Predicting Water Quality Parameters in Mahseer Fish Farming Using Machine Learning Techniques

Article · December 2024

DOI: 10.14445/23488549/IJECE-V11I11P123

Nuradin Mohamed Abdikadir
SIMAD University

Husein Osman
SIMAD University

Abdikarim Abi Hassan
SIMAD University

SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V11I11P123

Volume 11 Issue 11, 286-296, November 2024

Original Article

Predicting Water Quality Parameters in Mahseer Fish Farming Using Machine Learning Techniques

Nuradin Mohamed Abdikadir[1], Ahmad Shahidan Abdullah[2], Husein Osman Abdullahi[3], Abdikarim Abi Hassan[4]

[1,4]Faculty of Engineering, SIMAD University, Mogadishu, Somalia.
[3]Faculty of Computing, SIMAD University, Mogadishu, Somalia.
[2]Telecommunication Software and Systems (TeSS) Research Group, Faculty of Electrical Engineering
Universiti Teknologi Malaysia (UTM), Johor Bahru, Malaysia.

[1]Corresponding Author : nuradin@simad.edu.so

Received: 25 September 2024 Revised: 24 October 2024 Accepted: 22 November 2024 Published: 03 December 2024

Abstract *- Mahseer fish farming faces challenges in maintaining optimal water quality, essential for fish health and growth. Poor water quality can lead to stress, disease, and mortality, impacting productivity. This study compares Random forest Regression (RF) and Support Vector Regression (SVR) models in predicting water quality parameters, such as pH, dissolved oxygen (DO), and temperature. The RF model outperformed SVR, showing superior accuracy with lower Mean Squared Error (MSE) and Mean Absolute Error (MAE) and higher R-squared values (99% for DO, 98% for temperature, and 95% for pH). RF's superior performance makes it a reliable tool for tracking water quality trends and fluctuations. Recommendations for enhanced monitoring include extending data turbidity and capturing seasonal and long-term trends, integrating sensors for additional parameters like ammonia and turbidity, and developing a user-friendly mobile app for real-time data and alerts. These improvements aim to support the sustainability and productivity of Mahseer fish farming.*

Keywords *- Aquaculture, Machine learning, RF, SVR, Water quality.*

1. Introduction

Aquaculture, sometimes known as aqua farming, is the controlled commercial rearing of aquatic animals [1]. This method entails raising various aquatic animals in tanks, ponds, and ocean enclosures, including fish, crustaceans, mollusks, and aquatic plants. The availability of fish for human consumption is greatly increased via aquaculture, which is an essential part of the world's food production process [2].

As the world's demand for seafood grows, aquaculture—especially the production of native species like Mahseer—becomes increasingly important. Still, close monitoring of the environmental parameters in the aquaculture system is necessary to guarantee the long-term viability of fish populations. Water quality significantly impacts Fish growth and well-being [4]. Considering this, incorporating Machine Learning (ML) and the Internet of Things (IoT) into fish farming techniques has become a game-changing strategy for effective management and monitoring [5].

IoT and machine learning integration have emerged as a key trend in the information industry's evolution, potentially disrupting several industries completely [6]. Although the idea behind the Internet of Things (IoT) was not completely new to the Information and Communications Technology (ICT) sector, Kevin Ashton's 1999 presentation on the subject was a critical milestone [7]. Several factors, such as tank capacity, water temperature, pH levels, oxygen requirement, and fish diet, affect the effectiveness of aquaculture reproduction. In regions like Malaysia, which are renowned for their hot, muggy weather and occasional heatwaves, IoT in aquaculture seems to be a means of achieving more efficient, sustainable, and secure risk management [8].

Researchers have explored applying ML techniques to predict water quality parameters in Mahseer fish farming to address this challenge. The efficiency of machine learning techniques in evaluating river water quality, including the computation of the Water Quality Index (WQI), has been shown in recent research. [9]. In water resource studies, these methods are effective modeling tools for intricate non-linear processes [9]. ML algorithms have been extensively applied in intelligent fish aquaculture, providing new opportunities to realise digital fishery farming [10].

This project aims to assess how well different Machine Learning (ML) techniques predict important water quality metrics and provide a framework for forecasting water quality specifically for Mahseer fish farming. By conducting experiments within Mahseer aquaculture settings, the research seeks to derive insights into the practical applicability of these techniques in real-world scenarios [11].

