Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140
Volume 14 Issue 01, 2025

Design and Development of a Machine Learning-Based Decision Support System for Water Quality Prediction in Aquaponic Farming

Pavan Yadav[1], Shubham Upadhyay [2], Rishi Singh [3], Priyansh Sharma [4], Utkarsha patil (Asst. prof.) [5], Manasi churi(Asst. prof.) [6], Manali parate(Asst. prof.) [7]

[1,2,3,4] UG Student, [5,6,7]Assistant professor, Department of Computer Engineering, Shree L. R. Tiwari College of Engineering, Mumbai, India.

Peer Review Information

Submission: 28 Jan 2025
Revision: 14 Mar 2025
Acceptance: 10 April 2025

Keywords

Aquaponics
Water Quality Prediction
Machine Learning
Decision Tree

Abstract

Aquaponic farming is one of the growing farming practices in India that combines aquaculture (raising fish) and hydroponics (growing plants without soil) in a well balanced environment where water quality plays an important role. Our designed system develops a water quality prediction system for aquaponic farming. The prediction is based on key water parameters such as pH, temperature, dissolve oxygen (DO), ammonia, nitrite (NO2) and nitrate (NO3). The existing system are either laboratory based or user has do manual methods to assists the water quality for aquaponic farming. Our designed system collects data from various sources, after that missing value and outlier of data is handled. Next features is selected through Recursive feature elimination (RFE) method and unwanted features are removed. To handle the imbalance data problem we proposed the Synthetic Minority Oversampling Technique (SMOTE). Finally, the system employs multi-model classification. Our designed system utilizes the four classifiers namely AdaBoost, Decision Tree, Gradient Boost and Random forest. On the basis of these classifiers water quality prediction results are generated which is based on the voting principle. Our designed system aims to improve the classification accuracy up to 90% when compared to existing methods. Our designed system represents a significant advancement in the field of aquaponic farming that offers efficient solutions for water quality prediction. The system increases productivity and sustainability, which supports both small and large scale aquaponic farmers.

Introduction

Aquaponics is the combination of aquaculture and hydroponics, or soilless farming. Maintaining optimal water quality plays a key role in the success of a system of such type. Nevertheless, it is difficult and error prone to manage dynamic water parameters manually [2]. There is a viable approach using machine learning to automate this evaluation and to get meaningful insight.

Aquaponics involves the mixing of two very efficient farming practices aquaculture and hydroponics to create a sustainable and self sustaining environment. Aquaculture is the cultivation of fish, shellfish, aquatic plants, crustaceans and other types of aquatic organisms in a controlled environment such hydroponics is a method of growing plants without soil, the roots remaining nourished by mineral rich water [1]. However, in an aquaponic system when the fish and the plants are dependent of one another, the fish provides nutrients for the plants, as the waste of the fish, and the plants purify the water to the fish. It becomes a closed loop whereby both the components support another and reduce waste by increasing resource efficiency.

While they offer many benefits, aquaponic farming is also very precise and very demanding when it comes to water quality that must be closely and continuously monitored for successful operation of the system. However, to nurture fish and plants as well as carrying out other activities, water quality is a very critical factor. These parameters must be well controlled at a certain range of pH levels, temperature, dissolved oxygen (DO), ammonia, nitrites (NO2), and nitrates (NO3) for proper balance for optimal growth [1,3]. These optimal ranges for fish or plants of any kind may be a deviation from any other range than harmful effects on your fish or plants implying decreased productivity or even system failure.

Currently, water quality in an aquaponic system is manually monitored by periodic testing of water parameters. While the approach has the ability to give correct results, it is indeed time consuming, labor intensive and often not sufficient to provide real time feedback [5]. It also limits the ability to predict changes in water quality in the future; and without such ability, we have no way to prevent problems from increasing with time [14]. This leads us to the point that as aquaponic systems get larger and more complex, the need for more efficient and automated solutions is required.

