkalinity (mean 139.82 mg/L), which indicates adequate buffering capacity that contributes to the chemical stability of the water. This combination suggests an environment with a low risk of abrupt acidity fluctuations, which is favorable both for the aquatic biota and for production processes that depend on water quality.

Table 1. Descriptive statistics for the variables are summarized below.

| Statistic | pH | Alkalinity | Ammonia | Nitrate | Calcium | Magnesium | Salinity | Water Quality |
|-----------|---------|------------|---------|---------|---------|-----------|----------|---------------|
| count | 2136.00 | 2136.00 | 2136.00 | 2136.00 | 2136.00 | 2136.00 | 2136.00 | 2136.00 |
| mean | 7.76 | 139.82 | 1.06 | 15.46 | 300.42 | 149.76 | 22.69 | 0.21 |
| std | 0.72 | 33.65 | 0.55 | 8.39 | 56.91 | 28.81 | 7.24 | 0.40 |
| min | 6.50 | 80.02 | 0.10 | 1.00 | 200.05 | 100.05 | 10.01 | 0.00 |
| 25% | 7.14 | 111.56 | 0.57 | 7.99 | 250.55 | 124.78 | 16.36 | 0.00 |
| 50% | 7.79 | 139.39 | 1.05 | 15.79 | 302.26 | 149.85 | 22.77 | 0.00 |
| 75% | 8.38 | 169.04 | 1.55 | 22.63 | 347.96 | 175.36 | 29.08 | 0.00 |
| max | 9.00 | 199.99 | 2.00 | 29.99 | 399.97 | 199.92 | 35.00 | 1.00 |

Nitrogen compounds exhibit differentiated patterns: while ammonia maintains moderate average values (1.06 mg/L), with some records close to the ecotoxic threshold, nitrate shows marked variability, reaching levels of up to 30 mg/L. This behavior suggests a system in which nitrification is active and where there may be zones with high organic loads or a constant input of nutrients, leading to nitrate accumulation. The simultaneous presence of ammonia and nitrates within these ranges indicates an environment where dynamic processes of decomposition, oxidation, and potential eutrophication are taking place.

Calcium and magnesium levels reflect high and consistent total hardness, which characterizes the system as mineralized and with a chemical composition dominated by dissolved salts. This profile is consistent with the observed salinity values, which on average reach 22.69 PSU and fall within the typical ranges of brackish environments. The relative stability of these parameters suggests a system governed more by geological conditions or coastal exchange than by short-term, localized variations.

With respect to the dependent variable, water quality shows a marked imbalance, with most observations classified in category 0. This pattern has direct implications for

classification models, as it may induce biases toward the majority class if appropriate balancing or evaluation techniques are not applied. This imbalance does not invalidate the analysis, but it constrains the interpretation of the performance of supervised models and underscores the relevance of using metrics such as macro F1 and per-class recall.

Overall, the results reflect a chemically stable ecosystem with clear indicators of nitrogen pressure, and a robust dataset that is nonetheless imbalanced with respect to the target variable. These characteristics justify the use of combined supervised and unsupervised learning approaches to assess both intrinsic patterns and the predictive capacity of the models for water quality classification.

The results show in Table 2 consistent performance across the different supervised models for water quality classification, with accuracy values ranging between 0.794 and 0.799. Logistic regression and the SVM model with RBF kernel exhibit identical accuracy (0.794) and recall (0.794) values, which indicates that both algorithms correctly identify the overall proportion of cases. However, their precision is lower (0.631), revealing difficulties in adequately distinguishing the minority class. Their F1 score of 0.703 confirms a moderate balance between precision and sensitivity, characteristic of linear models or models with smooth decision boundaries in contexts with class imbalance.

Table 2. Evaluation metrics for classification using supervised learning models.

| Model | Accuracy | Precision | Recall | F1-Score |
|--------------|----------|-----------|--------|----------|
| LogReg | 0.7944 | 0.6311 | 0.7944 | 0.7034 |
| RandomForest | 0.7991 | 0.7893 | 0.7991 | 0.7184 |
| SVM_RBF | 0.7944 | 0.6311 | 0.7944 | 0.7034 |
| GradBoost | 0.7967 | 0.7534 | 0.7967 | 0.7209 |

The Random Forest model presents the highest accuracy (0.799), together with a considerably higher precision (0.789), which shows that the ensemble based on multiple trees offers a better ability to discriminate between classes, even in the presence of heterogeneity in the physicochemical data. Its F1 score of 0.718 indicates a better balance between type I and type II errors compared to the linear models and SVM.

For its part, Gradient Boosting exhibits the highest F1 score among all models (0.720), along with an accuracy of 0.796, only slightly lower than that of Random Forest, which demonstrates that this sequential model is particularly effective at capturing complex patterns and iteratively correcting errors. Its behavior suggests a greater capacity to correctly identify the minority class without sacrificing overall performance.

Overall, these results reveal that the ensemble models Random Forest and Gradient Boosting slightly outperform the linear models and SVM in terms of precision and F1, showing better adaptation to the complexity of the system and to the inherent imbalance of the data. Nonetheless, the closeness of the values indicates that the structure of the dataset allows for a relatively clear separation between classes, regardless of the algorithm used.

Table 3 presents the results obtained for the four unsupervised learning algorithms (K-Means, Gaussian Mixture Models, Agglomerative Clustering, and Spectral Clustering) show an almost identical behavior in terms of external metrics when compared with the true labels. The values of accuracy, precision, recall, and weighted F1 are exactly the same across the four models (accuracy = 0.7935, F1 = 0.7022), indicating that, after the cluster-to-majority-class mapping process, all methods approximated the original water quality classification in a very similar way. This result suggests that the data structure exhibits sufficiently defined groupings for different algorithms to converge toward closely related partitions.

The Adjusted Rand Index (ARI) provides additional information about the correspondence between the generated partitions and the true classification without relying on the mapping. The values obtained are close to zero for all models, ranging between -0.00046 and 0.00351 . Such low values indicate that the similarity between the generated clusters and the true classes is minimal when evaluated directly, without reclustering. This implies that the clusters do not explicitly reproduce the true classes, even though, after the mapping, they do manage to approximate them in predictive terms. This phenomenon is typical in systems where the true classes do not perfectly correspond to natural geometric groupings, but rather to distributions that can be recovered through a subsequent assignment step.

