International Journal of Marine Science, 2026, Vol.16, No.5, 271-282
http://www.aquapublisher.com/index.php/ijms

Feature Review Open Access

Data-Based Prediction of Fish Growth Based on Water Quality Parameters

Haimei Wang
Hainan Institute of Biotechnology, Haikou, 570206, Hainan, China
Corresponding author: hamei.wang@hitar.org

International Journal of Marine Science, 2026, Vol.16, No.5, doi: 10.5376/ijms.2026.16.0021
Received: 20 Jul., 2026
Accepted: 24 Aug., 2026
Published: 06 Sep., 2026

Copyright © 2026 Wang, This is an open access article published under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preferred citation for this article:
Wang H.M., 2026, Data-based prediction of fish growth based on water quality parameters, International Journal of Marine Science, 16(5): 271-282 (doi: 10.5376/ijms.2026.16.0021)

Abstract Accurate prediction of fish growth is essential for improving aquaculture productivity, optimizing feeding strategies, and achieving sustainable management of aquatic systems. Water quality parameters are closely associated with fish physiological processes, metabolic activity, and growth performance; however, the complex and nonlinear interactions among environmental factors make traditional prediction approaches insufficient for dynamic aquaculture conditions. This paper systematically reviews data-based approaches for predicting fish growth based on water quality parameters, focusing on data acquisition, feature construction, machine learning modeling, and practical applications in precision aquaculture. Key environmental variables, including temperature, dissolved oxygen, pH, salinity, ammonia nitrogen, nitrite nitrogen, and organic matter, are integrated with biological growth indicators to establish predictive models. Various statistical and artificial intelligence methods, such as multiple linear regression, random forest, support vector machines, gradient boosting algorithms, and deep learning models, provide effective tools for capturing complex relationships between environmental fluctuations and fish growth dynamics. Model interpretation approaches, including feature importance analysis and explainable artificial intelligence, further clarify the contribution of individual water quality factors to growth variation. A case-study framework demonstrates how continuous water quality monitoring combined with machine learning can identify critical environmental drivers, predict growth trends, and support adaptive aquaculture management. Overall, data-driven growth prediction models provide a powerful approach for transforming water quality information into actionable decisions, enabling intelligent feeding regulation, environmental control, and sustainable fish production.

Keywords Fish growth prediction; Water quality parameters; Machine learning; Precision aquaculture; Data-driven modeling

1 Introduction

Modern aquaculture increasingly depends on accurate prediction of fish growth because growth performance sits at the center of production planning, economic return, and biological management. As production systems intensify, farmers must make earlier and more precise decisions about stocking density, feeding schedules, harvest timing, and market coordination, and those decisions are more effective when growth can be forecast rather than only measured retrospectively at harvest (Siddique et al., 2025). This need is even stronger in technology-enabled and large-scale systems, where monitoring and control must shift from experience-based husbandry toward knowledge-based management supported by continuous sensing and predictive models. Traditional growth descriptors such as absolute growth, relative growth, and specific growth rate remain useful for reporting outcomes, but they simplify fish development and often fail to capture intermediate growth dynamics across life stages, limiting their value for forward-looking farm decisions. For this reason, fish growth prediction is no longer only a biological question; it has become a core management requirement for improving efficiency, sustainability, and responsiveness in modern aquaculture operations.

Among the many factors affecting fish performance, water quality is one of the most direct and operationally important determinants of growth. Aquaculture water conditions govern physiological homeostasis, and when parameters such as temperature, dissolved oxygen, pH, ammonia, salinity, or other chemical conditions drift outside suitable ranges, fish divert energy from growth toward stress adaptation and become more vulnerable to disease and poor production outcomes. Even slight changes in major parameters such as pH, temperature, and dissolved oxygen can induce physiological or behavioral stress, while deteriorated water quality reduces growth and reproductive capacity. Behavioral evidence strengthens this link, because fluctuations in temperature, turbidity,

dissolved oxygen, salinity, pH, and inorganic nitrogen alter swimming, feeding, avoidance, aggression, and other behaviors that reflect welfare status and strongly condition growth performance (Zhang et al., 2024). Empirical pond studies in tilapia likewise show that increases in temperature and dissolved oxygen are associated with higher growth, whereas increases in conductivity, pH, and ammonia can depress growth, confirming that growth prediction models must incorporate water-quality variation rather than treat growth as an isolated output.

These biological and environmental relationships have driven the development of data-driven prediction approaches for aquaculture. Recent work shows that intelligent monitoring systems can use continuous environmental and biological data to model the dynamic interactions underlying fish growth, replacing static one-time regressions with temporal forecasting frameworks (Saleh et al., 2026). More broadly, AI methods in aquaculture now analyze large volumes of sensor, camera, and management data to identify growth patterns, detect anomalies, and support real-time decisions on feeding, water regulation, and health management. Machine learning has therefore emerged as a key enabling technology for digital aquaculture, with applications spanning biomass evaluation, fish classification, behavioral analysis, and prediction of water-quality parameters themselves. At the same time, time-series models such as ARIMA and ARIMAX remain useful because they can capture growth trends and seasonality, while also incorporating exogenous drivers such as climatic and water-quality variables to refine forecasts for management purposes (Siddique et al., 2025).

Against this background, prediction of fish growth based on water quality parameters represents a necessary step toward more precise and sustainable aquaculture management. Recent forecasting studies in recirculating and pond-based systems show that growth can be predicted from water temperature, dissolved oxygen, pH, feeding inputs, and related environmental signals with good accuracy, and that models become more informative when they respect temporal structure and biological constraints (Zhang et al., 2025). Deep learning evidence further suggests that multivariate forecasting across multiple ponds can outperform recurrent baselines by 30-50% in error metrics while maintaining robust performance under variable environmental conditions, indicating strong potential for operational deployment (Saleh et al., 2026). However, the field still needs models that are not only accurate, but also interpretable, transferable across production settings, and explicitly centered on the most actionable water-quality indicators. Therefore, the objective of this paper is to examine how water-quality parameters can be used as the core explanatory variables in a data-based fish growth prediction framework, and to provide a basis for more reliable forecasting, better environmental regulation, and improved production efficiency in aquaculture.

