Early Fish Disease Detection Through Water Quality Monitoring and Machine Learning Techniques

K. Krupa Sagari[1], K. Pavani[2], Pothula Yedukondalu[3]

#1 Assistant Professor in the Department of MCA, SRK Institute of Technology, Vijayawada.

#2 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

Abstract_ Aquaculture fish infections pose a serious threat to the security of nutrition. Because of the lack of infrastructure, it is still difficult to identify diseased fish in aquaculture at an early stage. To stop the disease from spreading, it is imperative to identify affected fish as soon as possible. Since salmon aquaculture accounts for 70% (2.5 million tonnes) of the market and is the fastest-growing food production system worldwide, our goal in this effort is to identify the salmon fish disease in aquaculture. We are able to identify fish that are afflicted by different pathogens because to the combination of perfect image processing and machine learning mechanisms. There are two sections to this work. Image segmentation and pre-processing have been used in the basic section to decrease noise and exaggerate.

Index terms - *— Fish Disease Detection, Aquaculture, Image Processing, Machine Learning, GBM Image Segmentation, Image Preprocessing, Salmon Fish Disease, Disease Classification, Fish Health Monitoring.*

1. INTRODUCTION

Aquaculture has become an important source of food production worldwide, with salmon farming contributing significantly to the global seafood industry. Despite its rapid growth, the aquaculture sector faces major challenges due to diseases that affect fish health, productivity, and overall farm profitability. Timely identification of fish diseases is crucial to prevent large-scale outbreaks and ensure sustainable fish farming practices. Conventional disease diagnosis methods often depend on manual inspection by experts, which can be time-consuming, costly, and difficult to implement in remote aquaculture facilities. To overcome these limitations, this project presents an automated fish disease detection system that integrates image processing and machine learning techniques. The proposed system processes fish images through preprocessing and segmentation methods to improve image quality and isolate important disease-related features. These extracted features are then analyzed using a Support Vector Machine (SVM) classifier to determine whether a fish is healthy or affected by disease. By automating the detection process, the system provides a reliable, efficient, and economical solution for monitoring fish health and supporting early disease management in salmon aquaculture.

2. LITERATURE SURVEY

1. R. B. Wynn, V. A. I. Huvenne, T. P. Le Bas et al. (2014)

Title: Autonomous Underwater Vehicles (AUVs): Their Past, Present and Future Contributions to the Advancement of Marine Geoscience

This paper presents a comprehensive review of Autonomous Underwater Vehicles (AUVs) and their contributions to marine geoscience. The authors discuss the historical development of AUV technology, its current applications in underwater exploration, and future research opportunities. The study highlights the ability of AUVs to collect high-resolution seabed data, perform deep-sea surveys, and operate in environments that are difficult or dangerous for human divers. The paper emphasizes the growing importance of AUVs in ocean mapping, geological investigations, environmental monitoring, and marine resource exploration.

2. M. Dinc and C. Hajiyev (2015)
Title: Integration of Navigation Systems for Autonomous Underwater Vehicles

This research focuses on the integration of multiple navigation systems to improve the positioning accuracy and operational reliability of Autonomous Underwater Vehicles (AUVs). The authors investigate the combination of inertial navigation systems, acoustic positioning systems, and other sensor technologies to overcome the limitations of individual navigation methods. The proposed integrated navigation framework enhances vehicle localization, reduces navigation errors, and improves mission performance in underwater environments where GPS signals are unavailable. Experimental results demonstrate the effectiveness of sensor fusion techniques in achieving accurate and stable underwater navigation.

3. Y. Zhou, S. Cui, Y. Wang, and C. Ai (2015)
Title: Design of Autonomous Underwater Vehicle (AUV) Control Unit

This paper presents the design and implementation of a control unit for Autonomous Underwater Vehicles (AUVs). The proposed control system is responsible for managing navigation, motion control, communication, and sensor integration during underwater operations. The authors describe the hardware and software architecture of the control unit and discuss its ability to support autonomous mission execution. Experimental evaluations show that the designed control unit provides reliable performance, efficient processing, and stable operation under different underwater conditions, making it suitable for marine exploration and monitoring applications.

4. Y. Zhou, S. Cui, Y. Wang, and L. Zhai (2017)
Title: A Refined Attitude Algorithm for AUV Based on IMU

This paper proposes an improved attitude estimation algorithm for Autonomous Underwater Vehicles (AUVs) using data obtained from an Inertial Measurement Unit (IMU). Accurate attitude estimation is essential for maintaining vehicle stability and navigation performance during underwater missions. The proposed algorithm enhances the processing of IMU sensor data to achieve more precise measurements of roll, pitch, and yaw angles. Experimental results indicate that the refined algorithm significantly improves attitude estimation accuracy while reducing sensor noise and measurement errors. The study demonstrates the effectiveness of the approach in supporting reliable navigation and control of AUV systems in complex underwater environments.