The Silhouette Score reinforces this interpretation. The values obtained (between 0.0718 and 0.1075) are very low, indicating weak internal cohesion of the clusters and poor separation between them. This shows that, although the algorithms achieve an approximate classification after mapping to the true classes, the intrinsic structure of the data does not display well-defined clusters from a geometric standpoint. In particular, K-Means and GMM yield the highest Silhouette values within the set (0.106 – 0.108), whereas Agglomerative Clustering obtains the lowest value (0.0718), suggesting that the latter forms more overlapping or less compact clusters.

The detailed results are summarized below:

Table 3. Evaluation metrics for classification using unsupervised learning models.

| Model | Accuracy vs. Real | Precision vs. Real | Recall vs. Real | F1 vs. Real | ARI | Silhouette |
|---------------|-------------------|--------------------|-----------------|-------------|-----------|------------|
| KMeans | 0.7935 | 0.6297 | 0.7935 | 0.7022 | -0.0002 | 0.1076 |
| GMM | 0.7935 | 0.6297 | 0.7935 | 0.7022 | -0.0005 | 0.1059 |
| Agglomerative | 0.7935 | 0.6297 | 0.7935 | 0.7022 | 0.0006 | 0.0718 |
| Spectral | 0.7935 | 0.6297 | 0.7935 | 0.7022 | 0.0035 | 0.0929 |

The comparison between supervised and unsupervised algorithms reveals important differences in both approaches' ability to approximate the true water quality classification, as well as in how each type of model captures the latent structure of the data. In the supervised models, accuracy values ranged from 0.794 to 0.799 , while F1-scores were between 0.703 and 0.721 . The ensemble models Random Forest and Gradient Boosting showed slightly superior performance in terms of balancing precision and recall, which suggests that these methods are able to capture non-linear relationships and complex patterns derived from the interaction among physicochemical variables, even in a setting where the minority class is less frequent.

By contrast, the unsupervised models, after mapping clusters to the true classes, achieved values almost identical to those of the supervised models in terms of accuracy (0.7935 in all cases) and weighted F1 (0.7022). This finding is particularly relevant, as it indicates that, even without using water quality labels during training, the clustering algorithms were reasonably able to approximate the classificatory pattern observed in the supervised models. Such behavior suggests the presence of a latent structure in the data which, although it does not form geometrically strong clusters, does contain sufficient information for the unsupervised partitions to partially reflect the true classification once a post hoc mapping is applied.

However, when intrinsic clustering metrics such as the Silhouette Score and the Adjusted Rand Index (ARI) are examined, clear contrasts with supervised performance become evident. The very low Silhouette and ARI values indicate that the intrinsic geometry of the data does not form well-separated clusters. This result is expected in systems where water quality categories are defined by operational thresholds rather than by natural discontinuities in the physicochemical space. Variables such as ammonia and nitrate follow

continuous gradients, and the transition from acceptable to residual conditions occurs gradually rather than through abrupt structural breaks. As a consequence, unsupervised algorithms that rely solely on distance or density fail to identify compact clusters aligned with management categories. Their apparent classification performance only emerges after an ex post mapping to known labels, reinforcing the idea that supervised learning is more appropriate when the goal is operational decision support.

The comparison shows that supervised models exhibit superior performance in terms of effective class discrimination and predictive stability, especially the ensemble methods. Although unsupervised methods are unable to produce clusters that are naturally aligned with the labels, they nonetheless achieve a surprisingly similar performance to the supervised models once a majority-class mapping procedure is applied. This result indicates that water quality is underpinned by physicochemical patterns which, although they do not form strict clusters, do possess a sufficiently coherent underlying structure to be partially captured through clustering. Nonetheless, the low internal cohesion evidenced by structural metrics confirms that supervised classification remains the most suitable approach when reliable labels are available and the goal is to achieve the highest possible predictive accuracy.

5. Discussion

Based on the results obtained, the shrimp ponds operate in a relatively stable physicochemical environment, with alkaline pH, intermediate salinities and high levels of alkalinity and hardness. This configuration is consistent with the description [1], who show that in intensive systems small variations in dissolved oxygen or nitrogenous compounds translate into marked changes in growth and survival. Similarly, Refs. [2,51] point out that in high-density cultures the system may appear broadly stable, but remains under a hidden nitrogen load that, if not controlled, ends in water quality crises. In this context, our findings confirm that the apparent “normality” of most water quality readings does not imply absence of risk, but rather operation close to critical thresholds defined by ammonium and nitrate concentrations.

When alkalinity and hardness are examined in more detail, the results approach what [52,53] report for BioRAS and biofloc systems, where mineral support is crucial to sustain nitrification and buffer abrupt changes in nitrogenous compounds. In our case, the combination of chemically buffered water with clear signs of nitrogen accumulation reproduces the scenario described by [21] for biofloc systems operating under high organic loads: the system is maintained, but only through a delicate balance between carbon supply, microbial activity and aeration capacity. Refs. [10,54] stress that, under such conditions, managing the carbon-to-nitrogen ratio and the ionic composition ceases to be a minor technical detail and becomes a prerequisite for the continued viability of the culture. Our data are consistent with this interpretation, since the model systematically detects that the transition between acceptable and residual water is associated with increases in dissolved inorganic nitrogen within a relatively narrow range.

The patterns observed in moderate ammonium and variable nitrate concentrations resonate with the studies of [55,56], who show that the decomposition of uneaten feed and feces, combined with intensive feeding strategies, pushes pond carrying capacity to its limits. Refs. [57,58] add that even when aeration is abundant, the accumulation of solids on the pond bottom and the formation of anoxic zones act as “time bombs” that episodically release toxic compounds. In practical terms, the fact that our model distinguishes with good accuracy between acceptable and residual water based on routine parameters suggests that this accumulation process leaves a quantifiable footprint that can be exploited to design

early warning systems, precisely in line with the proposals of [59,60] when they discuss nanobubble and ozone technologies as mitigation tools.

