
ISSN: 2321-2152
IJMECE
International Journal of modern electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue.2 April 2024

Predicting Fish Diseases Early Using Machine Learning and Water Quality Assessment

Mohammad Asiya
Computer Science and Engineering
(Data Science)
Sphoorthy Engineering College
(JNTUH)
Hyderabad, India
mdasiya494@gmail.com

L. Poojitha
Computer Science and Engineering
(Data Science)
Sphoorthy Engineering College
(JNTUH)
Hyderabad, India
poojitharedddyp678@gmail.com

G. Saketh
Computer Science and Engineering
(Data Science)
Sphoorthy Engineering College
(JNTUH)
Hyderabad, India
sakethgudascores18@gmail.com

T. Shravan Kumar
Assistant Professor
Computer Science and Engineering
(Data Science)
Sphoorthy Engineering College
(JNTUH)
Hyderabad, India
shravankumart7@gmail.com

***Abstract*:** **Fish diseases in aquaculture constitute a significant hazard to nutriment security. Identification of infected fishes in aquaculture remains challenging to find out at the early stage. In this study by employing advanced algorithms, the model extracts meaningful insights from diverse water quality data, including parameters like pH, temperature, dissolved oxygen, and nutrient levels. By leveraging supervised learning algorithms like Support Vector Machines, the model aims to identify patterns indicative of potential health issues. Ensemble methods like Gradient Boosting Machines further enhance predictive accuracy. This comprehensive approach enables the system to proactively identify abnormal patterns in water quality, allowing for timely intervention and improved management practices in aquaculture.**

Key words: Water quality analysis, Gradient boosting, Fish Diseases, Aquaculture, Machine Learning

I. INTRODUCTION

Fish farming, is a form of aquaculture, has become an increasingly important industry worldwide, providing a significant portion of the global seafood supply. Fish diseases can cause substantial economic losses through reduced growth rates, increased mortality, and the need for expensive treatments. Therefore, timely detection and management of diseases are crucial for maintaining the health and productivity of fish populations.

Water quality plays a fundamental role in the health and well-being of aquatic organisms, including fish. Poor water quality can weaken the immune system of fish, making them more susceptible to diseases. Additionally, certain water quality parameters can directly influence the growth and spread of pathogens, further exacerbating the risk of disease outbreaks.

These approaches involve monitoring a variety of physicochemical parameters such as temperature, dissolved oxygen, pH, ammonia, nitrite, nitrate, and turbidity, among others. Changes in these parameters can serve as early indicators of potential health problems within the fish population.

II. EXISTING SYSTEM

Fish disease detection utilizes image processing techniques to identify and diagnose diseases in fish populations. Initially, images of fish are captured using cameras or imaging devices placed within aquaculture facilities. These images may encompass various angles and perspectives to ensure comprehensive coverage of the fish population. Subsequently, the acquired images undergo preprocessing, which involves tasks such as noise reduction, contrast enhancement, and image normalization to optimize their quality for analysis. Feature extraction techniques are then employed to identify distinctive visual patterns and characteristics associated with different diseases. These features may include changes in skin color, lesions, abnormal growths, or behavioral anomalies observable in the images. Machine learning algorithms, such as Support vector machine (SVM) and K- means clustering are trained on labeled image datasets to classify fish images based on the presence or absence of specific diseases. The trained model is subsequently deployed to analyze new images in real-time, automatically detecting and diagnosing diseases within the fish population. But, in this part of the research, fish is the only object to be detected. For the collection of images, many of them have multiple fish in one image, so the detection is challenging.