2 Relationship Between Water Quality Parameters and Fish Growth Dynamics

2.1 Effects of temperature, dissolved oxygen, PH, and salinity on fish physiological processes

Temperature and dissolved oxygen are primary regulators of fish physiology because they directly determine metabolic demand and aerobic capacity. Temperature strongly affects fish physiology, feed intake, survival, growth, reproduction, and energy acquisition, while rising temperature increases metabolic rate and therefore food demand (Islam et al., 2021; El-Hack et al., 2022). Dissolved oxygen then constrains whether this higher metabolic demand can be supported, since hypoxia reduces respiration, feeding activity, and growth rate and triggers behavioral and metabolic adjustments to cope with oxygen limitation (Canosa and Bertucci, 2023). Experimental evidence in rainbow trout further shows that high temperature impairs immunity, digestion, and growth, but that elevated dissolved oxygen can partly offset these adverse effects by supporting antioxidant defenses under thermal stress.

pH and salinity influence growth more indirectly by regulating acid-base balance, ion homeostasis, and osmoregulatory cost. Changes in pH alter ion balance and acid-base regulation, and both acidic and alkaline conditions can impair enzyme activity, respiratory efficiency, and stress tolerance. Salinity similarly affects metabolism, nutrition, reproduction, and growth because changes in external salinity alter osmotic regulation and the energy required to maintain internal balance (Seale et al., 2024). Evidence from Nile tilapia indicates that hypoxia and elevated salinity together reduce growth, feed utilization, digestive enzyme activity, and immune performance, showing that salinity stress becomes more damaging when oxygen supply is limited.

2.2 Influence of nitrogenous compounds, turbidity, and organic matter on growth and health status

Nitrogenous compounds are among the most important chemical constraints on fish growth in intensive culture because ammonia and nitrite directly disturb homeostasis. Elevated ammonia alters growth and biochemical, physiological, and immunological responses, and its toxicity becomes more severe when combined with hypoxia or unfavorable temperature conditions (Parvathy et al., 2022). Nitrite toxicity is also substantial in intensive and recirculating systems, where it disrupts ion regulation through the gills and impairs growth, blood oxygen-carrying capacity, oxidative balance, and immune function (Jie et al., 2025). More broadly, poor water quality characterized by high ammonia or nitrite can suppress feed intake, which helps explain why nitrogen accumulation rapidly translates into lower growth and higher disease risk (El-Hack et al., 2022).

Turbidity, suspended solids, and organic matter influence growth through both direct physical stress and indirect effects on microbial and nitrogen dynamics. In recirculating systems, high total suspended solids are associated with increased organic matter, higher bacterial activity, elevated ammonia and nitrite, and a 14% lower fish biomass increase over 128 days, indicating that particle accumulation degrades both biofilter performance and fish production. Reviews of biofloc and nitrification processes reach the same conclusion mechanistically: suspended organic matter and nitrogen compounds can severely affect fish physiology, whereas well-managed microbial communities convert toxic nitrogen into less harmful or even nutritionally useful forms (Robles-Porchas et al., 2020). Turbidity also matters behaviorally, because it is one of the physical water-quality variables that alters sensory perception, swimming, feeding, avoidance, and other behaviors closely linked to welfare and growth performance (Figure 1) (Zhang et al., 2024).

