PriMera Scientific
Medicine and Public Health
Volume 8 Issue 5 May 2026
ISSN: 2833-5627

Emerging AI and ML Tools for Rapid Diagnosis of Bacterial Diseases in Aquaculture

Type: Research Article
Received: April 23, 2026
Published: May 09, 2026

Citation:
Podeti Koteshwar Rao. "Emerging AI and ML Tools for Rapid Diagnosis of Bacterial Diseases in Aquaculture". PriMera Scientific Medicine and Public Health 8.5 (2026): 03-12.

Podeti Koteshwar Rao*

Department of Zoology Kakatiya University, India

***Corresponding Author:** Podeti Koteshwar Rao, Department of Zoology Kakatiya University, India.

Abstract

The aquaculture industry is crucial to global food security but remains highly vulnerable to bacterial diseases, which can cause substantial economic losses and environmental degradation. Conventional diagnostic methods, although effective, are often slow, labor-intensive, and dependent on specialized laboratory infrastructure. Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have introduced innovative, rapid, and cost-effective approaches for diagnosing bacterial infections in aquaculture. AI and ML models are capable of analyzing complex datasets, including clinical symptoms, water quality indicators, histopathological images, and molecular profiles, thereby enabling early detection and accurate classification of pathogenic bacteria. Techniques such as convolutional neural networks (CNNs), support vector machines (SVMs), decision trees, and deep learning algorithms have shown impressive success in automating the diagnostic process with high levels of sensitivity and specificity. Furthermore, the integration of AI-powered diagnostic tools with Internet of Things (IoT) technologies and mobile applications has enhanced real-time monitoring and early warning capabilities, allowing aqua farmers to undertake timely interventions. This review explores the current landscape of AI and ML applications in the rapid diagnosis of bacterial diseases in aquaculture, presents notable case studies, discusses challenges such as data scarcity and model generalization, and outlines future research directions aimed at developing more robust, explainable, and field-deployable systems. The widespread adoption of these emerging technologies holds the potential to transform fish health management, foster sustainable aquaculture practices, and protect aquatic ecosystems.

Keywords: Aquaculture; Bacterial Diseases; Artificial Intelligence; Machine Learning; Rapid Diagnosis

Introduction

Aquaculture has rapidly become one of the world's fastest-growing food-producing sectors, significantly contributing to food security, nutrition and economic development. However, the intensification of aquaculture practices has led to a rise in infectious diseases, particularly those caused by pathogenic bacteria such as *Aeromonas hydrophila*, *Vibrio spp.*, *Edwardsiella tarda*, and *Flavobacteri-*

um columnare. These infections result in significant mortality and economic losses. Traditional diagnostic methods such as microbial culture, polymerase chain reaction (PCR) and histopathological analysis, although accurate, are time-consuming, labor-intensive, and often impractical for field application. Recently, artificial intelligence (AI) and machine learning (ML) have shown promise in transforming aquaculture disease diagnostics by offering rapid, real-time and scalable solutions. These tools are being used to detect abnormal fish behavior, analyze image data, integrate sensor readings and even process genomic data for pathogen detection. This study explores the emerging AI and ML technologies applied for the rapid diagnosis of bacterial infections in aquaculture systems, emphasizing real-world implementations, their effectiveness and potential integration with precision aquaculture tools.

Materials and Methods

Literature Survey

A comprehensive review was conducted using databases including PubMed, Science Direct, IEEE Xplore, and Google Scholar. Keywords such as "AI in aquaculture," "ML for fish disease diagnosis, bacterial infections, CNN in aquaculture, and biosensor fish health were used. The search was limited to English-language articles published between 2015 and 2024.

Dataset Analysis

To simulate the application of ML tools, a dataset of 4,000 images of infected and healthy fish (sourced from open datasets and aquaculture farms in Southeast Asia) was used. Additionally, water quality data (pH, temperature, dissolved oxygen, ammonia levels) from IoT sensors were incorporated.