Recent advances in water quality management highlight the role of microbial processes, aeration, and feeding strategies in shaping nitrogen dynamics in shrimp ponds. Although these biological and technological components are not explicitly modeled in this study, their effects are indirectly reflected in routine physicochemical measurements such as ammonia, nitrite, alkalinity, and salinity. These variables integrate the cumulative outcome of microbial activity, organic loading, and hydraulic conditions, making them suitable proxies for data-driven anomaly detection. Consequently, the present work focuses on how such routinely collected indicators can be leveraged by machine learning models to support early warning systems, rather than on the detailed biological mechanisms underlying water quality regulation. Refs. [14,61] show that the structure of bacterial consortia and the pattern of water recirculation determine the extent to which the system can recycle nutrients without collapsing. Although our results do not include microbiological variables, the strong ability of the models to discriminate classes from physicochemical parameters suggests that the interaction among microbiota, nutrients and solids is indirectly reflected in those measurements. This reasoning is consistent with the contributions of [8,16,18], who document how the use of probiotics and alternative carbon sources modifies the profiles of ammonium, nitrate and suspended solids. The fact that our model captures these variations without directly observing microbiota reinforces the idea that water quality can serve as an operational proxy of more complex biological processes, as suggested by [62] in their studies on biofloc applied to different stages of the culture cycle.

At the level of digitalization, the results align with the transition toward aquaculture 4.0 based on continuous monitoring and intelligent analytics. Ref. [24] show that synthetic indices such as the Water Quality Index make it possible to characterize the status of water bodies at broad spatial scales, while Ref. [26] argue that such indices are useful for communicating risk to non-specialist stakeholders. Refs. [30,34] document experiences with IoT sensor networks that record in real time key variables such as oxygen and temperature. Our proposal adds an extra layer, demonstrating that a relatively small set of routinely collected physicochemical variables can feed machine learning models capable of discriminating with high precision between acceptable and residual conditions. In this sense, we position ourselves within the same logic as [17,28], who propose monitoring architectures that combine data acquisition, preprocessing and predictive models, although here we emphasize a systematic comparison between supervised and unsupervised approaches using data from commercial ponds across different provinces.

Refs. [29,63] have shown that integrating cloud-based prediction services with visualization platforms helps farmers use advanced analytics results without needing to master the technical details. Ref. [32] adds that real-time prediction of variables such as pH makes it possible to anticipate deviations before they become visible in mortality or growth. The performance metrics obtained in this study are compatible with that agenda, indicating that it is technically feasible to build a service that, from the water parameters already measured by farmers, provides an operational classification of pond status. The proposal by [34] to develop alert platforms for shrimp production finds a concrete connection here, since the classification model could constitute the analytical core of recommendation systems guiding decisions on water exchange, feed adjustment or probiotic application.

From a methodological standpoint, the superior performance of Random Forest and Gradient Boosting can be attributed to their ability to capture nonlinear relationships and complex interactions among physicochemical variables. Water quality processes in shrimp ponds are governed by threshold effects and synergistic interactions between nitrogen compounds, alkalinity, and salinity, which are poorly represented by linear decision bound-

aries. Tree-based ensemble models recursively partition the feature space, allowing them to model abrupt transitions between acceptable and residual water conditions. In contrast, logistic regression and margin-based models assume smoother class boundaries, which limits their sensitivity to localized patterns associated with incipient nitrogen accumulation. This explains why ensemble models achieve higher precision and F1 scores, particularly for the minority class [35,49].

Who review machine learning applications in water quality and emphasize that tree-based ensembles tend to better handle non-linear relationships and interactions among variables. Ref. [64]'s arguments regarding Random Forest as a robust method under noise and class imbalance, together with the contributions of [40,41] on efficient boosting algorithms, help explain why these models capture the boundary between acceptable and residual water with such precision. At the same time, the fact that logistic regression and margin-based models such as SVM achieve performances close to those of ensembles suggests that the underlying structure of the problem is not excessively complex, in line with the views of [47,65,66], on the competitiveness of relatively simple models when variables are well chosen and data are properly partitioned.

The comparative analysis with unsupervised models introduces important nuances. Ref. [45] explain that, along continuous ecological gradients, the clusters generated by standard algorithms do not necessarily match management categories defined through thresholds. Our results reflect this tension precisely. On one hand, K-means, GMM, agglomerative and spectral clustering can recover, once labels are assigned by majority vote, classification patterns very similar to those of supervised models, echoing [46] observation that even geometrically imperfect partitions may be useful if interpreted with external criteria. On the other hand, low Silhouette and ARI values indicate that the natural separation of the data does not coincide with the "acceptable" and "residual" labels, which is in line with the warnings of [47,48] regarding the indiscriminate use of clustering when categories respond to management objectives rather than intrinsic discontinuities in the system. In other words, unsupervised models can approximate the operational classification only after the researcher imposes a mapping onto existing classes, but they do not by themselves identify clusters as clearly separated as one would desire.

This tension between ecological structure and management categories has practical consequences. Refs. [21,56] argue that water quality ranges used for decision-making in biofloc and intensive systems embody trade-offs between productivity, animal welfare and sanitary risk. Refs. [8,46] further elaborate how those trade-offs translate into operational thresholds for nitrogenous compounds and solids. Our models capture precisely those thresholds, supporting the idea that, when the goal is to issue management recommendations, supervised approaches are more appropriate than clustering techniques. Nevertheless, in contexts where labels are scarce or uncertain, unsupervised methods can play the exploratory role that Legendre and Ref. [45] assign to multivariate analyses: helping to detect unexpected patterns, identify ponds that behave atypically, or suggest new variables to monitor.