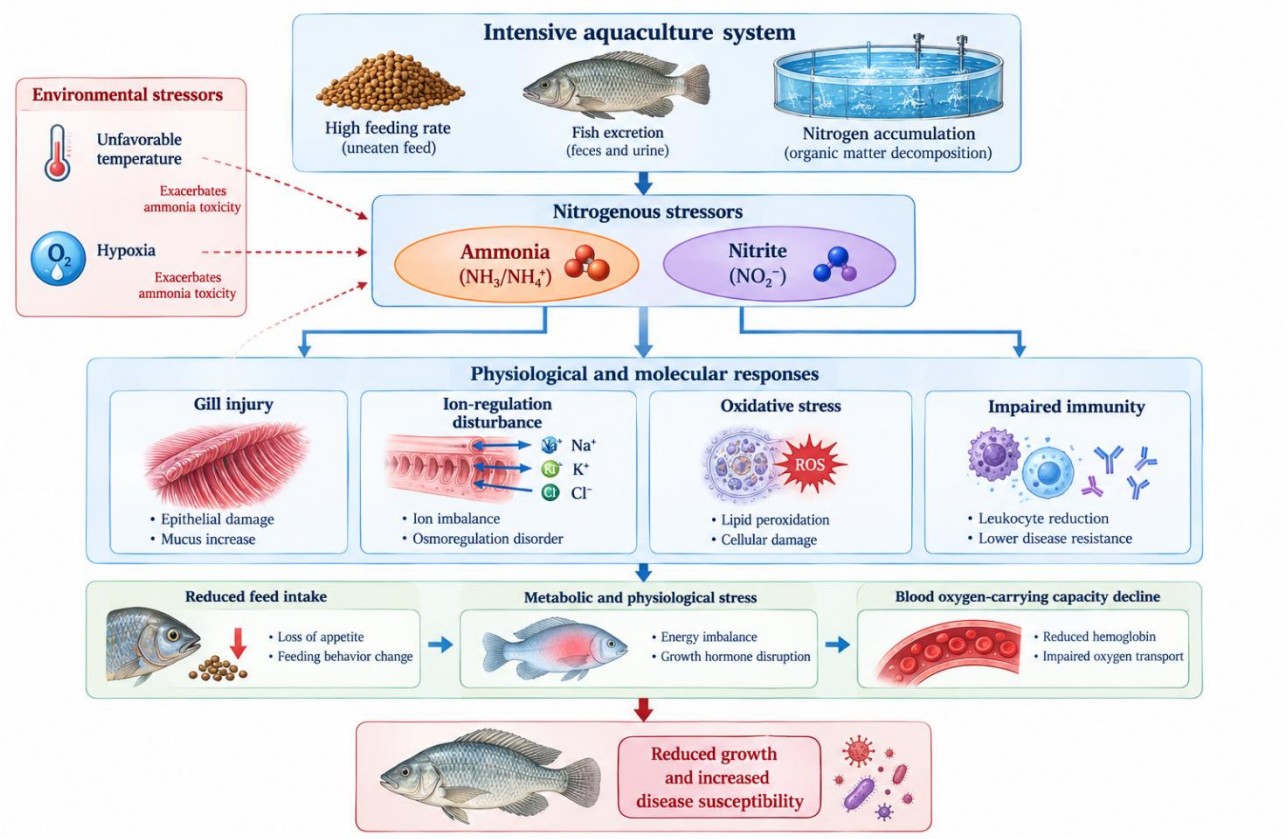


Figure 1 Conceptual mechanisms underlying the effects of ammonia and nitrite accumulation on fish growth and physiological performance in intensive aquaculture systems

2.3 Interactions between environmental factors, metabolic regulation, and growth performance

The effects of water quality on growth are not additive single-parameter responses, but integrated metabolic and endocrine reactions to multiple simultaneous stressors. Growth in teleost fish is regulated largely through the

GH/IGF axis, and environmental variables such as temperature, salinity, pollutants, and other external factors modify GH and IGF-I expression and thereby alter growth rate. Temperature is especially influential because it changes gene expression and metabolic enzyme activity and appears to regulate plasma GH and IGF-1 levels across seasons and species (Canosa and Bertucci, 2023). Under oxygen limitation, fish often maintain slower but still active growth by readjusting feeding behavior, cortisol, and IGF-related signaling, indicating that metabolic compensation rather than complete growth arrest can occur under moderate stress.

These interactions are highly relevant for predictive modeling because water-quality variables frequently modify one another and jointly shape performance outcomes. Dissolved oxygen depends partly on temperature and salinity as well as biological oxygen demand, so changes in one parameter can propagate through multiple physiological pathways (Canosa and Bertucci, 2023). Temperature extremes also amplify broader water-quality fluctuation and hydrological stress, while ammonia toxicity itself is modulated by pH, temperature, salinity, dissolved oxygen, species, and life stage (Islam et al., 2021; Parvathy et al., 2022). For growth prediction, this means that single-factor models will miss much of the biological signal, whereas multivariable approaches that incorporate interacting water-quality parameters are better aligned with the physiological reality of fish growth dynamics.

3 Data Acquisition and Feature Construction for Fish Growth Prediction

3.1 Collection and preprocessing of water quality and fish growth datasets

Reliable fish growth prediction begins with synchronized acquisition of water quality and growth measurements from aquaculture systems. IoT-based monitoring platforms now routinely collect temperature, pH, dissolved oxygen, turbidity, ammonia, and nitrate in real time, and these datasets are valuable because they can support decisions on water exchange, stocking density, feed conversion, and growth-rate prediction (Udanor et al., 2022). Publicly documented aquaponics datasets and cloud-linked sensor systems also address a persistent bottleneck in aquaculture analytics, namely the shortage of labeled water-quality data that are suitable for machine learning and decision support (Eneh et al., 2023).

Preprocessing is essential because fish growth studies rarely measure all variables at the same temporal resolution, and raw sensor streams often contain calibration drift and noise. In commercial flounder culture, temperature and dissolved oxygen were recorded every minute, feed was recorded daily, and fish weight was measured monthly, so temporal alignment was explicitly handled during preprocessing before modeling (Kim et al., 2025). Sensor correction and validation also improve downstream model quality, as low-cost aquaculture sensors achieved markedly higher agreement after regression-based calibration, with accuracy rising from 0.80 to 0.98 in one monitoring system (Jais et al., 2024).

3.2 Selection and optimization of environmental variables for predictive modeling

Environmental feature selection should focus on variables that are both biologically meaningful and empirically informative for growth. Across fish growth studies, water temperature, dissolved oxygen, pH, salinity, and feeding inputs are repeatedly identified as core predictors because they directly shape metabolic demand, oxygen supply, and nutrient utilization (Zhang et al., 2025). In more controlled indoor systems, variable-importance analysis can further refine this set, as abalone growth was influenced most strongly by time, dissolved oxygen, salinity, and nutrient supply, while temperature and pH contributed less under stable farming conditions.

Optimization of predictor sets also requires reducing redundancy and ranking variables by their actual predictive contribution rather than by convention alone. Machine-learning studies on aquaculture production show that feature-importance scoring can identify the most informative climatic inputs before model fitting, improving efficiency when many candidate variables are available but datasets remain limited. Similar evidence from aquaponics indicates that the most influential predictors can shift with data quality and production context, with turbidity, nitrate, pH, and temperature ranking highest in one sensor system, while ammonia, nitrate, total dissolved solids, and pH dominated after sensor improvement (Eneh et al., 2023).

3.3 Integration of temporal, spatial, and biological information for model development

Model development is stronger when environmental covariates are aligned to the actual exposure history of each fish rather than assigned as static background values. In rainbow trout, time between release and recapture was modeled as each fish's time-at-large, and environmental variables spanning multiple months were converted into time-weighted averages so that covariates reflected true ecological exposure (Fulton and Lyu, 2026). This biologically informed temporal integration improves ecological fidelity and avoids the distortion that occurs when growth observations are linked to arbitrary calendar points instead of exposure intervals (Saleh et al., 2026).

Spatial and biological information should also be incorporated because fish in different ponds, tanks, or workshops often experience distinct environments and management conditions that alter growth trajectories. Practical production studies show that differences among ponds and operators can generate systematic deviations in fish growth, which is why adaptive models that update with new measurements are useful for pond-specific prediction. Recent deep-learning frameworks extend this idea by combining continuous sensor streams, sliding windows, multi-resolution temporal resampling, and cross-pond validation, and they show that models can generalize across ponds when temporal and biological structure is encoded explicitly (Saleh et al., 2026).

