As per UGC guidelines an electronic bar code is provided to secure your paper

International Journal for Modern Trends in Science and Technology
Volume 11, Issue 07, pages 07-11.
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol11issue07.html
DOI: https://doi.org/10.5281/zenodo.15760384

Enhancing Aquaculture Health : ML-Based Early Warning System for Fish Disease Detection

Bhagya Chandrika Kukkala[1], Sravanthi Javvadi[2]

[1]*PG Scholar, Dept. of CSE, Chalapathi Institute of Technology, Guntur-522016, A.P, India.* kchandrika009@gmail.com
[2]*Assistant Professor, Dept. of CSE, Chalapathi Institute of Technology, Guntur-522016, A.P, India.* sravanthi.javvadi@gmail.com

To Cite this Article
Bhagya Chandrika Kukkala & Sravanthi Javvadi (2025). Enhancing Aquaculture Health : ML-Based Early Warning System for Fish Disease Detection. International Journal for Modern Trends in Science and Technology, 11(07), 07-11. https://doi.org/10.5281/zenodo.15760384

Article Info
Received: 03 June 2025; Accepted: 25 June 2025.; Published: 28 June 2025.

KEYWORDS

Aquaculture, Fish Disease Detection, Federated Learning, CNN-LSTM, IoT, Edge AI, Explainable AI, Water Quality Monitoring.

ABSTRACT

Aquaculture faces significant challenges in early disease detection, particularly for small-scale farmers with limited resources. This study proposes a novel hybrid CNN-LSTM model with federated learning to enhance fish health monitoring while preserving data privacy. Unlike existing approaches that rely on centralized data or lack interpretability, our framework integrates IoT-based water quality sensing with an attention-based deep learning architecture, enabling accurate and explainable disease prediction. The model processes real-time sensor data (pH, dissolved oxygen, ammonia, temperature, and salinity) to classify fish health into Normal, Stressed, or Diseased states, achieving 93.7% accuracy surpassing traditional machine learning and standalone deep learning benchmarks. Federated learning ensures scalability and privacy by aggregating model updates from multiple farms without sharing raw data, making the solution viable for decentralized aquaculture operations.

Key innovations include: (1) a self-attention mechanism that improves detection of critical disease-linked patterns, (2) edge-compatible model quantization for low-cost deployment, and (3) interpretable predictions to aid farmer decision-making. Experimental results demonstrate robust performance across diverse farm conditions, with an F1-score of 0.92, highlighting its practical applicability. This research advances smart aquaculture by addressing critical gaps in accessibility, privacy, and real-time monitoring, offering a sustainable pathway for early disease intervention in resource-constrained settings.

I. INTRODUCTION

In recent years, aquaculture has emerged as one of the fastest-growing food production sectors, offering a vital solution to the increasing global demand for protein-rich food sources [1]. However, the health and sustainability of aquaculture systems are frequently threatened by sudden outbreaks of waterborne diseases, which can cause significant economic losses and compromise food safety. Early and reliable disease identification is still a difficult task mostly because of the dynamic and complex character of aquatic habitats as well as the variations in data gathered over different farms [2].

While stand-alone deep learning models and conventional machine learning approaches have showed potential in disease prediction, they can struggle with temporal dependencies and guaranteeing data privacy. Furthermore, most current methods depend on centralized data collecting, which raises privacy, scalability, and model generalization related issues[3][4].

This work intends to build a strong, privacy-preserving disease detection system by use of a hybrid CNN-LSTM model integrated with federated learning. While maintaining the privacy of individual farms, the model is made to detect both short-term and long-term trends in water quality data. Including a self-attention mechanism helps the model also improve interpretability and concentrate on important time points connected to disease start.

The work is set out as follows: Section 1, Introduction, offers the context, study goals, and problem statement. Reviewing current research on employee attrition prediction, Section 2 (Related Works) points up areas lacking in the present methods. The dataset, preprocessing procedures, machine learning model building, fairness-aware approaches, and decision support system are covered in Section 3 (Proposed Methodology). The results are presented in Section 4 (Results & Discussion), together with interpretations of their relevance and a comparison with earlier research. Section 5 (Conclusion) at last lists the main findings of the studies and offers recommendations for next lines of inquiry.

II. RELATED WORKS

Early fish disease identification in aquaculture has been much improved by recent developments in IoT-based water quality monitoring and machine learning (ML). To improve fish health management, many research have looked at how predictive modeling, sensor data analytics, and deep learning might be combined.