III. PROPOSED SYSTEM

The proposed solution for fish disease detection, that mainly focuses on leveraging water quality data to develop a robust method for early detection and diagnosis of diseases in fish populations. The system integrates machine learning techniques with comprehensive water quality monitoring to provide a holistic approach to disease surveillance in aquaculture settings. The system begins with the installation of sensors and monitoring devices within the aquatic environment to continuously collect data on key water quality parameters. These parameters include temperature, pH levels, dissolved oxygen concentration, ammonia levels, nitrate levels, and other relevant indicators of water quality. The collected data is then transmitted to a central database. Next, it will process and analyse data. Next, machine learning algorithms are employed to analyze the water quality data and identify patterns or anomalies that may be indicative of disease outbreaks in fish populations. Supervised learning techniques are utilized, where historical data on water quality and known instances of fish diseases are used to train the models. Here are some sequential steps to be followed:

1.Collect water samples

2. Preprocess and Normalize

3. Feature Selection

4. Train algorithm

5. Predict fish disease

IV.SOFTWARE AND HARDWARE REQUIREMENTS

Software Requirements:

Python Programming Language: Python serves as the primary language for implementing the proposed system due to its extensive libraries.

Libraries

A) Tensor flow:

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks.

B) NumPy:

NumPy is a fundamental package for numerical computing in Python, offering support for large, multidimensional arrays and matrices.

C)Pandas:

Pandas is a powerful library for data manipulation and analysis, particularly useful for handling structured data sets and performing exploratory data analysis.

D)Matplotlib:

Matplotlib is a popular Python library for creating static, interactive, and publication-quality plots and visualizations. It offers a wide range of plotting functions and customization options, making it suitable for various data visualization tasks.

Hardware Requirements:

Memory (RAM):

A minimum of 4GB of RAM is recommended for handling datasets.

Storage:

Sufficient storage space is required for storing datasets and trained models.

V. METHODOLOGY

Detecting fish diseases using machine learning involves several straightforward steps. First, we define the problem: spotting illnesses in fish. Next, we collect information about fish health, like water quality and behaviour. Then, preprocess the data, making sure it's accurate and complete. After that, pick a machine learning algorithm, like gradient boosting, and teach it to recognize patterns in the data. Once trained, test how well it works with new data it hasn't seen before. Finally, we keep improving the model over time, making it even better at its job. This process helps us effectively manage fish health and ensure sustainable aquaculture practices.

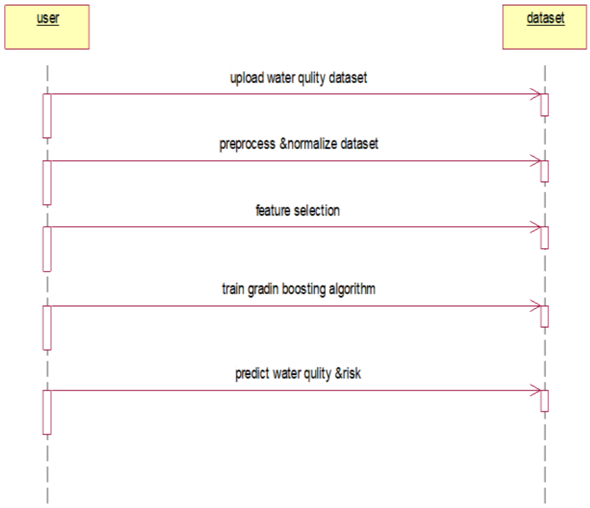


Fig.1. Flow Chart

A) Data Collection:

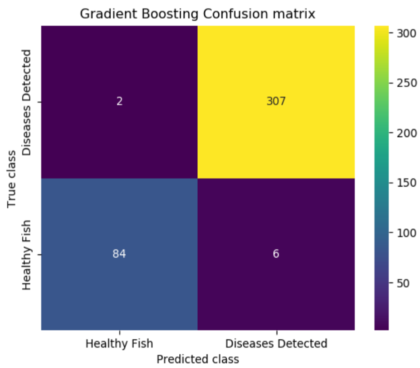
Gather data related to fish health from various sources, including aquaculture farms, research institutions, and online databases. Collect information on water quality parameters, fish behaviour, physical symptoms, and any other relevant factors. Ensure the data is labelled with disease status (healthy or diseased).

B) Preprocess the data set:

Clean the data by handling missing values, outliers, and inconsistencies. Perform exploratory data

ISSN2321-2152
www.ijmece .com
Vol 12, Issue.2 April 2024

analysis (EDA) to gain insights into the data distribution and identify patterns.