4 Machine Learning and Statistical Models for Predicting Fish Growth

4.1 Traditional statistical approaches for growth prediction based on environmental parameters

Traditional fish growth modeling has been dominated by the *von Bertalanffy* growth model, but evidence shows that relying on a single predefined equation is often too restrictive for environmentally variable aquaculture systems. A large comparative analysis found that the *von Bertalanffy* model was not the best-fitting model in 65.4% of cases, and recommended multi-model inference as a more robust way to handle model selection uncertainty in fish growth studies. This broader methodological shift is also visible across fisheries science, where growth modeling has moved away from fitting one default curve toward evaluating multiple candidate models and selecting biologically meaningful forms based on the data rather than convention.

In aquaculture experiments, traditional statistical models remain useful when environmental variables are incorporated directly into growth parameters. Multivariate analyses based on the *von Bertalanffy* framework showed that extended Gulland-Holt plots and multilinear regression on the growth parameter K explained 28-46% of variation in observed daily growth rate and were significantly related to Nile tilapia net yield. Even so, conventional formulations have important analytical limitations, because nonlinear *von Bertalanffy* fitting complicates formal comparison across populations, while direct fitting approaches depend on prior knowledge of asymptotic length and can shift variance structure depending on the value assumed.

4.2 Machine learning algorithms including random forest, support vector machine, and XGBoost

Machine learning models are increasingly favored because they can represent nonlinear interactions among water quality, fish size, and time without requiring a fixed growth equation. In rainbow trout growth forecasting, the evaluated machine-learning set included Linear SVR, kernel SVR, Random Forest, XGBoost, LightGBM, and ANN, and the strongest methods achieved 70%-80% error reductions relative to baseline growth models when environmental context was included (Fulton and Lyu, 2026). The best overall performance came from a stacked ensemble combining XGBoost with the *von Bertalanffy* model, showing that machine learning can outperform purely parametric approaches while still benefiting from mechanistic structure (Fulton and Lyu, 2026).

Within aquaculture decision systems, Random Forest, SVM, and XGBoost offer different operational strengths rather than a single universal optimum. Tree-based models such as Random Forest and XGBoost are effective for capturing threshold-style management rules and complex parameter interactions, while SVM with an RBF kernel performs well when water-quality scenarios are separable by clear boundaries in high-dimensional space. Comparative water-quality modeling further shows that optimized XGBoost can deliver very high predictive performance, including an R^2 of 0.9685 for water quality index regression, whereas in some application-specific datasets linear regression and SVR can match or exceed more complex models, underscoring the need to match algorithm choice to data structure and deployment goals.

4.3 Deep learning approaches for dynamic prediction using time-series water quality data

Deep learning approaches are especially well suited to dynamic prediction because fish growth and water quality both evolve as temporally dependent processes. In aquaculture water-quality prediction, LSTM, GRU, and CNN models are widely used because they capture nonlinear behavior and long-term temporal dependencies more effectively than conventional recurrent networks (Rasheed et al., 2022). Their practical value is clear in smart mariculture, where deep LSTM models built from preprocessed multivariate sensor data and correlation information reached short-term prediction accuracies of 98.56% for pH and 98.97% for water temperature, while maintaining strong long-term performance as well.

More recent model development has shifted from predicting isolated water variables to forecasting biological growth trajectories directly from continuous sensor streams. A Temporal Convolutional Network framework for multivariate, multi-horizon fish growth forecasting used sliding windows, multi-resolution resampling, and task-specific output heads to predict fish length and weight in real time, and it outperformed LSTM and GRU by 30%-50% in RMSE and MAE across forecasting windows (Saleh et al., 2026). Hybrid deep architectures also remain competitive for supporting this pipeline, as CNN-LSTM models improved both prediction accuracy and computation time over baseline LSTM, GRU, CNN, and attention-based recurrent models in aquaculture water-quality datasets, suggesting that convolutional-recurrent integration is particularly useful when growth prediction depends on upstream time-series estimation of water conditions (Rasheed et al., 2022).

5 Model Evaluation and Interpretation of Water Quality-Growth Relationships

5.1 Performance evaluation using prediction accuracy indicators

Prediction performance in fish growth modeling is typically assessed with complementary error and fit metrics rather than with a single indicator. In time-series aquaculture forecasting, RMSE, MAE, MAPE, AIC, and BIC are used together to quantify both deviation from observed values and overall model adequacy, and the best model is defined by the smallest error criterion values on validation data (Siddique et al., 2025). This multi-metric logic is consistent with fish growth studies that compare RMSE, MAE, R^2, adjusted R^2, and information criteria side by side, because models with similar absolute errors can still differ in parsimony, stability, and practical usefulness (Fulton and Lyu, 2026).

Recent fish growth models show that high predictive accuracy is achievable when dynamic water-quality information is integrated effectively. In a recirculating aquaculture system, the physics-informed L-PIGRU model reached an RMSE of 2.39, an MAE of 2.09, and a MAPE of 0.68%, indicating very low average prediction error under controlled deployment conditions (Zhang et al., 2025). In multivariate multi-horizon pond forecasting, a TCN architecture outperformed LSTM and GRU by 30%-50% in RMSE and MAE across temporal resolutions and forecast windows, showing that time-series structure strongly affects evaluation outcomes (Saleh et al., 2026).

Traditional and hybrid baselines remain important because they help determine whether added algorithmic complexity yields a meaningful gain. A biologically informed *von Bertalanffy* model explained over 96% of the variance in rainbow trout length, with RMSE 16.82 and MAE 11.60, providing a strong benchmark for newer approaches (Fulton and Lyu, 2026). Yet model comparison also shows that advanced methods can improve robustness: Bayesian deep kernel machine regression achieved substantially lower prediction errors than kernel-based alternatives in olive flounder, supporting the value of probabilistic nonlinear models when environmental covariates are complex and noisy (Kim et al., 2025).