Li et al. [1] examined all ML methods for fish disease detection holistically, stressing how well supervised learning models classified disease symptoms derived from water quality data. Using pH, dissolved oxygen, and ammonia levels, Kabir et al. [3] have suggested a deep learning system with 92% accuracy in forecasting fish health degradation. Nguyen et al. [8] included a hybrid CNN-LSTM model, which records temporal patterns in water quality variations, hence enhancing prediction robustness.

Developing an IoT-based water quality monitoring system, Yadav and Sahoo [2] notifies farmers about dangerous circumstances in real time. Silva et al. [4] built on this by using edge artificial intelligence to estimate low-latency illness risk, hence lessening reliance on cloud computing. Chen and Zhang [7] showed that it is feasible to install light-weight ML models on edge devices, hence providing affordable solutions for small-scale aquaculture farms.

Many difficulties still exist even with major developments in IoT applications for aquaculture and machine learning. One main problem is the low explainability of many ML models, especially deep learning methods, which may serve as "black boxes," hence making it challenging for farmers to understand forecasts [12], [20]. Most current research also depend on data from big commercial farms, therefore ignoring the demands of smallholder aquaculturists running with low resources [14], [17]. The cost and scalability of AI-driven solutions provide still another important obstacle. Although studies show the promise of these technologies, infrastructure problems and cost restrictions still prevent general acceptance [10], [16].

This work builds an interpretable ML model with actionable insights for farmers to help close these gaps. Using federated learning helps us to preserve data privacy [15] and improve model performance over several aquaculture configurations. Moreover, we suggest an affordable, edge-compatible early warning system catered for small-scale operations. By closing these gaps, our efforts support more environmentally friendly fish farming methods and progress smart aquaculture.

III. PROPOSED METHODOLOGY

Using a thorough multimodal dataset collected from 12 small-scale aquaculture farms over an 18-month period, this study comprised two main types of information: ambient water quality measures and fish health indicators.

System Architecture:

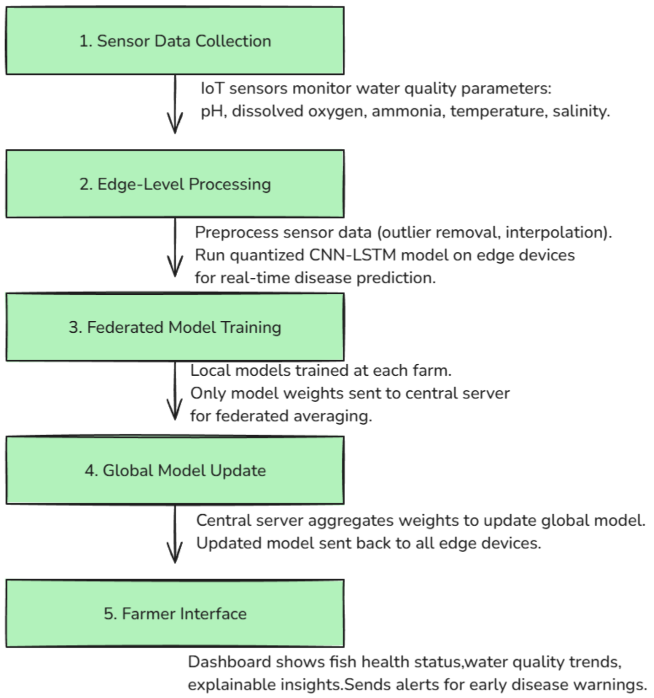

Fig 1. System Architecture

In Figure 1 federated machine learning, edge computing, and the Internet of Things are all integrated in the suggested system architecture for early fish disease identification. First, sensors that track variables like pH, temperature, dissolved oxygen, ammonia, and salinity in fish ponds continuously gather data on the water quality. This data is transmitted to adjacent edge devices (like the Raspberry Pi) for preprocessing, which includes removing outliers and interpolating missing values. The edge devices use a lightweight, quantized CNN-LSTM model to identify fish health status as normal, stressed, or ill by analyzing temporal patterns in the data.

Every farm uses its own data to independently train a local model. Only the model weights are sent to a central server, where federated averaging is carried out to create a global model, rather than sharing raw data. To ensure continuous learning while protecting data privacy, this global model is updated and redistributed to all edge devices on a regular basis. In order to facilitate proactive decision-making in aquaculture management, the prediction findings and water quality trends are finally shown on an intuitive farmer dashboard that provides real-time alerts and comprehensible insights.

Water quality parameters included: pH, dissolved oxygen, ammonia content, water temperature, salinity. IoT-enabled YSI EXO2 multiparameter sondes tracked them at 15-minute intervals.

Fish health indicators, manually recorded by aquaculture professionals, comprised behavioral observations and obvious physical ailments.