On the other hand, Maintaining key parameters such as temperature, pH, and DO within optimal ranges ensures a suitable environment for the fish. Table 1 summarizes the recommended ranges for these parameters based on previous research.

Monitoring and maintaining these water quality parameters within the optimal ranges are essential for promoting the health and productivity of Mahseer fish. This study assesses these parameters and their impact on fish health in Mahseer fish farming.

Table 1. Recommended water quality parameters for mahseer fish farming

Parameters	Value Ranges	References
Water Temperature	24ºC – 30ºC	[12],[13]
pH	6.5-8	[14]
DO	6-9 ppm	[15],[16]

2. Related Work

Aquaculture has seen a technological revolution with the development of IoT, allowing for real-time monitoring of water quality and prediction of water parameters to improve sustainability and profitability in the industry.

Under the guidance of P. Kirankumar (2021), a team from the National Institute of Technology Andhra Pradesh published a study on IOT and ML-based smart monitoring and water quality management in aquaculture [17]. Water quality is crucial for aquaculture profitability, which can be enhanced by implementing emerging technologies like sensors and ML to predict water parameters. These technologies can send alerts to farmers in advance. The use of such technologies can improve sustainability and profitability in aquaculture.

Azimbek Khudoyberdiev and colleagues (2023) report that fish farming is a rapidly growing sector expected to meet future seafood demands. A hybrid prediction-optimization-control system was implemented to maximize fish tank water quality while using the least energy possible. The system predicts and optimizes water quality parameters using fuzzy logic control, objective functions, and LSTM networks [18].

To track and control the quality of the water, the system also incorporates sensors, Internet of Things devices, and actuators. The fuzzy logic control module adjusts the actuators' operating level and activation duration based on expected water quality criteria. The system's ability to maintain ideal water quality while using the least energy was encouraging. In the study, five actuators were used.

Munesh Singh and colleagues (2022) reported on sustainable IoT solutions for freshwater aquaculture management. The paper evaluates the performance of intelligent forecasting algorithms for agriculture in India, focusing on water quality monitoring for aquaculture. Various ML models are compared, and the use of artificial neural networks for predictive analysis is discussed [19]. The proposed system integrates water quality sensors and actuators for real-time monitoring and control. The importance of tree-based models for non-linear relationships in water parameter monitoring is highlighted, along with the need to consider additional parameters for robust forecasting.

The model was trained with DO as the dependent variable and PH and temperature as independent variables. 160 samples were recorded from the testing platform, and 80 percent were used for training. The model's performance was evaluated using the remaining 20% of the samples.

Zeenat Reddy (2019) reports aquaculture involves cultivating marine creatures under controlled conditions. An IoT-based framework was developed to monitor water quality in aquaculture farms and provide alerts to farmers [20].

The framework includes a self-autonomous sensor node, base station, data storage module, and Android app for data visualization and alerts. The sensor node measures parameters like DO, water temperature, and pH, transmitting data to the base station via Zigbee communication. The base station then sends data to the cloud platform for analysis. The system is solar-powered and designed for the specific needs of aquaculture farmers in the Western Godavari Region of India. There were 50 aquaculture farmers in the analysis.

According to their recommendation, a system pilot deployment in Agra Canal, Okhla, demonstrated good performance with sensor accuracy of 80% for DO, 89% for pH, and 87% for water temperature. Bricks and stones were used to fix sensor location concerns.

3. Methodology

The study begins with a clear problem definition: accurately predicting essential water quality parameters critical for optimizing conditions in Mahseer fish farming. Comprehensive datasets are collected from IoT sensors deployed in the farming environment, focusing on data obtained from sources such as the Far East Planet (FEP) Tank One.

Data preprocessing involves addressing missing values in parameters such as Dissolved Oxygen (DO), temperature, and pH through mean imputation while treating outliers using methods like the Interquartile Range (IQR). The 'Date and Time' column is converted into a numerical format to facilitate analysis.