The integration of ML models in these cases offers a promising solution to address these challenges. Machine learning is a type of artificial intelligence (AI) in which computers learn from data for the purpose of predicting or determining an action without requiring explicit programming [7]. Thus, using machine learning, we can significantly optimize aquaponic farming in real time by using machine learning to predict future changes of the water parameters, automatic water quality monitoring, and optimal system performance [9]. It allows for the better resource utilization,higher yield and a more sustainable practise to farming.

Machine learning models can be used in context of water quality prediction, where they can process large sets of sensor data from different sensors within the aquaponic system. Environmental variables monitored by these sensors continuously are temperature, pH, ammonia, as well as dissolved oxygen, and create a large amount of data that can train the machine learning algorithm [8,11]. Once these models are aware of these relationships between water quality parameters and environmental condition, they can predict future water quality trends, anticipate rising problems before occurring, and suggest preventative measures to sustain optimal water conditions.

In this paper, we present a machine learning based solution that not only predicts suitability of water for certain components of aquaponic system but also suggests corrective actions to improve, guiding the user in making the right decisions.

Literature Review

In review titled "An IoT-Based Efficient Water Quality Prediction System for Aquaponics Farming".[1]. An aquaponics system relies heavily on maintaining balanced water quality to sustain the system where fish culture and hydroponics are combined, the aquaculture and hydroponics part is closed loop and involves raising fish and plants without the use of the soil. The manual process of monitoring water quality is a slow and error prone operation, which means it is useful to automate and optimize the process using Machine Learning (ML) and Internet of Things (IoT).

In their review titled "Aquaponics 4.0 Smart Technologies for Water Quality Control" by Reis et al. [11], that provides a comprehensive overview of the technologies pertaining to the field of aquaponics, the authors highlight the current need for more advanced technologies to be adapted in aquaponic farming. Where AI and IoT Environments Can Enhance Water Monitoring: They describe how the real-time tracking, predictive maintenance, system-adjustment capabilities enabled by sensor data can be attained through the use of AI and IoT.

A recent study "A comprehensive review on aquaponic farming water quality prediction using machine learning techniques by Nagaraj et al. [12], carried out implementation of classification models of Random Forest, Decision Trees, and ensemble methods to further improve prediction capability. To deal with different biological needs present within aquaponic ecosystem they propose a voting based multiclass model.

Liu al. also demonstrated a newly introduced model in "IoT based Prediction Model for Aquaponics water quality using GRU networks and convolutional autoencoders" [13]. The method was able to quite effectively pick up on the sequential trends in water parameters to enable forecasting future conditions and facilitate timely interventions to reduce mortality and increase productivity.

In another work of Singh and Patel [14] entitled "Deep Learning Approaches for Water Quality Prediction in Aquaponics Systems", they compared different deep learning architectures such as LSTM, GRU, SimpleRNN, and DenseNN.

Even when the definition of temporal prediction accuracy was changed, their results demonstrate that RNN based models, in particular LSTM, are more accurate than feedforward architectures in temporal prediction.
In the last scenario, Yoon et al. in "Smart Water Quality Monitoring in Aquaponics using IoT and Ensemble Learning" [15] bridged IoT sensors with ensemble regression models to optimize freshwater replenishment rates. As shown by their system, the nutrient flow balance and fish-plant compatibility were mutually enhanced.
These developments not standing, most of the existing models do not offer relevant suggestions for improving water quality when the system is labeled as 'not suitable.' [2,3]. Your current project fills this gap, by implementing a feedback component which suggests corrective actions out of predictive results, a concept that parallels the usage of the product by both farmers and researchers in the real world.