5.2 Identification of key environmental drivers affecting fish growth through feature importance analysis

Feature importance analysis consistently identifies a small core of environmental variables as the main drivers of fish or aquatic livestock performance. In freshwater pond production, dissolved oxygen, salinity, and temperature had the greatest influence on harvest performance, and their timing also mattered: oxygen and salinity variation late in the grow-out period and temperature variation soon after stocking most strongly separated high- and low-performing ponds (Rana et al., 2021). Time-series tilapia modeling similarly showed that ARIMAX gains explanatory power by incorporating climatic and water-quality covariates, with percent weight gain strongly influenced by water temperature and solar intensity (Siddique et al., 2025).

Feature rankings also show that environmental drivers must be interpreted alongside biological and temporal variables. In rainbow trout, Random Forest identified length at release, time at large, and weight at release as the three most important predictors, demonstrating that baseline fish condition and elapsed growth opportunity can outweigh any single environmental measurement (Fulton and Lyu, 2026). In cross-system modeling, species- and system-specific covariates remain necessary, because salinity, photoperiod, or stocking density can become critical predictors depending on farming context and should be tailored during model construction and re-validation (Kim et al., 2025).

Lower-ranked but highly relevant evidence broadens this picture by showing that feature importance can uncover both conventional and overlooked drivers. In Atlantic cod, Random Forest combined with SHAP identified dissolved oxygen, salinity, and temperature as important predictors, but also detected negative growth contributions from zinc, lead, and chromium, indicating that chemical pollutants can enter the same predictive hierarchy as standard water-quality parameters (Muta et al., 2025). In aquaculture water-quality classification, SHAP and DiCE likewise highlighted ammonia, nitrite, dissolved oxygen, turbidity, biochemical oxygen demand, temperature, pH, and CO_2 as the most influential indicators, reinforcing the value of ranked environmental variables for streamlined monitoring design (Naim et al., 2025)

5.3 Application of explainable artificial intelligence for interpreting prediction mechanisms

Explainable artificial intelligence is increasingly used to interpret how prediction models connect water-quality inputs to biological outcomes. SHAP is a game-theoretic method that treats each predictor as a contributor to the model output and can provide both global and local explanations, whereas LIME explains single instances by fitting a local surrogate model. This distinction matters in aquaculture applications because global interpretation is useful for identifying generally important environmental drivers, while local interpretation is better suited to understanding why a specific pond, tank, or prediction deviates from expectation.

Applied studies show that XAI can improve both interpretation quality and monitoring efficiency. In ensemble water-quality prediction, SHAP-based input ranking produced the most stable model performance among candidate variable-selection strategies, and the resulting explanations were detailed further with summary plots, force plots, target plots, and partial dependence plots. That work also showed that SHAP-informed variable selection can guide on-site monitoring toward the most informative measurements, reducing the cost of data collection while preserving predictive strength (Park et al., 2022).

XAI is most useful when it links global feature ranking to sample-level decision pathways, but its outputs must be interpreted cautiously. SHAP decomposition can trace cumulative feature contributions from a baseline expectation to an individual prediction and reveal anomalous or context-specific pathways, which makes it suitable for explaining nonlinear ecological models at both system and sample levels (Yang et al., 2025). However, SHAP and LIME are both model-dependent and sensitive to feature collinearity, so explanation results should be presented with clear assumptions and, where possible, compared across models to test robustness rather than treated as direct causal evidence.

5 Model Evaluation and Interpretation of Water Quality-Growth Relationships

5.1 Performance evaluation using prediction accuracy indicators

Prediction performance in fish growth modeling is typically assessed with complementary error and fit metrics rather than with a single indicator. In time-series aquaculture forecasting, RMSE, MAE, MAPE, AIC, and BIC are used together to quantify both deviation from observed values and overall model adequacy, and the best model is defined by the smallest error criterion values on validation data (Siddique et al., 2025). This multi-metric logic is consistent with fish growth studies that compare RMSE, MAE, R^2, adjusted R^2, and information criteria side by side, because models with similar absolute errors can still differ in parsimony, stability, and practical usefulness (Fulton and Lyu, 2026).

Recent fish growth models show that high predictive accuracy is achievable when dynamic water-quality information is integrated effectively. In a recirculating aquaculture system, the physics-informed L-PIGRU model

reached an RMSE of 2.39, an MAE of 2.09, and a MAPE of 0.68%, indicating very low average prediction error under controlled deployment conditions (Zhang et al., 2025). In multivariate multi-horizon pond forecasting, a TCN architecture outperformed LSTM and GRU by 30-50% in RMSE and MAE across temporal resolutions and forecast windows, showing that time-series structure strongly affects evaluation outcomes (Saleh et al., 2026).

Traditional and hybrid baselines remain important because they help determine whether added algorithmic complexity yields a meaningful gain. A biologically informed *von Bertalanffy* model explained over 96% of the variance in rainbow trout length, with RMSE 16.82 and MAE 11.60, providing a strong benchmark for newer approaches (Fulton and Lyu, 2026). Yet model comparison also shows that advanced methods can improve robustness: Bayesian deep kernel machine regression achieved substantially lower prediction errors than kernel-based alternatives in olive flounder, supporting the value of probabilistic nonlinear models when environmental covariates are complex and noisy (Kim et al., 2025).

5.2 Identification of key environmental drivers affecting fish growth through feature importance analysis

Feature importance analysis consistently identifies a small core of environmental variables as the main drivers of fish or aquatic livestock performance. In freshwater pond production, dissolved oxygen, salinity, and temperature had the greatest influence on harvest performance, and their timing also mattered: oxygen and salinity variation late in the grow-out period and temperature variation soon after stocking most strongly separated high- and low-performing ponds (Rana et al., 2021). Time-series tilapia modeling similarly showed that ARIMAX gains explanatory power by incorporating climatic and water-quality covariates, with percent weight gain strongly influenced by water temperature and solar intensity (Siddique et al., 2025).

Feature rankings also show that environmental drivers must be interpreted alongside biological and temporal variables. In rainbow trout, Random Forest identified length at release, time at large, and weight at release as the three most important predictors, demonstrating that baseline fish condition and elapsed growth opportunity can outweigh any single environmental measurement (Fulton and Lyu, 2026). In cross-system modeling, species- and system-specific covariates remain necessary, because salinity, photoperiod, or stocking density can become critical predictors depending on farming context and should be tailored during model construction and re-validation (Kim et al., 2025).