Table 2. Comparing the proposed system with related work

Goal of the approach	IoT sensors	Water quality prediction	Control method	Energy efficiency
Fish farm management [17]	Temperature, DO, pH	No	No	No
Recirculating aquaculture systems [21]	DO	Convolutional Neural Network	No	No
Environment monitoring [20]	Temperature, pH	No	Rule-based	No
Predictive Optimization for Fish Farming [18]	Temperature, Water level, pH, Conductivity	RNN-LSTM	Fuzzy logic	Yes
Monitoring and prediction [22].	Temperature, oxygen	ARIMA	No	No
Water quality prediction and suitability [23]	Temperature, conductivity, water level	Random Forest, Logistic Regression	No	No
Forecasting and maintaining water quality [24]	Temperature, conductivity, water level, pH, DO	Model Tree	Yes	No
Proposed System	Temperature, DO, pH	RF, SVR	Rule-based	Yes

The dataset is then prepared by segmenting it into training and testing sets, maintaining an appropriate ratio to support effective model development and evaluation. Random Forest Regression and Support Vector Regression models are implemented using Python within the Jupyter Notebook environment, with their performance optimized through hyperparameter tuning. Model accuracy is evaluated using evaluation measures such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2). The efficacy of each model is compared to predict water quality parameters vital for maintaining Mahseer fish health and effective farm management.

3.1. System Flowchart

The project's main objective is to develop a prediction model for evaluating water quality parameters in Mahseer fish farming using regression techniques. The complete procedure of this project's flow is depicted in the flowchart in Figure 1.

3.2. Data Gathering

The study requires crucial data gathering to gain preliminary insights and understand the dataset, aiming to evaluate data needs. The dataset, sourced from Far East Planet (FEP) Tank one in Excel format, consists of 22,813 records and 4 input attributes essential for predictive analysis, as illustrated in Figure 2.

3.3. Data Preparation

Data pre-processing is the initial step in preparing raw data for subsequent processing. The field data will be converted into a more comprehensible format. Data cleaning, transformation, and normalization are integral components of the data pre-processing phase, which is essential for training, testing, and analysis. During this procedure, the data type will be assessed, missing values will be identified, and additional measures will be implemented to ensure the data's reliability and suitability for the machine learning process.

3.4. Model Development and Implementation

Once the data was generated, it was divided into two distinct categories: the training data, which constituted seventy percent of the total (70%), and the testing data, which constituted thirty percent (30%). The performance of models is evaluated during the whole process of model creation and deployment by providing the ideal parameters of each model.

The data is partitioned into separate training and testing sets, and the parameters specified in each model are evaluated using the training and testing data. Random forest regression and support vector regression models will be used in this study. The primary tools utilized in this phase are Python libraries, specifically NumPy, Pandas, and Scikit-Learn.

3.5. Model Evaluation

Evaluating the precision of the model is a vital stage when determining the effectiveness of ML models. It assists in evaluating the accuracy of the model's predictions. Evaluation metrics vary depending on the type of problem. The errors indicate the extent to which the model is inaccurately predicting outcomes. The fundamental principle of accuracy evaluation is comparing the original target with the forecasted target using certain measures. The performance of models is evaluated using MSE MAE and R-squared (R^2).

MSE is a mathematical metric that calculates the squared difference between observed values and the predictions. The MSE measures the closeness of the ideal fit line to the given set of points. The average squared error is consistently higher than zero. The square function is employed to eradicate negative values. A prediction is more precise when the MSE approaches zero. The formula for MSE is.