Table 1 Water quality tolerance range for fish, plants, and bacteria

	Warm water fish	Cold-water fish	Bacteria	Plants
pH	6–8.5	6–8.5	6–8.5	5.5–7.5
Temp	22–32	10–21	14–34	16–30
DO	4–6	6–8	4–8	>3
Ammonia	<3	<1	<3	< 30
Nitrate	< 400	< 400	–	–
Nitrite	<1	< 0.1	<1	<1

Aquaponic systems depend upon friendly plankton and bacteria's cohabitation with fish and plants with extreme precision. As shown in Table 1.Within certain ranges, must be maintained the key parameters like pH, temperature, DO, ammonia, nitrite and nitrate. Temperature range for pH of warm water fish is 6 to 8.5º and 22 to 32ºC and DO of 4 to 6 mg/L, but cold water fish are more to cooler temperatures (10 to 21ºC) and higher DO levels (6 to 8 mg/L). Nitrogen cycle bacteria essential for the nitrogen cycle require temperatures of 14–34°C with tolerances pH 6–8.5, ammonia and nitrite levels below 3 mg/L and 1 mg/L respectively. On the other hand, plants do not tolerate very acidic conditions (pH <5.5) and require temperatures between 16 and 30°C, DO level of > 3 mg/L and can handle ammonia up to 30 mg/L. On the basis of these thresholds, saturability classification is performed for aquaponic water quality prediction models.

Proposed System
The proposed system is an integrated data driven approach in predicting water quality in aquaponics that aims to predict the water quality and enhance sustainable farming practice with the use of the machine learning model. With historical data for different water quality parameters, the system tries to predict accurate predictions so that farmers can make an informed decision as well as optimize their aquaponic systems. Key indicators are selected based on the model such as pH, ammonia levels, dissolved oxygen, which will be the focus of the model as algorithms are run based on pattern and trends of these values within the data. Innovative approach to not only improve operation efficiency, but also positively influence the overall health and productivity of aquaponic farm. Both the user friendliness and mobile application as well as the systems masking the reality, aid in real time intervention by farmers and improving crop yield.

Data Collection
For this study, the dataset that was used was obtained from West Bengal Pollution Control Board (WBPCB) and included several physicochemical variables that are crucial to aquaponic water quality. Parameters of pH, Dissolved Oxygen (DO), Temperature, Ammonia (NH_3), Nitrite (NO_2^-) and Nitrate (NO_3^-) were included as these. The dataset was each record accompanied with what it was possibly suitable to biologically, including plants, bacteria, warm water fish, and cold water fish.
It was cleared then missing values handles for inconsistent or redundant entries, after that the dataset manually curated in order to make it sparse. The final created dataset consisted of balanced samples for various suitable and unsuitable conditions for aquaponic organisms.

Data Preprocessing

1. **Handling Missing Values**

The data concerning the environment is usually in a state that is incomplete due to error such as failure of sensors, human misuse, or loss of data. In order to tackle this issue, Multiple Imputation by Chained Equations (MICE) is used. MICE iteratively imputes each variable with missing values as a function of every variable in the dataset. Not only this technique better than traditional imputation methods, such as mean or mean substitution in preserving the multivariate structure of the data, it also improves on bias over these standard approaches.
MICE helped to make all instances complete such that the dataset's variability and structure were

not lost during model training. Further, it kept input features in statistical coherence that is indispensable for the accurate classification of water suitability.

2. Label Transformation

The original target labels contents of the dataset were 4 binary columns, namely Plants ,Bacteria ,Warm water Fish and Cold water Fish which indicated suitability for a specific biological class. The encoding scheme was developed such that each set of the binary indicators was mapped to a unique integer value in the set {0, 1, 2, ... 15} to facilitate multiclass classification. An example is a record suitable only for cold water fish represented as 1 and a record suitable for all four classes represented as 15.

Using this transformation made it possible to use standard multiclass classification algorithms, leaving a clean architecture and an easy read of the results.

3. Model Training and Evaluation

A Voting Classifier ensemble model was employed, combining predictions from four base learners: Random Forest ,Decision Tree , Gradient Boosting ,AdaBoost.