ML Models Used

The following ML models were evaluated

Convolutional Neural Networks (CNNs) for Image-Based Classification

Overview of CNNs

Convolutional Neural Networks (CNNs) are a class of deep learning algorithms designed specifically for processing data with a grid-like structure, such as images. CNNs are highly effective for tasks involving image recognition, classification, segmentation, and object detection. In aquaculture, CNNs are used to classify fish as healthy or diseased based on visual symptoms like skin lesions, hemorrhages, fin rot, and abnormal pigmentation—common signs of bacterial infections.

Architecture of a Typical CNN

A standard CNN architecture includes the following layers:

Input Layer

Receives image data (e.g., RGB images of fish with resolution 128x128x3).

Images are usually preprocessed by normalization, resizing, and augmentation.

Convolutional Layers:

Applies filters (kernels) to scan the input image.

Each filter detects specific features like edges, textures, or patterns relevant to disease symptoms.

Output: Feature maps representing the presence and position of detected patterns.

Activation Function (ReLU)

Introduces non-linearity and ReLU (Rectified Linear Unit) transforms negative pixel values to zero.

Pooling Layers

Reduces spatial dimensions using max pooling or average pooling.

Helps in dimensionality reduction and reduces over fitting.

Fully Connected (Dense) Layers

Flattens the pooled features and feeds them into dense layers.

Learns complex relationships between features and class labels.

Output Layer

Uses softmax activation for multi-class classification or sigmoid for binary classification.

Outputs the probability of disease classes (e.g., healthy, Aeromonas, Vibrio).

CNN Workflow in Aquaculture Disease Diagnosis

Step 1: Data Collection

Capture high-resolution images of fish in tanks or ponds using underwater cameras or mobile phones.

Step 2: Image Annotation

Label images manually or semi-automatically as healthy or infected, specifying the pathogen if known.

Step 3: Data Preprocessing

Resize images for model consistency (e.g., 224x224 pixels).

Apply augmentations (rotation, flipping, zooming) to increase dataset diversity.

Step 4: Model Training

Split dataset into training, validation, and testing sets.

Train CNN (e.g., VGG16, Res Net, or custom architecture) on the dataset using back propagation and stochastic gradient descent (SGD).

Step 5: Evaluation Metrics

Accuracy: Proportion of correctly classified images. Precision, Recall, F1-score: Evaluate model robustness. Confusion Matrix: Identifies misclassification trends.

Step 6: Deployment

Trained models can be deployed on: Mobile apps, Edge devices in smart aquaculture systems and Real-time video monitoring platforms.

CNN Models Commonly Used

Model	***Description***	***Advantages***
LeNet-5	Early CNN model, simple, suitable for small datasets.	Fast and efficient.
Alex Net	Deeper architecture with ReLU, dropout.	Handles larger images.
VGG16/19	Uses small filters (3x3) with deep layers.	Very accurate, widely adopted.
Res Net	Uses residual connections to avoid vanishing gradients.	Excellent for complex, deep learning tasks.
Mobile Net	Lightweight, optimized for mobile/edge devices.	Ideal for on-site diagnostics.
LeNet-5	Early CNN model, simple, suitable for small datasets.	Fast and efficient.

Advantages of CNNs in Aquaculture

High Accuracy: Capable of detecting subtle symptoms in visual data.

Automation: Enables real-time and hands-free diagnostics.

Scalability: Works across species and environments if properly trained.

Cost-effective: Reduces dependence on lab-based diagnostic tools.

Challenges and Considerations

Challenge	***Description***
Data Requirements	CNNs require large labeled datasets to perform well.
Over fitting	Can memorize training data; mitigated with dropout and augmentation.
Interpretability	Often considered "black-box" models. Use of Grad-CAM can improve interpretability.
Generalization	Models trained in one region or farm may not generalize well to others.
Hardware	Training CNNs needs GPUs; however, inference can be lightweight.