Lower-ranked but highly relevant evidence broadens this picture by showing that feature importance can uncover both conventional and overlooked drivers. In Atlantic cod, Random Forest combined with SHAP identified dissolved oxygen, salinity, and temperature as important predictors, but also detected negative growth contributions from zinc, lead, and chromium, indicating that chemical pollutants can enter the same predictive hierarchy as standard water-quality parameters (Muta et al., 2025). In aquaculture water-quality classification, SHAP and DiCE likewise highlighted ammonia, nitrite, dissolved oxygen, turbidity, biochemical oxygen demand, temperature, pH, and CO_2 as the most influential indicators, reinforcing the value of ranked environmental variables for streamlined monitoring design (Naim et al., 2025).

5.3 Application of explainable artificial intelligence for interpreting prediction mechanisms

Explainable artificial intelligence is increasingly used to interpret how prediction models connect water-quality inputs to biological outcomes. SHAP is a game-theoretic method that treats each predictor as a contributor to the model output and can provide both global and local explanations, whereas LIME explains single instances by fitting a local surrogate model. This distinction matters in aquaculture applications because global interpretation is useful for identifying generally important environmental drivers, while local interpretation is better suited to understanding why a specific pond, tank, or prediction deviates from expectation.

Applied studies show that XAI can improve both interpretation quality and monitoring efficiency. In ensemble water-quality prediction, SHAP-based input ranking produced the most stable model performance among candidate variable-selection strategies, and the resulting explanations were detailed further with summary plots, force plots, target plots, and partial dependence plots. That work also showed that SHAP-informed variable

selection can guide on-site monitoring toward the most informative measurements, reducing the cost of data collection while preserving predictive strength (Park et al., 2022).

XAI is most useful when it links global feature ranking to sample-level decision pathways, but its outputs must be interpreted cautiously. SHAP decomposition can trace cumulative feature contributions from a baseline expectation to an individual prediction and reveal anomalous or context-specific pathways, which makes it suitable for explaining nonlinear ecological models at both system and sample levels (Yang et al., 2025). However, SHAP and LIME are both model-dependent and sensitive to feature collinearity, so explanation results should be presented with clear assumptions and, where possible, compared across models to test robustness rather than treated as direct causal evidence.

7 Applications of Data-Driven Growth Prediction in Precision Aquaculture

7.1 Integration of sensor networks, iot platforms, and real-time environmental monitoring

Data-driven growth prediction in precision aquaculture depends first on continuous, real-time acquisition of environmental data through distributed sensor systems. IoT platforms have transformed aquaculture by enabling automated monitoring of pH, temperature, dissolved oxygen, turbidity, and salinity, while also supporting remote data access for rapid operational response (Flores-Iwasaki et al., 2025). This monitoring layer is biologically relevant because the most widely deployed sensors track exactly the variables most closely tied to growth and survival, especially pH, temperature, dissolved oxygen, and, in more advanced systems, nitrogen compounds.

Practical deployment increasingly links sensors to microcontrollers, cloud platforms, and user-facing dashboards so that growth-relevant water conditions can be interpreted in near real time. Typical architectures include sensing, processing, connectivity, and application layers, with platforms such as ThingSpeak and Firebase enabling remote visualization, storage, and system interaction (Flores-Iwasaki et al., 2025). Multi-node monitoring is especially important in pond systems because it captures spatial and temporal heterogeneity that single-point sampling can miss, including oxygen-deficient zones and local temperature stratification that directly affect production performance (Figure 2) (Yavuzcan and Mansour, 2026).

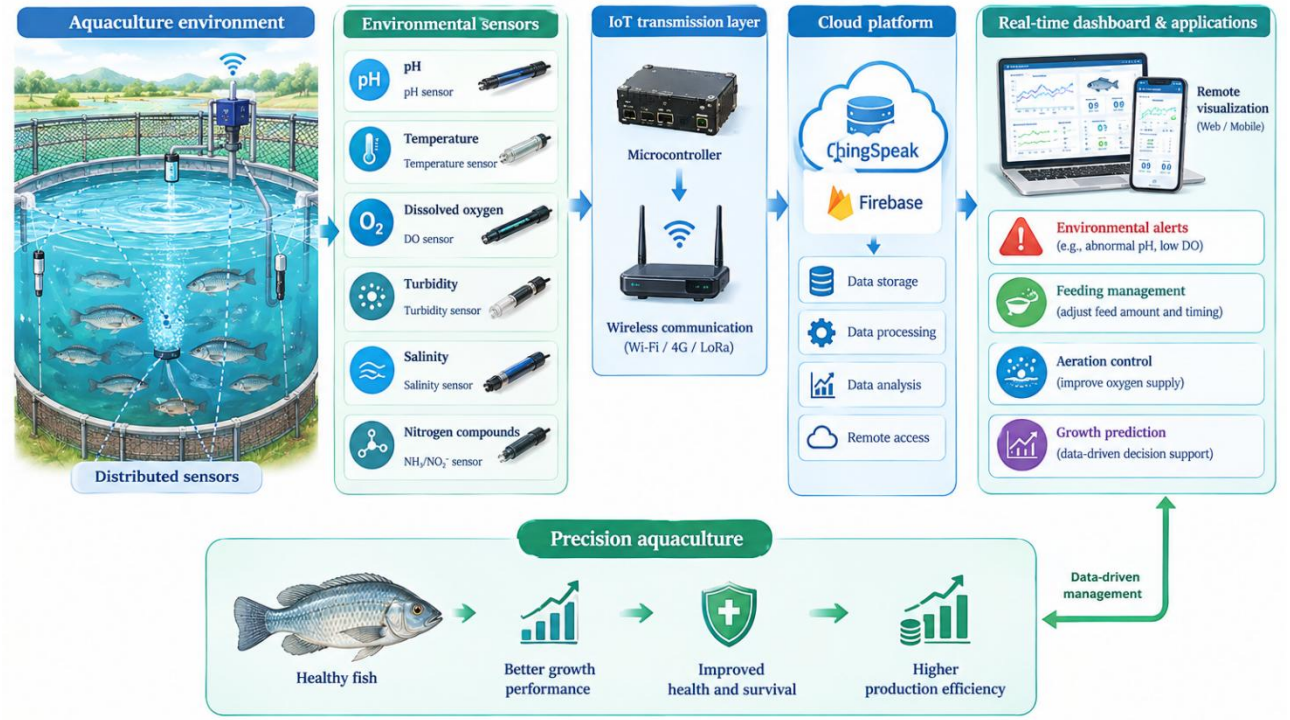


Figure 2 IoT-enabled architecture for real-time environmental monitoring and data-driven management in precision aquaculture