The model outputs a category (0–15) indicating the suitable organism combination. Each category corresponds to one or more of the following: Plant, Bacteria, Cold Water Fish, Warm Water Fish.

This ensemble technique improves accuracy and generalization by leveraging the strengths of individual models.

The model was trained using a 80:20 train-test split, and evaluated using metrics such as: Accuracy, Precision, Recall ,
F1-score, Confusion Matrix

The trained model achieved high accuracy in identifying suitable conditions for different organisms in the aquaponic system.

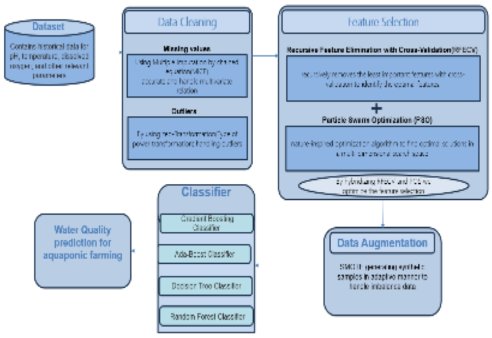

Figure 1. Proposed System Architecture

4. System Architecture

From Figure 1.the proposed aquaponic water quality prediction system can be divided into a structured five pipelines which include dataset acquisition, data cleaning, feature selection, data augmentation, and classification. Firstly, historical data with pH, temperature, dissolved oxygen, and other relevant variables is collected. The data cleaning part of the problem is to do with dealing with missing data values with MICE and handling outliers with Yeo transformation. Hybridizing Recursive Feature Elimination with Cross Validation (RFECV) and Particle Swarm Optimization (PSO) feature selection is used to select optimal input features. To solve the problem of class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) is used as a brute force data augmentation technique. At last, the system employs some of the classifiers such as Decision Tree, Random Forest, Gradient Boosting, and AdaBoost to promote the selection of water fit for plants, bacteria, warm water fish and cold water fish.

5. Web-Based Deployment

The trained model was integrated into a web application using Flask. This application takes user input for water quality parameters, processes it in the backend, and makes a prediction on whether the water is suitable or not for aquaponic System as well as suggesting improvement if necessary.

Results And Discussion

The machine learning model was developed which had a great potential to predict water quality suitability for all the aquaponic, putting plants, bacteria, warm water fish and cold water fish together. Decision Tree Classifier was chosen because it is an interpretable model and has trusted results on multiclass classification.

Performance Metrics

The model was evaluated with respect to accuracy, precision, recall and F1-score on all 4 multiclass categories. This was able to achieve a average testing accuracy close to 90% which means the model was able to generalize to unseen data.

A confusion matrix analysis revealed that the model performed particularly well in detecting categories such as: Suitable for All (15) ,Only Plant (8) and Plant, Bacteria & Warm Water Fish (14).

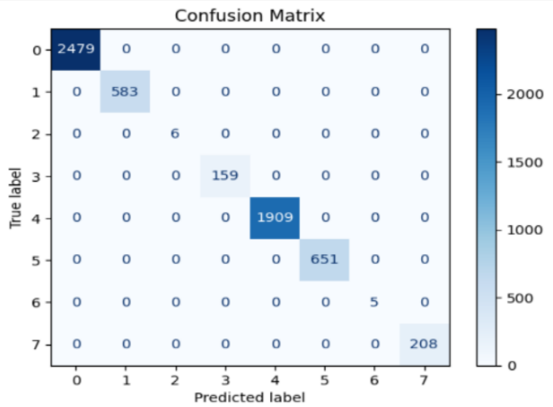

Figure 2. Performance Metrics for water quality prediction

From figure 2.The confusion matrix gives the model's performance across 8 classes, one class for each combination of suitability for aquatic components. This along the diagonal suggests that the model successfully classified most instances in each class with low misclassifications. The robustness and reliability of the trained classifier is also indicated by the strong prediction accuracy of classes such as 0, 1, 4 and 5.