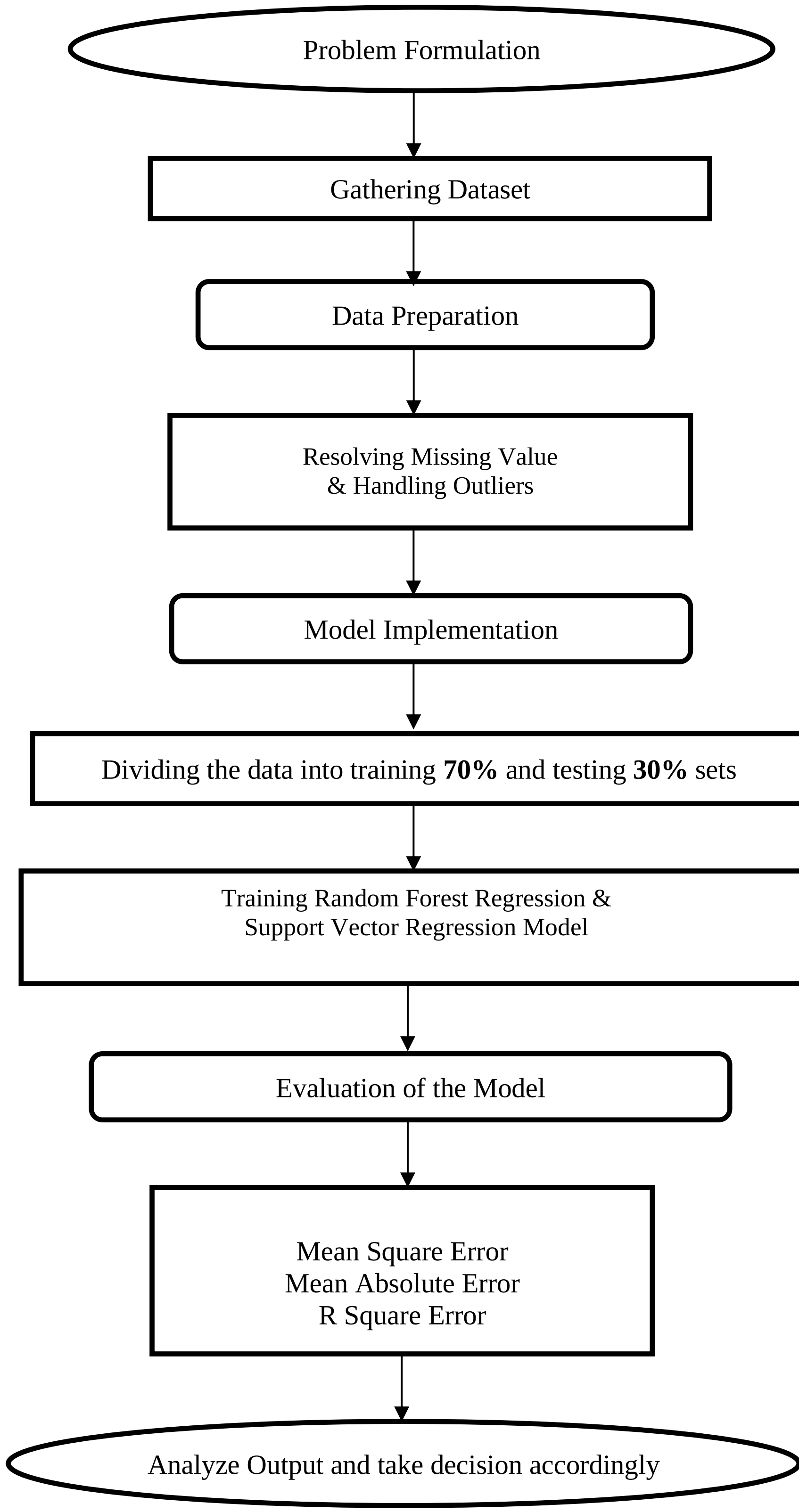


Fig. 1 System flowchart

```
df = pd.read_csv("WMS.csv")
df
```

	DO	Temperature	pH	Date and Time
0	11.37	5.0	19.0	2023-09-10: 12:26:05 AM
1	11.37	5.0	19.0	2023-09-10: 12:26:10 AM
2	11.47	5.0	19.0	2023-09-10: 12:28:06 AM
3	11.45	5.0	19.0	2023-09-10: 12:28:11 AM
4	11.48	5.0	19.0	2023-09-10: 12:29:05 AM
...	...	...	...	...
22808	9.60	NaN	NaN	2023-09-26: 11:54:04 PM
22809	9.61	NaN	NaN	2023-09-26: 11:55:03 PM
22810	9.60	NaN	NaN	2023-09-26: 11:56:04 PM
22811	9.62	NaN	NaN	2023-09-26: 11:57:05 PM
22812	9.61	NaN	NaN	2023-09-26: 11:58:04 PM

22813 rows × 4 columns

Fig. 2 Mahseer fish dataset

$$\text{MSE} = \frac{1}{n}\sum(\boldsymbol{actual\ values} - \boldsymbol{predicted\ values})^2$$

An analytical statistic called the MAE determines the average absolute difference between the actual and projected values. The computational approaches used by the MAE and MSE are different. The MAE measures the absolute difference between the expected and actual values. The following is the mathematical formula for MAE.

$$\text{MAE} = \frac{\sum_{1=1}^{n}(\boldsymbol{actual\ values} - \boldsymbol{predicted\ values})}{n}$$

R2 denotes the coefficient of determination. R-squared calculates how much of the variability in a dependent variable can be explained by independent variables. R-squared measures the degree of agreement between the observed and projected values to gauge how well the regression model fits the data.

$$R^2 = 1 - \frac{\boldsymbol{Sum\ of\ square\ of\ Residuals}}{\boldsymbol{Total\ sum\ of\ squares}}$$

3.4. Implementing ML Models

Feature selection and target variable definition are crucial in the next stages when dataset preparation is finished. The dataset is then divided into training and testing sets to assess the model's performance.