7.2 Decision-support systems for feeding optimization and water quality regulation

Once sensor data are available continuously, predictive models can be embedded in decision-support systems to optimize feeding and maintain water quality within growth-supporting ranges. Feeding is a dynamic control variable influenced by fish size, temperature, dissolved oxygen, and stocking density, so effective management requires continuous adjustment rather than fixed schedules (Elmessery et al., 2025). This is why decision-support systems and digital twins are becoming central in aquaculture: they convert sensor and model outputs into actionable recommendations for forecasting, scenario testing, and operational optimization in domains such as feeding and water-quality control (Shubairi et al., 2026).

Applied studies show that intelligent control systems can improve both biological and economic outcomes when feeding and environmental regulation are optimized together. In recirculating systems, reinforcement-learning control simultaneously managed feeding and water quality, improving feed conversion by 15.5%, maintaining water quality within target ranges 96.8% of the time, and reducing operational costs by 31.5% (Elmessery et al., 2025). Related control work also found that model predictive control reduced mortality and food consumption under different ammonia exposure conditions, underscoring that growth prediction is most useful when coupled to active intervention rather than passive observation.

7.3 Future development of intelligent aquaculture management based on artificial intelligence

The next stage of precision aquaculture is moving from monitoring systems toward integrated AI management frameworks that combine prediction, automation, and adaptive control. Reviews of AI in aquaculture describe this transition as a broader digital transformation in which machine learning, deep learning, and hybrid AI methods improve water-quality prediction, feed optimization, disease detection, and behavioral analysis at the farm level (Yang et al., 2025; Ratan et al., 2026). At the system level, this evolution aligns with the longer-term shift from experience-based fish farming toward knowledge-based production regimes built on sensing, modeling, and automated decision support.

Future development is likely to center on digital twins, edge computing, explainable AI, and multimodal AIoT platforms, but adoption still depends on solving persistent infrastructure and validation challenges. Current reviews identify digital twins as a major route for synchronized virtual testing and predictive control, while also emphasizing the need for continuous calibration, cross-farm generalization, and stronger data standardization (Shubairi et al., 2026). Emerging frameworks also point to federated learning, edge AI, blockchain, and expert-knowledge-integrated systems as scalable next-generation tools, although barriers in sensor reliability, system integration, rural connectivity, and implementation cost remain substantial (Ratan et al., 2026).

8 Conclusions and Future Perspectives

A major advance has been the transition from static, single-target growth estimation to dynamic forecasting frameworks that learn from continuous environmental data. Multivariate deep-learning architectures now predict fish length and weight across multiple future horizons from IoT sensor streams, and Temporal Convolutional Networks have outperformed recurrent baselines by 30%-50% in RMSE and MAE while maintaining strong cross-pond generalization. In parallel, hybrid approaches have improved interpretability by combining data-driven learning with biological constraints, as shown by the L-PIGRU model, which explicitly links fish growth to temperature, dissolved oxygen, pH, and feeding while achieving low prediction error in recirculating systems.

Another important advance is the broader integration of AI methods into aquaculture monitoring and prediction pipelines. Machine learning is now used not only for growth estimation but also for biomass assessment, fish identification, behavior analysis, and water-quality prediction, making growth forecasting part of a wider digital fishery framework. This expansion is reinforced by evidence that AI can extract actionable patterns from large volumes of farm data, allowing farmers to identify growth trends, feeding responses, and environmental stress signals earlier than with manual observation alone.

These predictive advances have practical value because they support earlier and more precise intervention in feeding, water regulation, and risk control. Digital technologies in aquaculture are being adopted specifically to

improve operational efficiency, environmental performance, and risk management, while decision-support systems translate sensor and model outputs into recommendations for water quality, feeding, and health monitoring. At farm level, AI-based monitoring systems can continuously track water quality and fish condition, issue timely alerts, and support proactive actions such as adjusting feeding schedules or modifying water parameters before losses accumulate. The sustainability implications are broader than prediction accuracy alone. Sustainable aquaculture depends on production systems that optimize efficiency while protecting coastal and inland ecosystems, and integrated technology-based management can reduce waste, improve nutrient use, and conserve water and space. At the sector scale, improved water-quality forecasting and disease prevention are expected to optimize economic and environmental resources, while better environmental control is also central to avoiding negative impacts from poor water management and unsustainable exploitation.

Future progress will depend on integrating heterogeneous data streams into unified predictive systems rather than treating sensors, models, and controls as separate tools. Effective digital twins require the seamless coupling of data, models, and control, with multimodal inputs that include sensor streams, physics-based models, domain knowledge, and user interaction within one computational framework. In aquaculture specifically, digital twin development is still early, but the main enabling components already exist, including sensor technologies, knowledge-based models, and ongoing efforts to integrate models across biological and engineering domains. The next stage is likely to move from decision support toward partially autonomous and eventually agentic aquaculture systems. Digital twin platforms can already simulate future environmental conditions from historical and real-time data to anticipate temperature or salinity shifts that affect growth, enabling managers to act before stress and yield loss occur. More advanced control architectures are emerging as well: a recent RAG-LLM plus DQN framework achieved 1.8% higher fish growth than expert-led management with under 2% decision error, suggesting that autonomous, AI-guided control is becoming technically feasible, although data integration, standardization, validation, and interoperability remain major barriers to reliable large-scale adoption.