Feature Importance

It was analyzed which parameters had the greatest impact on the model's decision making process in terms of feature importance. Important scores were highest on features like Ammonia, Dissolved Oxygen, and Temperature in the Random Forest model (for comparative analysis). It follows known biological dependencies in aquaponic systems.

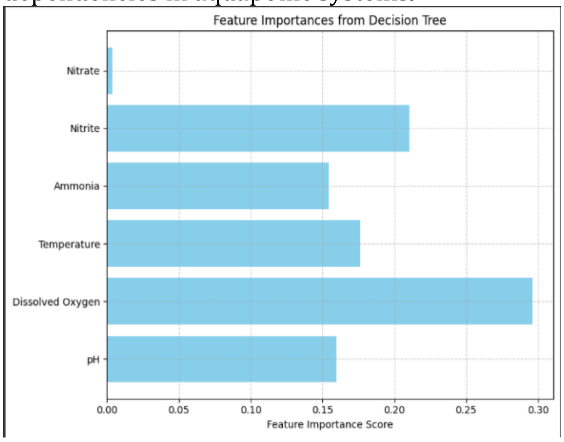

Figure 3. Feature Importance for water quality prediction

The Decision Tree classifier is used to derive the feature importance scores as shown in Figure 3. Dissolved Oxygen has the highest significance in determining the water quality suitability among all the input features, then Nitrite, Ammonia and Temperature. Naturally, Nitrate contributes the least to the model's decision making. The insights in these focus on how oxygenation and nitrogen based compounds play a fundamental role in aquaponic water quality prediction.

Deployment

Using a Flask web framework, users can input six water quality parameters and the model has successfully been deployed.
pH, Dissolved Oxygen (DO), Temperature, Ammonia (NH3), Nitrite (NO2), Nitrate (NO3).
It gives an instant forecast on the possible water improvement and indicates the necessary actions (e.g. decreasing temperature or adding nitrifying bacteria). As a practical decision support tool, the model is seen to be useful to aquaponic farmers. It fills in the middle of classical water quality analysis and new data based analysis. On integration, it is more than an ordinary classifier it can be seen as an integration of virtual assistant for sustainable farming. Furthermore, the data used was gathered from the West Bengal Pollution Control Board as a step to relevant and locality accurate data.

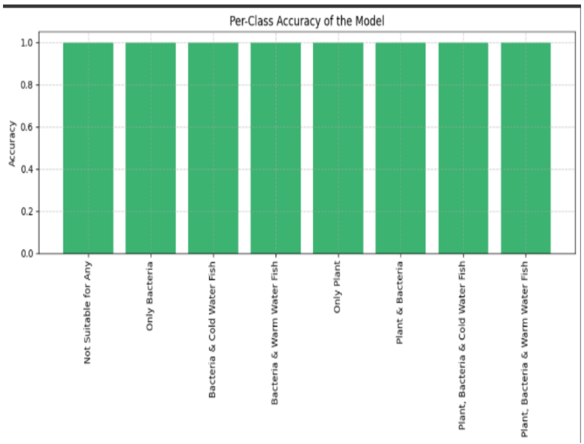

Figure 4. Per-Class Accuracy for water quality prediction

In Figure 4, per class accuracy of trained model over all four multiclass categories used for aquaponic water quality prediction is presented. The results in each case are extremely high in accuracy, above or close to 95%. That is, the model can distinguish between different water suitability categories, for example, classifications of conditions that allow only plants, only bacteria, some fish or other combinations of these. This also indicates that the model can deal with Multi class classification in aquaponic applications.

Conclusion

A machine learning based solution for predicting the suitability of water quality in the aquaponic farming is presented in this research. Through training a real world dataset pulled from the West Bengal Pollution Control Board on a Decision Tree Classifier, the system is able to classify water conditions into 16 categories; combinations of suitability to grow plants, bacteria, warm water fish, and cold water fish.