This partitioning ensures that the model is trained on a reliable dataset while verifying its generalization on unknown data by reserving 30% of the data for testing and 70% for training. The *train_test_split* function from *sklearn.model_selection* facilitates this partitioning seamlessly. Visual aids, such as Figure 3, can effectively illustrate these steps, clarifying the feature selection process and dataset partitioning, crucial for subsequent model development and evaluation.

3.6.1. Random Forest Regression

The Random Forest Regressor module from the sklearn package predicts water quality parameters in Mahseer fish farming. It offers various parameters that can be adjusted to obtain the most accurate results. Figure 4 shows the procedure of building the random forest using the test dataset, aligned with the specific parameters for Mahseer fish farming.

3.6.1. Support Vector Regression

The Support Vector Regression (SVR) module from the sklearn package predicts water quality parameters in Mahseer fish farming. It offers various parameters that can be adjusted to obtain the most accurate results. Figure 5 shows the procedure of building the SVR model using the test dataset, aligned with the specific parameters for Mahseer fish farming.

3.6.2. Comparison between the Two Models

Support Vector Regression and Random Forest Regression operate very differently. An ensemble learning technique called Random Forest Regression builds several decision trees during training and provides the average prediction of these trees.

This approach leverages the power of multiple models to improve performance and reduce overfitting, making it robust and capable of handling non-linearity in the data. However, it can be computationally expensive and may not achieve the same level of precision as some specialized models.

On the other hand, Support Vector Regression performs regression tasks by utilizing the ideas of Support Vector Machines (SVMs). With an emphasis on the points within a given margin, it seeks to identify the hyperplane that best fits the data. SVR works especially well in high-dimensional settings and resists overfitting in these situations. It is also well-suited for non-linear relationships, especially when using kernel functions like the Radial Basis Function (RBF).

Nevertheless, SVR can be sensitive to the choice of parameters and kernel, and its computational cost can be high for large datasets. Overall, while both models can handle non-linear relationships and provide robust predictions, the choice between Random Forest Regression and Support Vector Regression depends on the specific characteristics of the dataset and the computational resources available.

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor

# Convert 'Date and Time' column to datetime
new_df['Date and Time'] = pd.to_datetime(new_df['Date and Time'])

# Extract day, month, year, hour, minute, and second
new_df['day'] = new_df['Date and Time'].dt.day
new_df['month'] = new_df['Date and Time'].dt.month
new_df['year'] = new_df['Date and Time'].dt.year
new_df['hour'] = new_df['Date and Time'].dt.hour
new_df['minute'] = new_df['Date and Time'].dt.minute
new_df['second'] = new_df['Date and Time'].dt.second

# Define features and target variable
doX = new_df[['Date and Time', 'year', 'month', 'day', 'hour', 'minute', 'second']]
doy = new_df['DO']

# Split the data
doX_train, doX_test, doy_train, doy_test = train_test_split(doX, doy, test_size=0.3, random_state=42)
```

Fig. 3 Splitting data

```
from sklearn.svm import SVR

svm = SVR()

svm.fit(X_train,y_train)

▾ SVR
SVR()

pred2 = svm.predict(X_test)
```

Fig. 4 Random forest regression training

```
from sklearn.ensemble import RandomForestRegressor

rf = RandomForestRegressor(random_state=1)

rf.fit(X_train,y_train)