Acknowledgments

I extend my sincere gratitude to the anonymous reviewers for their valuable and insightful comments, which have greatly strengthened this paper.

Conflict of Interest Disclosure

The author affirms that this research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Abd El-Hack M.E., El-Saadony M.T., Nader M.M., Salem H.M., El-Tahan A.M., Soliman S.M., and Khafaga A.F., 2022, Effect of environmental factors on growth performance of Nile tilapia (*Oreochromis niloticus*), International Journal of Biometeorology, 66(11): 2183-2194. https://doi.org/10.1007/s00484-022-02347-6

Canosa L.F., and Bertucci J.I., 2023, The effect of environmental stressors on growth in fish and its endocrine control, Frontiers in Endocrinology, 14: 1109461. https://doi.org/10.3389/fendo.2023.1109461

Elmessery W.M., Abdallah S.E., Oraiath A.A.T., Espinosa V., Abuhussein M.F.A., Szűcs P., Eid M.H., El-Shinawy D.M., Aljumayi H., Mahmoud S.F., and Elwakeel A.E., 2025, A deep deterministic policy gradient approach for optimizing feeding rates and water quality management in recirculating aquaculture systems, Aquaculture International, 33(4): 253. https://doi.org/10.1007/s10499-025-01914-z

Eneh A.H., Udanor C.N., Ossai N.I., Aneke S.O., Ugwoke P.O., Obayi A.A., Ugwuishiwu C.H., and Okereke G.E., 2023, Towards an improved internet of things sensors data quality for a smart aquaponics system yield prediction, MethodsX, 11: 102436. https://doi.org/10.1016/j.mex.2023.102436

Flores-Iwasaki M., Guadalupe G.A., Pachas-Caycho M., Chapa-Gonza S., Mori-Zabarburú R.C., and Guerrero-Abad J.C., 2025, Internet of things (IoT) sensors for water quality monitoring in aquaculture systems: a systematic review and bibliometric analysis, AgriEngineering, 7(3): 78. https://doi.org/10.3390/agriengineering7030078

Fulton L.V., and Lyu P., 2026, Integrating biological and machine learning models for rainbow trout growth: Balancing accuracy and interpretability, PLoS One, 21(3): e0336890. https://doi.org/10.1371/journal.pone.0336890

Islam M.J., Kunzmann A., and Slater M.J., 2022, Responses of aquaculture fish to climate change-induced extreme temperatures: A review, Journal of the World Aquaculture Society, 53(2): 314-366.
https://doi.org/10.1111/jwas.12853

Saleh H., AlMohimeed A., Abumalloh R.A., Verma P., Rowan N.J., McCann M., Breslin J.G., and El-Sappagh S., 2026, Multivariate multi-horizon fish growth forecasting using temporal convolutional networks in IoT-enabled aquaculture, Smart Agricultural Technology, 2026: 102189.
https://doi.org/10.1016/j.atech.2026.102189

Seale A.P., Cao K., Chang R.J.A., Goodearly T.R., Malintha G.H.T., Merlo R.S., Peterson T.L., and Reighard J.R., 2024, Salinity tolerance of fishes: experimental approaches and implications for aquaculture production, Reviews in Aquaculture, 16(3): 1351-1373.
https://doi.org/10.1111/raq.12900

Shubairi M.N., Furtado P., and Carvalho R.F., 2026, Decision support systems for digital twins in aquaculture: A systematic literature review, Aquacultural Engineering, 2024: 102728.
https://doi.org/10.1016/j.aquaeng.2026.102728

Siddique M.A.B., Mahalder B., Haque M.M., and Ahammad A., 2025, Impact of climatic and water quality parameters on Tilapia (*Oreochromis niloticus*) broodfish growth: Integrating ARIMA and ARIMAX for precise modeling and forecasting, PLoS One, 20(3): e0313846.
https://doi.org/10.1371/journal.pone.0313846

Udanor C.N., Ossai N.I., Nweke E.O., Ogbuokiri B.O., Eneh A.H., Ugwuishiwu C.H., Aneke S.O., Ezuwgu A.O., Ugwoke P.O., and Arua C., 2022, An internet of things labelled dataset for aquaponics fish pond water quality monitoring system, Data in Brief, 43: 108400.
https://doi.org/10.1016/j.dib.2022.108400

Yang L., Zhou W., Zhang C., and Tang F., 2025, A comparative machine learning study identifies Light Gradient Boosting Machine (LightGBM) as the optimal model for unveiling the environmental drivers of yellowfin tuna (Thunnus albacares) distribution using SHapley Additive exPlanations (SHAP) analysis, Biology, 14(11): 1567.
https://doi.org/10.3390/biology14111567

Yavuzcan H., and Mansour A.T., 2026, Technological innovations are driving precision aquaculture towards a future of transparent, automated, and sustainably efficient production, Animal Reports, 2(1): 16-34.
https://doi.org/10.64636/ar.37

Zhang D., Xie Z., Deng W., Lu E., Zhang S., and Gao H., 2025, A physics-informed hybrid neural network for interpretable and accurate fish growth prediction in recirculating aquaculture systems, Computers and Electronics in Agriculture, 239: 110914.
https://doi.org/10.1016/j.compag.2025.110914

Zhang K., Ye Z., Qi M., Cai W., Saraiva J.L., Wen Y., Liu G., Zhu Z., Zhu S., and Zhao J., 2025, Water quality impact on fish behavior: A review from an aquaculture perspective, Reviews in Aquaculture, 17(1): e12985.
https://doi.org/10.1111/raq.12985

Disclaimer/Publisher's Image caption

The statements, opinions, and data contained in all publications are solely those of the individual authors and contributors and do not represent the views of the publishing house and/or its editors. The publisher and/or its editors disclaim all responsibility for any harm or damage to persons or property that may result from the application of ideas, methods, instructions, or products discussed in the content. Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.