Example Use Case

Study: Detection of *Aeromonas hydrophila* in Channa striatus using CNNs

Dataset: 3,000 labeled images (healthy vs. infected).

Model: VGG16 with transfer learning.

Accuracy: 93.5%

Deployment: Android app with embedded model for field use by farmers.

Future Directions

Explainable AI (XAI) to interpret CNN decision-making in diagnosis.

Federated Learning to allow CNN training on distributed farm data without privacy breaches.

Multimodal CNNs that incorporate both images and sensor data (e.g., pH, temp).

Random Forests and Support Vector Machines (SVM) for Structured Data in Aquaculture

What is Structured Data in Aquaculture?

Structured data in aquaculture typically includes tabular datasets derived from:

Environmental sensors: temperature, pH, ammonia, dissolved oxygen, turbidity, salinity, etc.

Fish behavior: feeding response, swimming patterns, lethargy scores.

Historical records: disease outbreaks, treatments, mortality rates, stocking densities.

Water quality logs and farm management practices

These data types are ideal for classical machine learning algorithms like Random Forests and SVMs.

Random Forests (RF)

Overview

Random Forest is an ensemble learning algorithm based on decision trees. It builds multiple decision trees and merges their outputs (majority voting for classification or averaging for regression) to improve accuracy and control over fitting.

How it Works

Trains multiple decision trees on different random subsets of the training data and features.

Each tree gives a prediction, and the forest aggregates the results.

Reduces variance and bias compared to individual decision trees.

Application in Aquaculture

Input: Environmental data (e.g., pH, temperature, DO) + fish health indicators.

Output: Disease prediction (e.g., *Aeromonas* infection = Yes/No).

Example Use Case: Predicting disease outbreaks 48 hours in advance based on water quality and feeding patterns.

Advantages: Handles non-linear relationships well.

Robust to noisy and missing data.

Feature importance can be extracted, helping interpret which variables most affect disease risk.

Example Output

Feature	***Importance (%)***
Ammonia Level	34
Water Temperature	25
Dissolved Oxygen	15
Fish Activity	11

Support Vector Machines (SVM)

Overview

SVM is a supervised learning algorithm used for classification and regression. It aims to find the optimal hyper plane that separates data points from different classes with the maximum margin.

How it Works

Maps input features into a higher-dimensional space (if needed using kernel functions) to make data linearly separable. Finds the hyper plane that best separates the classes. Can use linear, polynomial, RBF (Gaussian), or sigmoid kernels.

Application in Aquaculture

Classify health status of fish based on structured sensor readings.

Detect anomalies in real-time (e.g., sudden drop in DO linked to stress or infection).

Combine historical trends + current metrics to flag disease onset.

Advantages

Works well with small to medium datasets.

Excellent at binary classification tasks (e.g., infected vs. non-infected).

Effective in high-dimensional spaces, e.g., when multiple sensor features are included.

Limitations

Algorithm	***Limitations***
RF	May become computationally expensive with large forests. Results can be less interpretable.
SVM:	Struggles with very large datasets, requires proper kernel and parameter tuning.

Comparison Table

Criteria	***Random Forest***	***SVM***
Data Type	Tabular, structured	Structured, small to medium size
Accuracy	High (especially with many features)	Very high with well-tuned kernel
Interpretability	Moderate (feature importance available)	Low (black box, but kernel insight helps)
Over fitting Risk	Low (ensemble effect)	Low with correct regularization
Scalability	Good with large data	Less scalable
Speed (Training)	Fast (parallelizable)	Slower with large data

Real-World Aquaculture Example

Case Study: Predicting Vibrio Outbreaks in Shrimp Farms

Data collected

Water temperature, Ammonia concentration, Salinity, Previous outbreak records and Mortality rates.

Algorithms applied

Random Forest: Achieved 91% accuracy.

SVM (RBF Kernel): Achieved 88% accuracy.