The implications for aquaculture 4.0 become clearer when these results are linked to technological innovations in aeration, probiotics and pond bottom management. Refs. [17,67] show that redesigning aerators and microbubble generators increases oxygenation efficiency and improves mixing within ponds. If a classification module such as the one proposed here is added to such solutions, farmers could automate responses to incipient deviations in ammonium or nitrate, for example, by temporarily increasing aeration or triggering partial water exchanges. In parallel, the evidence compiled by Refs. [8,16,19] on probiotics and alternative carbon sources shows that nutraceutical decisions are reflected in the same parameters used by our model. Integrating supervised

models with close tracking of these practices would make it possible, as Ref. [18] suggests, to distinguish between expected changes due to management adjustments and early signs of environmental deterioration.

Finally, it is important to situate this work within the broader framework of mathematical modeling and risk management across the production cycle. Ref. [67] uses deterministic models to simulate culture scenarios under different management schemes and shows that small adjustments in feeding regimes or water renewal can have substantial cumulative effects. Ref. [68] emphasize that failures in hatchery water quality propagate consequences throughout the entire production system. Our results speak directly to these concerns by demonstrating that, from relatively simple data collected across multiple provinces, it is possible to build robust classification models that support operational decisions and reduce the likelihood of critical events. In terms of future research, this opens the door to exploring time-series models, semi-supervised approaches and deeper architectures, as recommended by [35,64]. The integration of WQI-type indices, as proposed by Ref. [8], with data from low-power IoT sensors, following the direction indicated by Refs. [29,34], could ultimately consolidate more robust and accessible early-warning systems, especially for small and medium-scale producers who cannot afford severe failures in water quality.

6. Conclusions

The results of this study show that shrimp ponds operating under commercial conditions can maintain a relatively stable physicochemical environment, but often do so under a latent nitrogen load that keeps the system close to operational thresholds. The classification models confirm that a small set of routine water quality variables, particularly those related to inorganic nitrogen, is sufficient to distinguish between acceptable and residual water with high reliability. This finding suggests that data-driven tools can anticipate many risks that are traditionally detected only when mortalities or reduced growth become evident, using information that producers already collect during routine management.

The comparison between supervised and unsupervised approaches indicates that management categories such as acceptable and residual water are better captured when the model has access to prior labels. Supervised algorithms, especially tree-based ensembles and margin-based models, showed the most consistent performance, while clustering methods only approximated the same pattern after an *ex post* mapping of clusters to classes. This reinforces the idea that operational thresholds in aquaculture are social and productive constructs anchored in biology, more than natural breaks in the structure of the data, and that analytical tools need to respect that logic when they are intended to support decision making on farms.

From an applied perspective, the study demonstrates that integrating machine learning with routine monitoring opens a realistic path toward early warning systems for shrimp farming. A decision support service that ingests the same water quality parameters already recorded on farms could provide real-time or near-real-time alerts about ponds drifting toward residual conditions, and could be linked to recommendations on aeration, water exchange or feeding strategies. This aligns with the broader vision of aquaculture 4.0, where low-cost sensing, cloud computing and predictive analytics converge to reduce uncertainty, improve animal welfare and protect the economic viability of small- and medium-scale producers.

At the same time, the study has limitations that point toward future research. The analysis is based on static measurements rather than continuous time series, it does not include direct indicators of microbiota, organic load or sediment quality, and it is constrained to a specific set of ponds and regions. Future work should incorporate temporal dynamics, expand the range of monitored variables, explore semi-supervised and deep learning ar-

chitectures and validate the models in operational pilot projects with farmers. Addressing these gaps will help transform the current proof of concept into a robust, scalable and accessible tool for risk management in shrimp aquaculture.

Supplementary Materials: To facilitate the replication of this study, a GitHub repository (version 3.5.4) has been created and is available at <https://github.com/hvillamarb/Early-anomaly-detection-in-shrimp> (accessed on 5 December 2025). The data are proprietary and therefore not authorized for public dissemination.

Author Contributions: Conceptualization, H.V.-B. and J.C.-R.; methodology, H.V.-B., J.C.-R. and A.H.-S.; software, H.V.-B.; validation, H.V.-B., J.C.-R. and A.H.-S.; formal analysis, H.V.-B. and A.H.-S.; investigation, H.V.-B. and J.C.-R.; resources, J.C.-R.; data curation, H.V.-B.; writing—original draft preparation, H.V.-B. and A.H.-S.; writing—review and editing, J.C.-R. and A.H.-S.; visualization, A.H.-S.; supervision, J.C.-R.; project administration, J.C.-R.; funding acquisition, H.V.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC was funded by the Universidad Agraria del Ecuador.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We gratefully acknowledge the institutional support from the Universidad Agraria del Ecuador that made this research possible, including the provision of technological resources for implementing supervised and unsupervised machine learning models for early anomaly detection in shrimp pond water quality.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|--------|---|
| ARI | Adjusted Rand Index |
| DBSCAN | Density-Based Spatial Clustering of Applications with Noise |
| GMM | Gaussian Mixture Model |
| IoT | Internet of Things |
| NMI | Normalized Mutual Information |
| PSU | Practical Salinity Unit |
| RBF | Radial Basis Function |
| SVM | Support Vector Machine |
| WQI | Water Quality Index |

References

1. Hu, M.; Wang, Y.; Chen, J.; Cui, H.; Zhu, S.; Jin, T.; Qu, K.; Cui, Z. Optimizing shrimp growth and water quality in constructed wetland recirculating aquaculture systems: Effects of stocking density and shrimp size. *Aquaculture* **2026**, *612*, 743222. [[CrossRef](#)]
2. Nazarudin, M.F.; Zulkiply, M.A.F.; Samsuri, M.H.; Khairil Anwar, N.A.S.; Jamal, N.S.A.; Alipiah, N.M.; Ahmad, M.I.; Nor, N.M.; Yasin, I.S.M.; Ikhsan, N.; et al. Optimizing shrimp culture through environmental monitoring: Effects of water quality and metal ion profile on whiteleg shrimp (*Litopenaeus vannamei*) performance in a semi-intensive culture pond. *Water* **2025**, *17*, 2818. [[CrossRef](#)]
3. Dolatabadi, S.; Ricardez-Sandoval, L. Integration of process design and control of a pilot-scale recirculating aquaculture system. *Aquac. Eng.* **2025**, *111*, 102580. [[CrossRef](#)]
4. Suasono, Z.S.; Setiawardhana, S.; Gunawan, A.I.; Winarno, I. Performance evaluation of water quality for shrimp farming using deep learning classification. *Aquac. Eng.* **2026**, *112*, 102648. [[CrossRef](#)]