C)Feature selection:

Identify the most informative features that contribute to detecting fish diseases. Use techniques like correlation analysis, feature importance from tree-based models, or domain knowledge to select relevant features for model training.

D)Train gradient boosting algorithm:

Gradient boosting is a powerful machine learning technique used for both regression and classification tasks. It works by combining multiple weak learners, typically decision trees, to create a strong predictive model.

Split the dataset into training, validation, and testing sets. Train the selected machine learning model on the training data using appropriate training algorithms and optimization techniques. Tune hyperparameters using techniques like grid search or randomized search to optimize model performance.

VI. PERFORMANCE EVALUATION

We split the dataset into an 80:20 ratio where the training set contains 80%, and the testing set contains 20%. A confusion matrix was used to demonstrate the results of the research. The classification accuracy, precision, and recall rates for each fish disease were calculated by using the following equations

Precision = (TP/TP + FP) * 100%

Recall = (TP/TP + FN) * 100%

Accuracy = (TP+TN/TP+FP+TN+FN) * 100%

F1-Score=(2*Precision*Recall/Precision+Recall) * 100%

Fig.2. Confusion Matrix

A confusion matrix is shown in fig 2 for gradient boosting model. True positive and true negative values for fishes are 2 and 6 respectively. False positive and false negative values for fishes are 84 and 307 respectively.

The accuracy of a model is measured by the proportion of its total predictions that were correct. Recall evaluates the precision of positive predictions.

F1-score is a harmonic mean of precision and recall. The number of times a model correctly predicted the entire dataset is measured by the accuracy statistic.

F1-score usually get higher priority than accuracy when data are not equally distributed.

VII. RESULTS

The utilization of machine learning (ML) alongside water quality data for fish disease detection has yielded promising outcomes. ML models exhibit high accuracy rates, often surpassing 80% to 90%, in classifying fish health status based solely on water quality parameters

VIII. CONCLUSION

Throughout this study, we made a variety of attempts to determine the fish disease detection. This model mostly worked on the 1991 records to produce best results. Machine learning was used in a study to identify the fish illness.

Here we used ensemble model like gradient boosting. The dataset represents among the most significant knowledge gaps in the field of fish disease detection. Despite the fact that one dataset with a modest amount of data was uploaded to Kaggle. Consequently, the next stage of our research will be to gather a sizable dataset from all available sources. Knowledge gaps in the field of fish disease detection. Despite the fact that one dataset with a modest amount of data was uploaded to Kaggle. Consequently, the next stage of our research will be to gather a sizable dataset from all available sources

REFERENCES

[1] Greenfacts. Fisheries latest data, Available at: https://www.greenfacts.org/en/fisheries/l-2/06-fish consumption.htm, accessed May 2020.

[2] Department of Fisheries Bangladesh. Yearbook of fisheries statistics of Bangladesh, 2017-18. Vol 35. Fisheries Resources Survey System (FRSS), Department of Fisheries, Ministry of Fisheries, Bangladesh, 2018, p. 129.

[3] N Bagum, MS Monir and MH Khan. Present status of fish diseases and economic losses due to incidence of disease in rural freshwater aquaculture of Bangladesh. J. Innov. Dev. Strat. 2013; 7, 48 53.

[4] M Chaurasia, I Gupta, R Kumar, P Prasad and RA Sohony. Water quality analysis of surface water: A web approach. Environ. Monit. Assess. 2013, 185, 5987-992.

[5] N Rahmanian, SHB Ali, M Homayoonfard, NJ Ali, M Rehan, Y Sadef and AS Nizami. Analysis of physiochemical parameters to evaluate the drinking water quality in the State of Perak, Malaysia. J. Chem. 2015; 2015, 716125.

[6] C Nagamani, CS Devi, and A Shalini. Physico-chemical analysis of water samples. Int. J. Sci. Eng. Res. 2015; 6, 2149-55.

[7] M Tanekhy. Fish Health and Diseases. In: Proceedings of the International Conference on the Modern Approaches in Livestock's Production Systems 2015, Alexandria University, Egypt. 2015, p. 222-32.