3.METHODOLOGY

i) Proposed Work:

The authors propose a machine learning-based approach for predicting fish health by analyzing water quality parameters. Poor water quality, including the presence of harmful toxins, pollutants, or disease-causing microorganisms, can negatively affect fish health, leading to infections, respiratory problems, and even mortality. To address this issue, water samples are collected and analyzed using a predictive machine learning model. The study utilizes a water quality dataset obtained from the Kaggle platform and employs the Gradient Boosting algorithm for classification and prediction. By learning the relationship between water quality indicators and fish health conditions, the model can determine whether the aquatic environment is suitable for healthy fish growth. Experimental results indicate that the Gradient Boosting model achieves a prediction accuracy of over 95% on the test dataset, demonstrating its effectiveness in supporting early detection of unfavorable water conditions and improving aquaculture management..

ii) System Architecture:

The above architecture diagram illustrates the workflow of the Fish Condition Prediction System based on water quality analysis using the Gradient Boosting algorithm. The process begins with the User Interface, where users can upload the water quality dataset, provide test data, view prediction results, and analyze model accuracy. The uploaded raw dataset is then passed to the Preprocess and Normalize Dataset module, where missing values are handled, non-numeric data is converted into numerical format, and all attributes are normalized to improve model performance.

After preprocessing, the Feature Selection module selects the input features (X) and target class label (Y). The dataset is then divided into 80% training data and 20% testing data. The training data is provided to the

Gradient Boosting Algorithm, which learns the relationship between water quality parameters and fish health conditions. Once training is completed, a trained model is generated and stored for future predictions.

The Data Layer manages the storage of the raw dataset, processed dataset, and train-test split dataset throughout the workflow. The Model Storage component stores the trained Gradient Boosting model, model parameters, performance metrics such as Accuracy, Precision, Recall, and F1-Score, along with training logs and reports.

In the final stage, the Predict Fish Condition module accepts new test data, applies the same preprocessing steps, and uses the trained model to predict the fish health status. The prediction result is displayed as either Healthy Fish or Disease-Affected Fish. The system is implemented using technologies such as Python, Pandas, Scikit-learn, Gradient Boosting, and Flask, providing an efficient and reliable solution for automated fish health prediction based on water quality analysis..

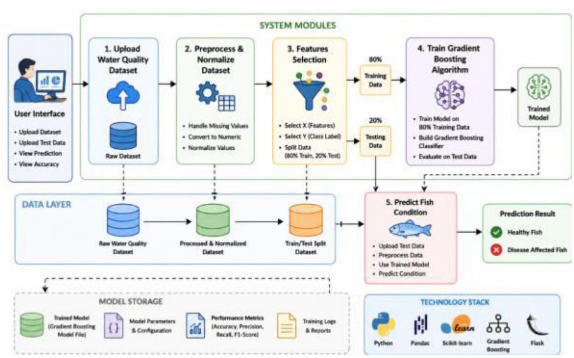

Fig1 proposed architecture

iii) Modules:

1) Upload Water Quality Dataset: using this module we will upload dataset to application

2) Preprocess & Normalize Dataset: using this module will convert all non-numeric data to numeric data and then normalize all values

3) Features Selection: using this module application will select X training features and Y class label and then split dataset into train and test where application using 80% dataset for training and 20% for testing

4) Train Gradient Boosting Algorithm: 80% training data will be input to Gradient Boosting algorithm to train a model and this model will be applied on test data to calculate prediction accuracy

5) Predict Fish Condition: using this module we will upload test data and then algorithm will predict weather fish is healthy or disease affected..

iv) Algorithms:

Gradient Boosting Algorithm

Gradient Boosting is a powerful machine learning algorithm used for classification and prediction tasks. It works by combining multiple weak learning models, usually decision trees, to create a strong predictive model. The algorithm builds trees sequentially, where each new tree attempts to correct the errors made by the previous trees. By continuously minimizing prediction errors through gradient descent optimization, Gradient Boosting improves the overall accuracy and performance of the model.

In this project, the Gradient Boosting algorithm is used to classify the condition of fish based on water quality parameters. After preprocessing, normalization, and feature selection, the training dataset is provided to the Gradient Boosting model. The algorithm learns the relationship between water quality attributes and fish health conditions, enabling it to accurately predict whether a fish is healthy or disease-affected. The trained model is then tested using unseen data to evaluate its performance and prediction accuracy.