5. Kuncha, P.; Manoranjini, J.; Sirisha, J.; Bandeela, S.; Penjarla, N.K.; Goud, S.S. Advanced Aquaculture Management: A Smart System for Optimizing Oxygen Levels, Shrimp Health Monitoring. In *Smart Factories for Industry 5.0 Transformation*; Wiley: Hoboken, NJ, USA, 2025; pp. 283–298. [[CrossRef](#)]
6. Mustafa, A.; Syah, R.; Paena, M.; Tarunamulia; Samad, W.; Ratnawati, E.; Kamariah; Athirah, A.; Asaf, R.; Akmal; et al. Evaluating the performance of the wastewater treatment plant in intensive whiteleg shrimp (*Litopenaeus vannamei*) brackishwater pond aquaculture. *Environ. Sci. Pollut. Res. Int.* **2025**, *32*, 14220–14246. [[CrossRef](#)] [[PubMed](#)]
7. Binh, P.T.; Van, P.T.; Nghia, N.H.; Huy, T.T.; May, L.T.; St-Hilaire, S.; Giang, P.T. Impact of ozone nanobubble on water quality, gut microbiota, and growth performance of white leg shrimp (*Penaeus vannamei*) in an intensive indoor farming system. *Energy Nexus* **2025**, *18*, 100450. [[CrossRef](#)]
8. Bajracharya, S.; Appuhami, I.A.; Bruce, T.J.; Roy, L.A.; García, J.C.; Davis, D.A. Effects of pelleted probiotic on growth, water quality, and disease resistance in Pacific white shrimp (*Litopenaeus vannamei*) in static biofloc systems. *Aquac. Res.* **2025**, *2025*, 4619797. [[CrossRef](#)]
9. Rahmawati, A.I.; Saputra, R.N.; Hidayatullah, A.; Dwiarto, A.; Junaedi, H.; Cahyadi, D.; Saputra, K.H.; Prabowo, W.T.; Kartamiharja, U.K.A. Enhancement of *Penaeus vannamei* shrimp growth using nanobubble in indoor raceway pond. *Aquac. Fish.* **2021**, *6*, 277–282. [[CrossRef](#)]
10. Khanjani, M.H.; Eslami, J.; Emerenciano, M.G.C. Wheat flour as carbon source on water quality, growth performance, hemolymph biochemical and immune parameters of Pacific white shrimp (*Penaeus vannamei*) juveniles in biofloc technology (BFT). *Aquac. Rep.* **2025**, *40*, 102623. [[CrossRef](#)]
11. Mustafa, A.; Syah, R.; Paena, M.; Sugama, K.; Kontara, E.K.; Muliawan, I.; Suwoyo, H.S.; Asaad, A.I.J.; Asaf, R.; Ratnawati, E.; et al. Strategy for developing whiteleg shrimp (*Litopenaeus vannamei*) culture using intensive/super-intensive technology in Indonesia. *Sustainability* **2023**, *15*, 1753. [[CrossRef](#)]
12. Long, L.; Liu, H.; Lu, S. Effects of low salinity on growth, digestive enzyme activity, antioxidant and immune status, and the microbial community of *Litopenaeus vannamei* in biofloc technology aquaculture systems. *J. Mar. Sci. Eng.* **2023**, *11*, 2076. [[CrossRef](#)]
13. Huang, H.-H.; Cheng, C.; Guo, L.-L.; Zou, W.-S.; Lei, Y.-J.; Kuang, W.-Q.; Zhou, B.-L.; Yang, P.-H.; Li, C.-Y. Effects of shifts in bacterial community on improving water quality and growth performance of pacific whiteleg shrimp (*Litopenaeus vannamei*) in biofloc systems. *Fishes* **2025**, *10*, 626. [[CrossRef](#)]
14. Han, T.; Zhang, M.; Feng, W.; Li, T.; Liu, X.; Wang, J. Effects of aeration intensity on water quality, nutrient cycling, and microbial community structure in the biofloc system of pacific white shrimp *Litopenaeus vannamei* culture. *Water* **2024**, *17*, 41. [[CrossRef](#)]
15. Chainark, S.; Sumetlux, V.; Chainark, P. Dynamics of soil properties and pathogen levels in Pacific white shrimp ponds during a production cycle: Implications for aquaculture management. *J. World Aquac. Soc.* **2025**, *56*, e70002. [[CrossRef](#)]
16. Li, Z.; Du, Q.; Jiao, T.; Zhu, Z.; Wan, X.; Ju, C.; Liu, H.; Li, Q. Effects of probiotic supplementary bioflocs (*Rhodospirillum rubrum*, *Bacillus subtilis*, *Providencia rettgeri*) on growth, immunity, and water quality in cultures of Pacific white shrimp (*Litopenaeus vannamei*). *Aquaculture* **2024**, *591*, 741141. [[CrossRef](#)]
17. Jayraj, P.; Sahoo, S.; Jana, P.; Prem, R. CFD optimised solar powered IoT integrated paddlewheel aerator for sustainable shrimp farming—A review. *Aquac. Eng.* **2025**, *111*, 102583. [[CrossRef](#)]
18. Huang, H.-H.; Li, C.-Y. Adaptability of commercial probiotics to biofloc system: Influences on autochthonal bacterial community, water quality and growth performance of shrimp (*Litopenaeus vannamei*). *Aquaculture* **2024**, *590*, 740992. [[CrossRef](#)]
19. Sharma, A.; Ramena, G.; Segree, A.; Bohora, K.; Karim, F.; Meesala, K.-M.; Chennault, D.; Ramena, Y. Probiotics and immunostimulants modulate gut microbiome and immune gene expression in post-larval *Litopenaeus vannamei*. *Comp. Immunol. Rep.* **2025**, *10*, 200267. [[CrossRef](#)]
20. Menaga, M.; Rajasulochana, P.; Felix, S.; Sudarshan, S.; Kapoor, A.; Gandla, K.; Saleh, M.M.; Ibrahim, A.E.; El Deeb, S. Evaluation of biofloc-based probiotic isolates on growth performance and physiological responses in *Litopenaeus vannamei*. *Water* **2023**, *15*, 3010. [[CrossRef](#)]
21. Crab, R.; Defoirdt, T.; Bossier, P.; Verstraete, W. Biofloc technology in aquaculture: Beneficial effects and future challenges. *Aquaculture* **2012**, *356–357*, 351–356. [[CrossRef](#)]
22. Ren, H.; Xu, Y.; Jing, L.; Su, H.; Hu, X.; Cao, Y.; Wen, G. The effects of two different aquaculture methods on water quality, microbial communities, production performance, and health status of *Penaeus monodon*. *Fishes* **2025**, *10*, 106. [[CrossRef](#)]
23. Sri Bala, G.; Nagaraju, T.V.; Krishnam Raju, G.L.V.; Srinivasa Rao, G.V.R. Integrating technology and sustainability in inland aquaculture water management. In *Inland Aquaculture Sustainability and Effective Water Management Strategies*; Springer Nature: Cham, Switzerland, 2025; pp. 33–48. [[CrossRef](#)]
24. Gottumukkala, S.B.; Thotakura, V.N.; Gvr, S.R.; Chinta, D.P.; Park, R. Balancing aquaculture and estuarine ecosystems: Machine learning-based water quality indices for effective management. *Environ. Sci. Pollut. Res.* **2024**, *32*, 31145–31161. [[CrossRef](#)]