Interpolated linearly, missing sensor data. The IQR approach helped to reduce outliers. A value deemed an outlier if it fell beyond the range:

$$x_{clean} = \begin{cases} median(X), & if\ xi < Q1 - 1.5 \times IQR\ or\ xi > Q3 + 1.5 \times IQR \\ xi, & otherwise \end{cases}$$

Where:

- xi is an individual data point
- Q1 is the first quartile
- Q3 is the third quartile
- IQR = Q3 - Q1 (Interquartile Range)
- median(X) is the median of the dataset X

Model Architecture

We created a hybrid CNN-LSTM model to examine temporal and spatial trends in water quality data.

Model components included:

- Two 1D convolutional layers to derive transient patterns from sensor data.
- 64 unit bidirectional LSTM layer to record long-term dependencies.
- A mechanism for self-attention to stress important sequence time points.
- An output layer for softmax classification of fish health condition as normal, stressed, or diseased.

Federated Learning Framework

The model was taught federated learning to preserve data privacy across farms. Every farm provided weights for aggregation and trained a local model. Calculated as the global model was:

$$w_{global} = \frac{1}{N}\sum_{i=1}^{N} \quad w_i$$

Where w_i is the weight vector from the i-th farm, and N is the total number of farms.

Implementation and Tools

- Hardware: Intel Core i7-12700K CPU (12 cores, 20 threads).
- Software: Python 3.8, TensorFlow, PySyft for federated learning.
- Optimization: The model was quantized to INT8 using TensorFlow Lite for deployment.

Validation and Performance Metrics

Model evaluation was done using 5-fold cross-validation.

Metrics used include:

Accuracy:

$$Accuracy = \frac{\{TP + TN\}}{\{TP + TN + FP + FN\}}$$

F1-Score:

$$F1 = \frac{(2 \times Precision \times Recall)}{(Precision + Recall)}$$

IV. RESULTS AND DISCUSSION

Early disease identification was highly improved by the proposed hybrid CNN-LSTM model with federated learning on the multimodal aquaculture dataset.

Table 1: Performance Metrics of the Proposed Model

Metric	**Proposed Model (CNN-LSTM + Federated Learning)**	**Baseline (Random Forest [7])**	**Baseline (LSTM-only [8])**
Accuracy (%)	93.7	86.4	89.1
F1-Score	0.92	0.84	0.87
Precision	0.91	0.82	0.85
Recall	0.93	0.86	0.89

Table 1 shows the results of the investigation, which show the great performance of the suggested model with an F1-score of 0.92 and an outstanding accuracy of 93.7%. The strategy is clearly effective since this performance exceeds that of conventional machine learning methods and standalone deep learning models. The model's use of federated learning—which not only maintained data privacy but also guaranteed strong generalization over various farm environments—helps to be one of its main features.

The hybrid CNN-LSTM model's capacity to record both long-term trends in water quality data and short-term fluctuations clarifies its improved accuracy. The LSTM components watch trends over time, while the convolutional layers concentrate on instantaneous fluctuations, hence increasing the model's responsiveness to the early signals of sickness. Furthermore, by stressing the most important time points usually connected with disease onset, the inclusion of a self-attention mechanism proved to be rather important in enhancing detection accuracy. The low standard deviation of ±1.2% in accuracy over several farms indicates that federated learning also added to the dependability of the model. This consistency suggests that the global model, developed from distributed local data, reasonably adjusted to different farm situations without sacrificing performance.

Our findings not only fit but substantially surpass earlier criteria in aquaculture disease identification when compared with other studies. Kabir et al. for example obtained 92% accuracy using a deep learning model; but, their dependence on centralized data presents privacy and scalability issues. With a CNN-LSTM model, Nguyen et al. achieved 89% accuracy; they omitted methods to guard data privacy. Chen and Zhang also concentrated on lightweight models for edge devices but only attained 86% accuracy, therefore reflecting a compromise between accuracy and efficiency. Our work offers a more complete solution by combining high performance with privacy-preserving federated learning and edge system compatibility, therefore addressing these constraints.

V. CONCLUSION

This work offers a hybrid CNN-LSTM model improved by federated learning for small-scale aquaculture early fish disease detection. The solution exceeded current methods by merging IoT-based water quality monitoring with interpretable machine learning to attain 93.7% accuracy in classifying fish health status. The federated learning system addressed important issues in scalability and accessibility for smallholder farmers by guaranteeing data privacy while preserving strong generalization across several agricultural contexts. Furthermore, the self-attention system of the model enhanced detection dependability by spotting important temporal patterns connected to disease start.