▾ RandomForestRegressor
RandomForestRegressor(random_state=1)

pred = rf.predict(X_test)
```

Fig. 5 Support vector regression training

4. Results and Discussion

The findings and ramifications of using machine learning methods, particularly Random Forest Regression (RFR) and Support Vector Regression (SVR), to forecast water quality metrics in Mahseer fish farming are covered in this section. The study evaluated the performance of these models based on their predictive accuracy and effectiveness in capturing complex relationships within the dataset. The findings highlight the potential for enhancing water quality management practices in aquaculture through advanced predictive analytics.

4.1. Predictive Analysis of DO

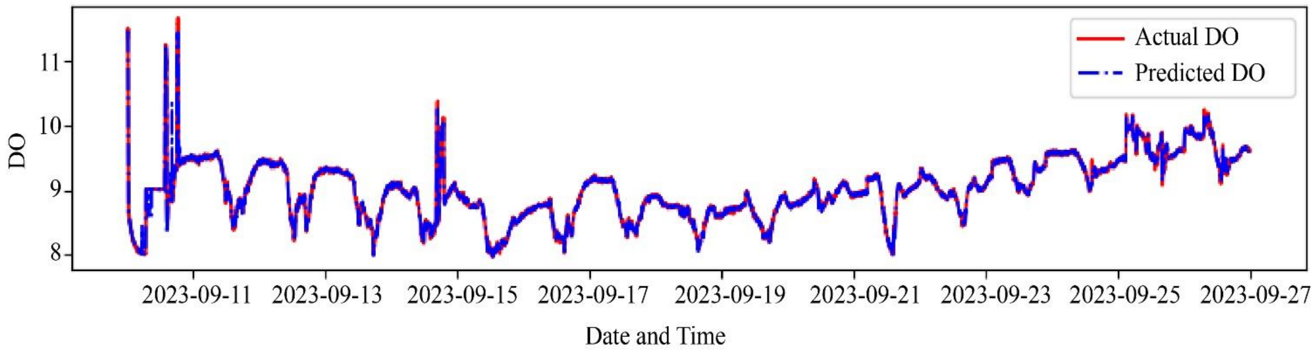
The RF model (Figure 6a) shows high accuracy, with predicted values closely following actual values, capturing trends and fluctuations effectively.

The SVR model (Figure 6b) also shows good alignment but slightly more deviations, indicating that the RF model is more convincing for predicting DO levels.

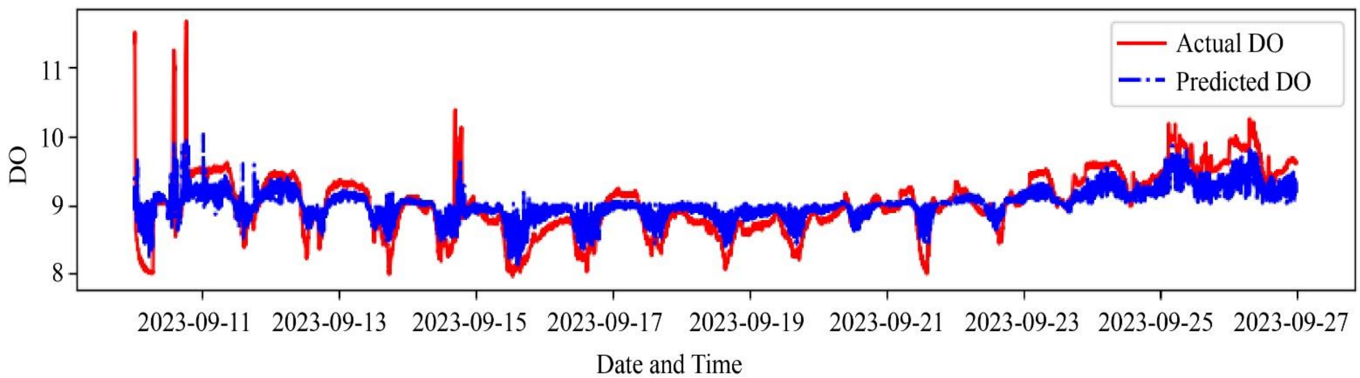
4.1.1. Statistical Evaluation DO

Table 3 illustrates the performance data for the Random Forest (RF) and Support Vector Regression (SVR) models, which reveal notable variations in their prediction accuracy and fitness. The RF model demonstrates superior performance, with an MSE of 0.00196, an MAE of 0.0181, and an R-squared (R^2) value of 0.9904. These results highlight the RF model's exceptional accuracy, as indicated by the very low MSE and MAE values and its excellent fit, with 99.04% of the variance in the dependent variable explained by the model.