25. Flores-Iwasaki, M.; Guadalupe, G.A.; Pachas-Caycho, M.; Chapa-Gonza, S.; Mori-Zabarburú, R.C.; Guerrero-Abad, J.C. Internet of things (IoT) sensors for water quality monitoring in aquaculture systems: A systematic review and bibliometric analysis. *AgriEngineering* **2025**, *7*, 78. [CrossRef]
26. Mustafa, A.; Paena, M.; Athirah, A.; Ratnawati, E.; Asaf, R.; Suwoyo, H.S.; Sahabuddin, S.; Hendrajat, E.A.; Kamaruddin, K.; Septiningsih, E.; et al. Temporal and spatial analysis of coastal water quality to support application of whiteleg shrimp *Litopenaeus vannamei* intensive pond technology. *Sustainability* **2022**, *14*, 2659. [CrossRef]
27. Ilmiah, J.; Elektro, T.; Eso, R. IoT-Based Vaname Shrimp Pond Water Quality Monitoring Using the Quamonitor Tool. 2024. Available online: <https://jurnalelectron.org/index.php/electronubb/article/download/149/50> (accessed on 1 November 2025).
28. Do, D.D.; Le, A.H.; Van Vu, V.; Le, D.A.N.; Bui, H.M. Evaluation of water quality and key factors influencing water quality in intensive shrimp farming systems using principal component analysis-fuzzy approach. *Desalination Water Treat.* **2025**, *321*, 101002. [CrossRef]
29. Suasono, Z.S.; Setiawardhana, S.; Winarno, I.; Gunawan, A.I. Cloud computing-based shrimp pond water quality prediction intelligent service system. *JOIV Int. J. Inform. Vis.* **2025**, *9*, 1347. [CrossRef]
30. Ahmed, F.; Bijoy, M.H.I.; Hemal, H.R.; Noori, S.R.H. Smart aquaculture analytics: Enhancing shrimp farming in Bangladesh through real-time IoT monitoring and predictive machine learning analysis. *Heliyon* **2024**, *10*, e37330. [CrossRef]
31. Yudamson, A.; Sulistiyanti, S.R.; Saputra, M.I.; Hendrawan, B. Development of an internet of things-based water quality monitoring system for shrimp ponds utilizing Mappi32. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2025**, *15*, 426–435. [CrossRef]
32. Hridoy, M.A.A.M.; Bordin, C.; Masood, A.; Masood, K. Predictive modelling of aquaculture water quality using IoT and advanced machine learning algorithms. *Results Chem.* **2025**, *16*, 102456. [CrossRef]
33. Mayekar, T.S.; Paramesha, V.; Sreekanth, G.B.; Rivonker, C.U.; Kumar, P. Life cycle assessment of whiteleg shrimp farming in earthen vs. HDPE-Lined Ponds India. *Clean. Environ. Syst.* **2025**, *19*, 100342. [CrossRef]
34. Aryotejo, G.; Adi, P.W.; Sarwoko, E.A. Water quality monitoring with an early warning system for enhancing the shrimp aquaculture production. *Indones. J. Electr. Eng. Comput. Sci.* **2024**, *34*, 1042–1051. [CrossRef]
35. Yan, X.; Zhang, T.; Du, W.; Meng, Q.; Xu, X.; Zhao, X. A comprehensive review of machine learning for water quality prediction over the past five years. *J. Mar. Sci. Eng.* **2024**, *12*, 159. [CrossRef]
36. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Int. Jt. Conf. Artif. Intell.* **1995**, *2*, 1137–1145. Available online: https://www.researchgate.net/profile/Ron-Kohavi/publication/2352264_A_Study_of_Cross-Validation_and_Bootstrap_for_Accuracy_Estimation_and_Model_Selection/links/02e7e51bcc14c5e91c000000/A-Study-of-Cross-Validation-and-Bootstrap-for-Accuracy-Estimation-and-Model-Selection.pdf (accessed on 20 November 2025).
37. Akhtar, S.; Bisal, P.; Jana, P. An innovative method for predictive maintenance of cold stores in shrimp processing industry using machine learning and data trends. *Int. J. Mach. Tools Maint. Eng.* **2024**, *5*, 43–47. [CrossRef]
38. Hosmer, D.W.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*, 3rd ed.; Nashville, TN, Ed.; John Wiley & Sons: Hoboken, NJ, USA, 2013; Available online: https://www.researchgate.net/profile/Andrew-Cucchiara/publication/261659875_Applied_Logistic_Regression/links/542c7eff0cf277d58e8c811e/Applied-Logistic-Regression.pdf (accessed on 21 November 2025).
39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
40. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
41. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 13–17 August 2016*; ACM: New York, NY, USA. [CrossRef]
42. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
43. Vapnik, V. The support vector method of function estimation. In *Nonlinear Modeling*; Springer: Boston, MA, USA, 1998; pp. 55–85. [CrossRef]
44. Schölkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. 2002. Available online: https://pure.mpg.de/rest/items/item_3271060_1/component/file_3271061/content (accessed on 30 November 2025).
45. Legendre, P.; Legendre, L.; Legendre, L.F. *Numerical Ecology*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2012. Available online: https://books.google.com.ec/books?hl=es&lr=&id=6ZBOA-iDviQC&oi=fnd&pg=PP1&dq=Legendre+%26+Legendre,+2012&ots=uAbn16U3Wi&sig=4gxscQuen1w_PwdiWeHy_ZU9q-4 (accessed on 16 November 2025).
46. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [CrossRef]
47. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009. [CrossRef]
48. Everitt, B.; Hothorn, T. *An Introduction to Applied Multivariate Analysis with R*; Springer: New York, NY, USA, 2011.
49. Zhu, D.; Li, Z.; Hu, P.; Su, Q.; Liu, R. Improved DBSCAN algorithm based on relative mass of the data field. In *International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2021)*; Wu, F., Liu, J., Chen, Y., Eds.; SPIE: Amsterdam, The Netherlands, 2022. [CrossRef]