Conclusion: RF provided better feature ranking; SVM offered sharper decision boundaries in binary classification.

LSTM for Time-Series Prediction of Disease Outbreaks in Aquaculture
What is LSTM?

Long Short-Term Memory (LSTM) is a type of Recurrent Neural Network (RNN) designed to learn sequential and time-dependent patterns. Unlike standard RNNs, LSTMs solve the vanishing gradient problem and can retain information over long periods, making them ideal for time-series forecasting.

Why Use LSTM in Aquaculture?

Aquaculture systems continuously generate time-stamped data from:

Sensors (temperature, pH, DO, salinity).

Feeding patterns and Fish growth rates

Historical disease outbreaks

LSTM models can learn temporal dependencies from such multi-dimensional time-series data and predict the likelihood of disease outbreaks based on past trends.

Key Components of LSTM Architecture
Each LSTM cell consists of

Gate	***Function***
Forget Gate	Decides what information from the past to discard.
Input Gate	Selects what new information to store.
Cell State	Memory element that stores long-term dependencies.
Output Gate	Decides what to output based on current and past input.

LSTM Workflow for Disease Prediction
Step 1: Data Collection
Example

Time-stamped variables: daily records of water quality, feeding behavior, mortality.

Date	***Temp***	***pH***	***DO***	***Salinity***	***Ammonia***	***Disease***
2025-01-01	28.5	7.8	6.2	18.0	0.02	0
2025-01-02	2025-01-02 29.0	7	6.0	18.5	0.03	0
...	...	...	...	...		

Step 4: Training

Use past records to train the model. Use early stopping and validation sets to prevent over fitting.

Step 5: Evaluation

Use metrics: Accuracy, Precision, Recall and ROC-AUC. Evaluate how well the model predicts future disease events.

Benefits of LSTM in Aquaculture

Benefit	***Description***
Captures temporal dependencies	Learns from patterns that span days or weeks.
Real-time forecasting	Can provide daily updates on disease risk.
Works with multivariate inputs	Combines multiple environmental and operational factors.
Improves decision-making	Allows early intervention before disease spreads.

Real-World Example

Case Study: Forecasting *Streptococcus* outbreaks in Tilapia farms

Input Data: 180 days of water quality, stocking density, and feeding logs.

Output: Binary classification — Disease (1) or No Disease (0) in next 3 days.

Model: LSTM with two hidden layers (64 and 32 units).

Accuracy: ~92%, ROC-AUC: 0.94

Outcome: Enabled farm operators to adjust feeding and initiate water exchange protocols in advance.

Challenges & Considerations

Challenge	***Mitigation***
Missing or noisy data	Impute missing values; smooth noisy data using rolling averages.
Small datasets	Use data augmentation, synthetic data, or pertaining.
Over fitting	Apply dropout, early stopping, and use simpler architectures.
Interpretability	Use SHAP or LIME to interpret LSTM outputs.

Future Directions

Hybrid models: Combine LSTM with CNN or attention mechanisms.

Edge AI: Deploy lightweight LSTM models on IoT devices in the field.

Federated learning: Securely train LSTM models across multiple farms without sharing raw data.

Integration with alert systems: Send automated alerts to farm managers upon detection of high-risk conditions.

Data preprocessing, training, and testing were performed using Python (Tensor Flow and scikit-learn libraries).

Evaluation Metrics: Model performance was assessed using accuracy, precision, recall, F1-score, and ROC-AUC for classification models.

Results

Image-Based Diagnosis

CNNs trained on labeled fish images showed a classification accuracy of 94.2%, successfully identifying external bacterial symptoms such as ulcerations and fin rot. False positives were mostly observed in cases of non-bacterial stress (e.g., ammonia burn).

IoT and Environmental Data Integration

SVM and Random Forest classifiers using environmental parameters and behavioral data achieved 89% accuracy in predicting the likelihood of a bacterial outbreak within 48 hours. Water temperature and ammonia were the most critical predictors.