One of the major advantages of Gradient Boosting is its ability to handle complex and nonlinear relationships within data. It provides high predictive accuracy, reduces bias and variance, and performs well even with large datasets. Due to these capabilities, Gradient Boosting is widely used in data mining, predictive analytics, healthcare, environmental monitoring, and aquaculture applications for reliable classification and decision-making.

.

3. EXPERIMENTAL RESULTS

Performance evaluation was carried out using standard metrics such as Accuracy, Precision, Recall, and F1-Score, confirming that the proposed system provides reliable and accurate results. These findings highlight that the system can effectively assist in early pneumonia diagnosis and support clinical decision-making.

Accuracy: A test's accuracy is its capacity to distinguish healthy from ill cases. Find the percentage of instances with genuine positives and negatives to assess test accuracy.

Accuracy = TP + TN /(TP + TN + FP + FN)

$$Accuracy = \frac{(TN + TP)}{T}$$

Precision: Classification accuracy or positive cases constitute precision. The formula for accuracy is:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$

Recall: A model's recall measures its ability to recognize all appropriate machine learning class instances. The ratio of accurately predicted positive observations to total positives indicates a model's class instance detection skill.

$$Recall = \frac{TP}{(FN + TP)}$$

mAP: Mean Average Precision ranks quality. It considers the number and order of relevant ideas. Calculating MAP at K uses the arithmetic mean of each user or query's Average Precision (AP).

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k$$

$$AP_k = \text{the AP of class } k$$

$$n = \text{the number of classes}$$

F1-Score: A high F1 score suggests an accurate machine learning model. Integrating recall and precision improves model correctness. Accuracy measures how often a model predicts a dataset correctly.

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$

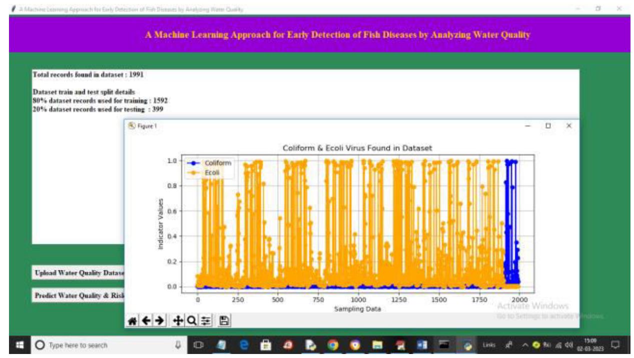

Fig 2:In above screen we can see dataset total values and then we can see training and testing dataset size and in graph x-axis represents number of records and y-axis represents presence quality of 'Coliform and Ecoli' virus where yellow line is for Ecoli and blue line for Coliform and now close above graph and then click on 'Train Gradient Boosting Algorithm' button to train algorithm and get below output

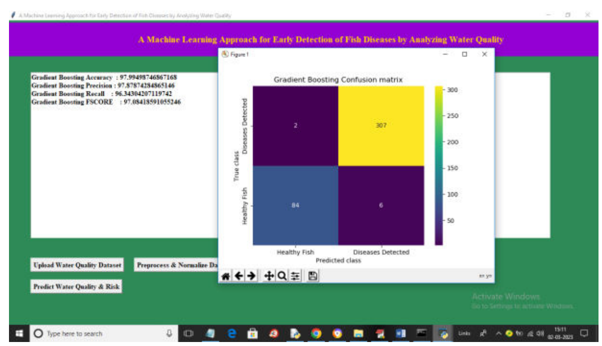

Fig 3:In above screen with Gradient Boosting we got 97% accuracy and in confusion matrix graph x-axis represents Predicted Labels and y-axis represents True Labels and blue colur boxes represents Incorrect prediction count which is 2 only and different colour boxes contains correct prediction count. Now close above graph and then click on 'Predict Water Quality & Risk' button to get below output

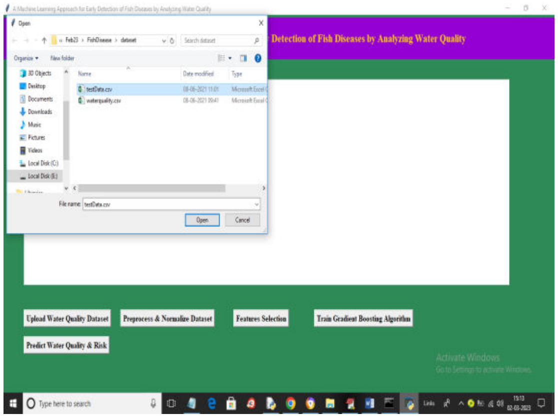

Fig 4:In above screen selecting and uploading 'testData.csv' file and then click on 'Open' button to load dataset and get below output