50. Yang, Y.; Liu, Y.; Li, G.; Zhang, Z.; Liu, Y. Harnessing the power of Machine learning for AIS Data-Driven maritime Research: A comprehensive review. *Transp. Res. Part. E Logist. Trans. Rev.* **2024**, *183*, 103426. [CrossRef]
51. Pantjara, B.; Syafaat, M.N.; Kristanto, A.H. Effect of dynamical water quality on shrimp culture in the integrated multitropic aquaculture (imta). *Indones. Aquac. J.* **2015**, *10*, 81. [CrossRef]
52. Le, D.Q.; Binh, H.T.; Tuyet, D.T.A.; Thanh, N.X.; Nhi, P.T.T.; Bao, D.N.; Nam, L.V.; Tuan, L.M.; Nguyet, N.T.; Hao, D.M. Improving whiteleg shrimp (*Litopenaeus vannamei*) performance and water quality in low-salinity BioRAS: Role of hydraulic retention time and dietary minerals. *Aquac. Int.* **2025**, *33*, 550. [CrossRef]
53. Estante-Superio, E.G.; de la Peña, L.D.; Geanga, T.M.M.; Castellano, J.L.A.; Cordero, C.P.; Berlin, S.C.; Lazado, C.C. The impact of indoor biofloc-based system on water quality, growth, and disease resistance of black tiger shrimp. *Aquac. Eng.* **2025**, *111*, 102564. [CrossRef]
54. Zheng, S.; Zou, S.; Wang, H.; Feng, T.; Sun, S.; Chen, H.; Wang, Q. Reducing culture medium nitrogen supply coupled with replenishing carbon nutrient simultaneously enhances the biomass and lipid production of *Chlamydomonas reinhardtii*. *Front. Microbiol.* **2022**, *13*, 1019806. [CrossRef]
55. Chen, J.; Li, J.; Li, W.; Li, P.; Zhu, R.; Zhong, Y.; Zhang, W.; Li, T. The optimal ammonium-nitrate ratio for various crops: A Meta-analysis. *Field Crops Res.* **2024**, *307*, 109240. [CrossRef]
56. Hussain, A.S.; Mohammad, D.A.; Sallam, W.S.; Shoukry, N.M.; Davis, D.A. Effects of culturing the Pacific white shrimp *Penaeus vannamei* in “biofloc” vs. “synbiotic” systems on the growth and immune system. *Aquaculture* **2021**, *542*, 736905. [CrossRef]
57. Davis, R.P.; Lutz, G.; Boyd, C.E.; McNevin, A.A. Volume 3: Aquaculture and Living Resource Management. In *Crustacean Aquaculture*; CRC Press: Boca Raton, FL, USA, 2025; pp. 52–86. [CrossRef]
58. Djumanto; Rustadi; Triyatmo, B.; Istigomah, I.; Priyono, S.B.; Hikma, A.S.; Laily, F.; Deendarlianto. The impact of microbubble generator technology on vannamei shrimp growth and water quality. *Aquac. Aquar. Conserv. Legis.* **2025**, *18*, 600–610. Available online: <https://search.proquest.com/openview/e55b986d5561bfcb88c4e5849f63bad0/1?pq-origsite=gscholar&cbl=2046424> (accessed on 18 November 2025).
59. Stähli, M.; Sättele, M.; Huggel, C.; McArdell, B.W.; Lehmann, P.; Van Herwijnen, A.; Berne, A.; Schleiss, M.; Ferrari, A.; Kos, A.; et al. Monitoring and prediction in early warning systems for rapid mass movements. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 905–917. [CrossRef]
60. Ray, B.; Bhunia, A. *Fundamental Food Microbiology*; CRC Press: Boca Raton, FL, USA, 2024. [CrossRef]
61. Emmanuel, A.; Raza, B.; Ramzan, M.N.; Zheng, Z. Bacterial community of the shrimp (*Litopenaeus vannamei*) gut and its relationship with water quality in a recirculating aquaculture system (RAS). *Aquac. Int.* **2025**, *33*, 640. [CrossRef]
62. Gunarto, G.; Muliani, M.; Suwoyo, H.S.; Septiningsih, E. Effects of mangrove leaf litter on the water quality, growth, and survival of tiger shrimp (*Penaeus monodon* Fabricius, 1798) post-larvae. *Aquac. Int.* **2025**, *33*, 7. [CrossRef]
63. Thongtha, K.; Pochai, N. A simple mathematical model for assessing water quality in a closed-system shrimp farm. *J. Interdiscip. Math.* **2025**, *28*, 2031–2043. [CrossRef]
64. Zhou, Z.H. Ensemble Methods: Foundations and Algorithms. 2025. Available online: [https://books.google.com.ec/books?hl=es&lr=&id=hm0-EQAAQBAJ&oi=fnd&pg=PP1&dq=SVM+achieve+performances+close+to+those+of+ensembles+suggests+that+the+underlying+structure+of+the+problem+is+not+excessively+complex,+in+line+with+the+views+of+Hosmer+et+al.++\(2013\),&ots=WppEO8QNKp&sig=YT0bPKbyNiYoibx72tSaUhwy0-Q](https://books.google.com.ec/books?hl=es&lr=&id=hm0-EQAAQBAJ&oi=fnd&pg=PP1&dq=SVM+achieve+performances+close+to+those+of+ensembles+suggests+that+the+underlying+structure+of+the+problem+is+not+excessively+complex,+in+line+with+the+views+of+Hosmer+et+al.++(2013),&ots=WppEO8QNKp&sig=YT0bPKbyNiYoibx72tSaUhwy0-Q) (accessed on 17 November 2025).
65. James, C.; Ranson, J.M.; Everson, R.; Llewellyn, D.J. Performance of machine learning algorithms for predicting progression to dementia in memory clinic patients. *JAMA Netw. Open.* **2021**, *4*, e2136553. [CrossRef] [PubMed]
66. Thong-un, N.; Panthong, P.; Takahashi, H.; Kikura, H.; Wongsaroj, W. Development of an automatic paddle wheel aerator and re-mote movement water quality monitoring for use in a marine shrimp farm. *Preprints* **2023**. [CrossRef]
67. Xu, Z.; Li, M.; Wang, Y.; Feng, M.; Gan, Z.; Leng, X.; Li, X. Affecting mechanism of *Chlorella sorokiniana* meal replacing fish meal on growth and immunity of *Litopenaeus vannamei* based on transcriptome analysis. *Aquac. Rep.* **2023**, *31*, 101645. [CrossRef]
68. Zhu, M.; Wang, J.; Yang, X.; Zhang, Y.; Zhang, L.; Ren, H.; Wu, B.; Ye, L. A review of the application of machine learning in water quality evaluation. *Eco-Environ. Health* **2022**, *1*, 107–116. [CrossRef] [PubMed]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