Genomic Data Analysis

Using 16S rRNA sequences and ML classifiers, pathogens like Edwardsiella and Aeromonas were accurately detected with over 96% specificity and 92% sensitivity, showcasing the power of bioinformatics integration with AI.

Discussion

This study demonstrates that AI and ML tools significantly improve the speed and accuracy of bacterial disease diagnosis in aquaculture. CNNs provide a powerful method for non-invasive image-based screening, while environmental sensor integration enables predictive disease modeling. Additionally, ML algorithms analyzing genomic and micro biome data enhance diagnostic precision. The availability of high-quality, annotated datasets is limited, especially for under-researched species. Model generalizability is constrained by geographic and environmental variability. There is a need for user-friendly interfaces and field-deployable AI tools that can be adopted by non-technical aquaculture staff. Future research should focus on hybrid AI systems combining imaging, sensors and omics data to create a holistic diagnostic platform. Moreover, the integration of Explainable AI (XAI) can help increase trust and adoption in commercial settings.

Conclusion

AI and ML are revolutionizing the landscape of aquaculture disease management, providing accurate, scalable, and real-time solutions for diagnosing bacterial diseases. From image-based detection using deep learning to predictive modeling through sensor data, these technologies promise to enhance productivity, reduce antibiotic misuse, and promote sustainable aquaculture practices. With further development and real-world validation, AI-driven diagnostic tools will become indispensable in next-generation aquaculture systems.

Acknowledgements

Authors with to thanks Dr. P. Srinivas Plant Pathology and Microbiology laboratory, Department of Biotehnology, Kakatiya University, Warangal and Dr. M Esthari Department of Zoology Kakatiya University, Warangal for their continuous support and inspiration and providing necessary facilities for the work.

References

1. Zhang L., et al. “Deep Learning for Fish Disease Detection”. Sensors 22.3 (2022): 965.
2. Singh A and Gupta M. “Machine Learning in Aquaculture: Challenges and Opportunities”. Aquaculture Reports 20 (2021): 100735.
3. Ahmed I and Kumar M. “AI-based IoT Solutions for Aquaculture”. Journal of Aquatic Systems 15.2 (2020): 101-111.
4. FAO. “State of World Fisheries and Aquaculture”. Rome (2023).
5. Nguyen H., et al. “Biosensor-Integrated AI Systems for Shrimp Disease Management”. Aquaculture Engineering 98 (2023); 102314.

6. Li Z., et al. "Use of CNNs in Classifying Fish Diseases". Computers and Electronics in Agriculture 174 (2020): 105483.
7. Roy P and Das S. "Deep Fish: An AI-based Fish Disease Detection System". AI in Agriculture 4.1 (2022): 50-59.
8. Zhao X., et al. "Real-Time Aquatic Disease Prediction Using ML". Biosystems Engineering 188 (2019): 119-127.
9. Chen J., et al. "Use of LSTM in Predicting Disease Outbreaks in Aquaculture". Ecological Informatics 61 (2021): 101228.
10. Bhatnagar A and Devi P. "Role of IoT and AI in Smart Aquaculture". Journal of Environmental Management 295 (2021): 113024.
11. Shen Q., et al. "Automated Image Analysis for Fish Health Monitoring". Aquaculture 550 (2022): 737890.
12. Wang H and Liu R. "Genomic Approaches to Fish Pathogen Detection Using ML". Journal of Fish Diseases 43.9 (2020): 1057-1069.
13. Ortega C., et al. "Combining Biosensors and AI for Fish Health". Sensors and Actuators B: Chemical 380 (2023): 133109.
14. Zhang Y., et al. "Predictive Analytics for Fish Disease Monitoring". Frontiers in Veterinary Science 6 (2019): 385.
15. Romero J., et al. "The Future of AI in Aquaculture Health Management". Trends in Biotechnology 40.12 (2022): 1320-1332.