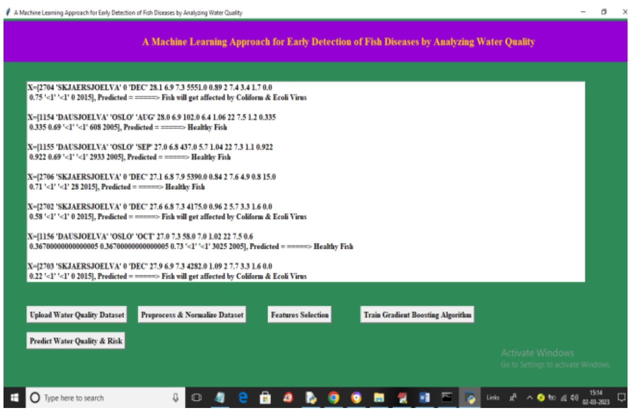

Fig 5:In above screen in square bracket we can see test data values and after arrow symbol we can see predicted values as healthy or disease affected fish

4. CONCLUSION

The proposed Fish Condition Prediction System provides an effective and automated solution for identifying whether a fish is healthy or disease-affected using water quality parameters. The system utilizes data preprocessing, normalization, feature selection, and the Gradient Boosting algorithm to build an accurate predictive model. By training the model on historical water quality data and evaluating it on unseen test data, the system demonstrates reliable prediction performance and assists in early disease detection. The implementation of machine learning techniques reduces manual effort, improves decision-making, and supports better aquaculture management. Overall, the proposed system offers a practical, efficient, and accurate approach for fish health monitoring, helping farmers and aquaculture experts take timely preventive measures to improve fish productivity and reduce losses caused by diseases... care and intelligent computer-aided medical diagnosis.

REFERENCES

1. R. B. Wynn, V. A. I. Huvenne, T. P. Le Bas et al., "Autonomous underwater vehicles (AUVs): their past, present and future contributions to the advancement of marine geoscience," Marine Geology, vol. 352, pp. 451–468, 2014.

View at: Publisher Site | Google Scholar

2. M. Dinc and C. Hajiyev, "Integration of navigation systems for autonomous underwater vehicles," Journal of Marine Engineering & Technology, vol. 14, no. 1, pp. 32–43, 2015.

View at: Publisher Site | Google Scholar

3. Y. Zhou, S. Cui, Y. Wang, and C. Ai, "Design of autonomous underwater vehicle (AUV) control unit," in 2015 ASEE Gulf-Southwest Annual Conference, pp. 25–27, ASEE Gulf-South, San Antonio, TX, 2015.

View at: Google Scholar

4. Y. Zhou, S. Cui, Y. Wang, and L. Zhai, "A refined attitude algorithm for AUV based on IMU," in 15th International Conference on Scientific Computing (CSC'17), pp. 16–22, CSREA Press ©, Las Vegas, NV, 2017.

View at: Google Scholar

5. SAUVC, https://sauvc.org/#competition.

6. A. Cadena, P. Teran, G. Reyes, J. Lino, V. Yaselga, and S. Vera, "Development of a hybrid autonomous underwater vehicle for benthis monitoring," in Proceedings of 2018 4th International Conference on Control, Automation and Robotics (ICCAR), pp. 20–23, IEEE, Auckland, New Zealand, 2018.

View at: Publisher Site | Google Scholar

7.M. Eichhorn, H. C. Woithe, and U. Kremer, "Parallel of path planning algorithms for auvs concepts, opportunities, and program – technical implementation," in 2012 Oceans - Yeosu, MTS/IEEE, Yeosu,South Korea, 2012.

Author Profiles

Ms. K. Krupa Sagari is working as an Assistant Professor in the Department of MCA at SRK Institute of Technology, Vijayawada. She has completed her MCA and M.Sc. (Mathematics). She has 6 years of teaching experience at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. Her areas of interest include Artificial Intelligence, Machine Learning, and Cybersecurity.

Ms.K.Pavani Working as Assistant & Head of Department of MCA ,in SRK Institute of technology in Vijayawada. She done with B .tech, MCA ,M. Tech in Computer Science .She has 10 years of Teaching experience in SRK Institute of technology, Enikepadu, Vijayawada,NTR District. Her area of interest includes Machine Learning with Python and DBMS.

Mr. Pothula Yedukondalu is an MCA Student in the Department of Computer Application at SRK Institute Of Technology, Enikepadu, Vijayawada, NTR District. He has Completed Degree in B.Sc.(electronics) from ASN Degree College Tenali. His area of interest are DBMS and Machine Learning with Python.