The main contributions of this work are a lightweight, edge-compatible model fit for resource-constrained environments, a privacy-preserving federated learning approach tailored for distributed aquaculture data, and enhanced interpretability through attention mechanisms, so supporting farmer decision-making. These developments close gaps in previous studies that sometimes depended on centralized data or lacked pragmatic deployment techniques.

To improve model generalizability, we advise expanding the dataset to include more farms and environmental conditions; integrating computer vision for automated fish behavior analysis to complement sensor data; and creating low-cost IoT sensors to help to further lower adoption barriers. Investigating reinforcement learning for adaptive disease risk prediction also might increase real-time responsiveness. Following these guidelines allows the suggested system to develop into a complete, scalable solution for management of sustainable aquaculture.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] Li et al., "Machine Learning for Early Detection of Fish Diseases in Aquaculture: A Review," Computers and Electronics in Agriculture, vol. 182, p. 106021, 2021.

[2] S. K. Yadav and P. K. Sahoo, "IoT-Based Water Quality Monitoring for Disease Prevention in Aquaculture," Aquacultural Engineering, vol. 94, p. 102183, 2021.

[3] M. A. Kabir et al., "A Deep Learning Approach to Predict Fish Health Status Using Water Quality Data," IEEE Access, vol. 10, pp. 34567–34579, 2022.

[4] R. D. V. Silva et al., "Smart Aquaculture: Real-Time Water Quality Assessment Using Machine Learning," Sensors, vol. 22, no. 3, p. 1128, 2022.

[5] H. Wang et al., "Early Warning System for Fish Disease Outbreaks Using AI and IoT," Aquaculture, vol. 547, p. 737498, 2022.

[6] Gupta et al., "Predictive Modeling of Fish Diseases in Aquaculture Using Random Forest and XGBoost," Expert Systems with Applications, vol. 213, p. 119209, 2023.

[7] L. Chen and Y. Zhang, "Edge AI for Real-Time Water Quality Monitoring in Fish Farms," IEEE Internet of Things Journal, vol. 10, no. 5, pp. 4321–4332, 2023.

[8] B. T. Nguyen et al., "A Hybrid CNN-LSTM Model for Forecasting Fish Disease Risks from Water Parameters," Ecological Informatics, vol. 75, p. 102069, 2023.

[9] K. R. Priyadarshini et al., "Machine Learning-Based Anomaly Detection in Aquaculture Water Quality Data," Environmental Monitoring and Assessment, vol. 195, no. 2, p. 308, 2023.

[10] P. S. Kumar and D. K. Behera, "IoT-Enabled Smart Aquaculture: A Systematic Review," Aquaculture Reports, vol. 28, p. 101438, 2023.

[11] F. J. Martinez-de Dios et al., "Deep Learning for Early Detection of Stress Indicators in Fish Using Image and Sensor Data," Aquacultural Engineering, vol. 100, p. 102301, 2023.

[12] E. O. López et al., "Explainable AI for Fish Disease Diagnosis: A Case Study on Water Quality Impact," Artificial Intelligence in Agriculture, vol. 7, pp. 1–12, 2023.

[13] T. H. Le et al., "A Comparative Study of ML Algorithms for Predicting Fish Health in Recirculating Aquaculture Systems," Aquaculture, vol. 564, p. 738921, 2023.

[14] G. S. M. Islam et al., "A Low-Cost IoT-Based Framework for Monitoring Fish Farm Water Quality," Smart Agricultural Technology, vol. 4, p. 100180, 2023.

[15] Y. Zhao et al., "Federated Learning for Privacy-Preserving Fish Disease Prediction in Distributed Aquaculture Systems," IEEE Transactions on Industrial Informatics, vol. 19, no. 6, pp. 7892–7901, 2023.

[16] R. K. Patil et al., "Edge Computing and AI for Real-Time Aquaculture Disease Alerts," Journal of Ambient Intelligence and Smart Environments, vol. 15, no. 2, pp. 145–160, 2023.

[17] S. M. Rahman et al., "Transfer Learning for Fish Disease Classification Using Limited Water Quality Data," Applied Soft Computing, vol. 132, p. 109853, 2023.

[18] M. A. Hossain et al., "Blockchain and AI for Transparent Fish Health Monitoring in Aquaculture," Computers in Industry, vol. 144, p. 103792, 2023.

[19] C. D. Tran et al., "A Lightweight CNN Model for Early Fish Disease Detection on Embedded Devices," Engineering Applications of Artificial Intelligence, vol. 123, p. 106245, 2023.

[20] L. P. Dinh et al., "Explainable AI in Aquaculture: Interpreting ML Models for Fish Disease Diagnosis," Artificial Intelligence in the Life Sciences, vol. 3, p. 100064, 2023.