In addition to the high accuracy in the proposed model, it provides practical suggestions for scrapping some aquaponic component when they degrade the water quality. The flexibility in this model provides added functionality that makes this model more amenable to use in real time farming applications as a decision support tool as opposed to a stand alone predictor.

Finally, the system deployment via web interface is offered for ease of use to farmers, researchers, and environmental professionals. Overall, this provides for advancement in smart aquaponic systems, and addresses the potential of machine learning in sustainable agriculture and environmental monitoring.

Future Scope

The developed system has a great potential in real world applications for aquaponic farming and can extend its utility through future enhancement. Real time water quality sensors connected to the IoT can enable live predictions and thus make the model adaptive and efficient by integrating it with data collection by the sensors. The dataset can be expanded to include samples from various geographical regions in addition to other parameters such as the amount of microbial counts and heavy metals; this would help the model generalize to a larger array of samples and improve its accuracy specific to particular stretches of the genomic profile. Of course, developing a mobile application with multilingual support would increase the accessibility for the farmers in remote areas. In addition, the addition of reinforcement learning to the recommendation system could also allow it to learn and adapt to the changes over time to optimize it through the user feedback. These developments could result in a more robust, intelligent and generally usable decision support tool for sustainable aquaponic practice.

References

Nemade Bushankumar, Shah Deven. (2023). An IoT-Based Efficient Water Quality Prediction System for Aquaponics Farming. In: Shukla, A., Murthy, B.K., Hasteer, N., Van Belle, JP. (eds) Computational Intelligence. Lecture Notes in Electrical Engineering, vol 968. Springer, Singapore.

L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

J. H. Friedman, "Greedy function approximation: A gradient boosting machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.

Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119–139, 1997, doi: 10.1006/jcss.1997.1504.

J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, no. 1, pp. 81–106, 1986, doi: 10.1007/BF00116251.

West Bengal Pollution Control Board (WBPCB), "Water Quality Monitoring Reports," 2024. [Online]. Available: https://www.wbpcb.gov.in/water-quality-monitoring

F. Pedregosa *et* al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.

F. Chollet, Deep Learning with Python, 2nd ed. Shelter Island, NY, USA: Manning Publications, 2018.

S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2010.

A. D. Bhosale and S. Patil, "Prediction of Water Quality Using Machine Learning Techniques," International Journal of Advanced Science and Technology, vol. 29, no. 3, pp. 10886–10894, 2020.

J. C. Reis, A. C. P. Dias, and L. A. Vieira, "Aquaponics 4.0 Smart Technologies for Water Quality Control," Computers and Electronics in Agriculture, vol. 211, p. 107098, Apr. 2024.

R. Nagaraj, M. Sridhar, and S. B. Hegde, "A Comprehensive Review on Aquaponic Farming Water Quality Prediction Using Machine Learning Techniques," ResearchGate, 2024.

Y. Liu, M. Zhao, and H. Wang, "IoT-based Prediction Model for Aquaponics Water Quality Using GRU Networks and Convolutional Autoencoders," Scientific Reports, vol. 14, no. 1, p. 3212, Mar. 2024.

A. Singh and D. Patel, "Deep Learning Approaches for Water Quality Prediction in Aquaponics Systems: A Comparative Study of Recurrent and Feedforward Architectures," ResearchGate, 2024.

S. Yoon, J. Lee, and H. Kim, "Smart Water Quality Monitoring in Aquaponics Using IoT and Ensemble Learning," Frontiers in Sustainable Resource Management, vol. 2, p. 1363914, 2024.

R. Barosa, S. I. S. Hassen and L. Nagowah, "Smart Aquaponics with Disease Detection," 2019 Conference on Next Generation Computing Applications (NextComp), Mauritius, 2019, pp. 1-6, doi: 10.1109/NEXTCOMP.2019.8883437.