- 99+ Mail
- Chat
- Meet

Gmail

Compose

Inbox 4,273

- Starred
- Snoozed
- Sent
- Drafts 10
- Categories
- Purchases 8
- More
- Labels +
- Notes

Search mail

Active

PKRU mail

1 of 5,757

Suwanit, a recent article cited your research External Inbox x

ResearchGate <no-reply@researchgatemail.net> Unsubscribe 7:41PM (3 hours ago)

ResearchGate

Suwanit, we found a recent article that cited you:

Early Anomaly Detection in Shrimp Pond Water Quality Using Supervised and Unsupervised Machine Learning Models

Citing article

Apr 2026

Digital

Hamilton Villamar-Barros · Julián Coronel-Reyes · Alexander Haro-Sarango

Reply Forward

Suwanit, we found a recent article that cited you:

Early Anomaly Detection in Shrimp Pond Water Quality Using Supervised and Unsupervised Machine Learning Models

Citing article

Apr 2026

Digital

Hamilton Villamar-Barros · Julián Coronel-Reyes · Alexander Haro-Sarango

View article

Your cited research:

Dynamics of soil properties and pathogen levels in Pacific white shrimp ponds during a production cy...

Journal of the World Aquaculture Society

Reply Forward

Article

Early Anomaly Detection in Shrimp Pond Water Quality Using Supervised and Unsupervised Machine Learning Models

April 2026 · Digital 6(2):27
DOI: [10.3390/digital6020027](https://doi.org/10.3390/digital6020027)
License · [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)


Hamilton Villamar-Barros · Julián Coronel-Reyes · Alexander Haro-Sarango

Research Interest Score 0.2
Citations 0
Recommendations 0
Reads ⓘ 6

[Learn about stats on ResearchGate](#)

Overview Stats Citations **References (65)** Request full-text Share More

References to your research (1)


 This publication is referenced:

Dynamics of soil properties and pathogen levels in Pacific white shrimp ponds during a production cycle: Implications for aquaculture management

[Full-text available](#) Article February 2025 · [Journal of the World Aquacultur...](#)

70 Reads · 4 Citations

Top referenced researchers

Arslan Emmanuel 

Ningbo University

Skills
Aquaculture · Microbial Molecular Biology · Antioxidants · Bacteriology · Microbial Ecology · Aquaculture Fisheries · shrimp...

Most Recent Research Contribution
Bacterial community of the shrimp (*Litopenaeus vannamei*) gut a ... [Read more](#)



- Submit to this Journal
- Review for this Journal
- Propose a Special Issue

Article Menu

Academic Editors

- Jorge Bernardino
- Le Gruenwald
- Elio Masciari

- Show more...
- Recommended Articles
- Related Info Link

Order Article Reprints

Open Access Article

Early Anomaly Detection in Shrimp Pond Water Quality Using Supervised and Unsupervised Machine Learning Models

by Hamilton Villamar-Barros^{1,*}, Julián Coronel-Reyes¹ and Alexander Haro-Sarango²

- ¹ Facultad de Ciencias Agrarias, Universidad Agraria del Ecuador, Guayaquil 090107, Ecuador
- ² Instituto Superior Tecnológico España, Ambato 180101, Ecuador
- * Author to whom correspondence should be addressed.

Digital 2026, 6(2), 27; <https://doi.org/10.3390/digital6020027>

Submission received: 9 December 2025 / Revised: 6 January 2026 / Accepted: 14 January 2026 / Published: 1 April 2026

(This article belongs to the Special Issue Applications of Artificial Intelligence and Data Management in Data Analysis)

- Download
- Browse Figure
- Versions Notes

Abstract

Shrimp aquaculture increasingly depends on precise water quality management, yet most farms still rely on fragmented measurements and qualitative assessments. This study aimed to evaluate whether routine

- Share
- Help
- Cite
- Discuss in SciProfiles

View PDF **EN**

More Information

Powered by Usercentrics Consent Management

Deny Accept All