In contrast, the SVR model shows a higher MSE of 0.0795 and MAE of 0.2145, suggesting less precise predictions. Additionally, the R^2 value of 0.6124 for the SVR model indicates a moderate fit, with only 61.24% of the variance explained. In general, the RF model is a more dependable option for prediction tasks in this scenario since it significantly outperforms the SVR model in identifying the underlying patterns in the data.



(a) RF



(b) SVR

Fig. 6 (a) Actual vs. Predicted DO values using Random Forest (RF) (b) Actual vs. Predicted DO values using Support Vector Regression (SVR)

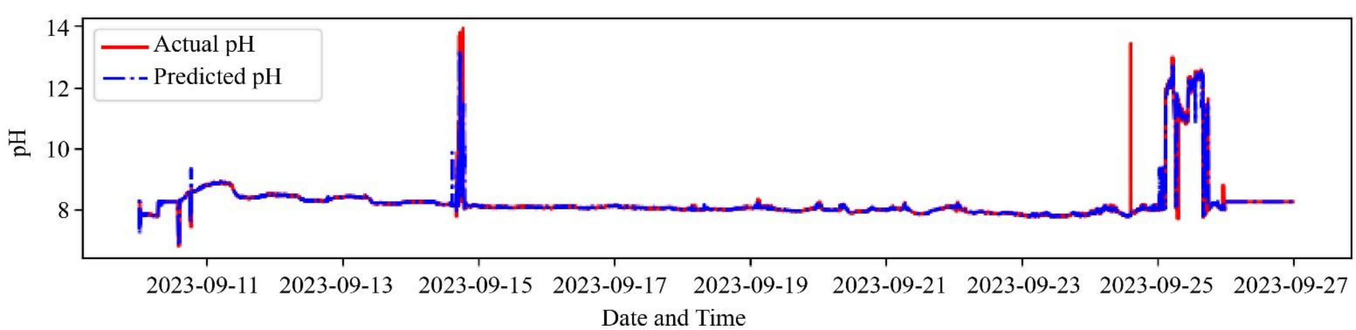
Table 3. DO Performance metrics for RF and SVR

Metrics	DO	
	RF	SVR
MSE	0.00196	0.0795
MAE	0.0181	0.2145
R^2	0.9904	0.6124

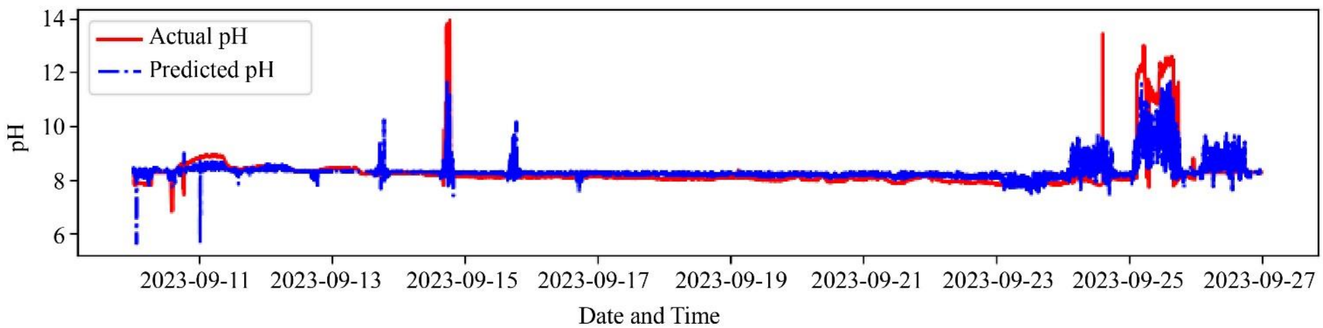
4.2. Predictive Analysis of pH

The RF model (Figure 7a) shows high accuracy, with predicted values closely following actual values and effectively capturing trends and fluctuations. It demonstrates strong predictive ability throughout most of the timeframe, though it struggles with sharp spikes around September 15th and 25th.

The SVR model (Figure 7b) also aligns well with actual pH values but has slightly more deviations than the RF model. It performs well during stable periods but faces challenges during the same sharp fluctuations. While both models perform well, the RF model is more realistic for predicting pH levels.



(a) RF



(b) SVR

Fig. 7 (a) Actual vs. Predicted pH values using Random Forest (RF) (b) Actual vs. Predicted DO values using Support Vector Regression (SVR)

4.2.1. Statistical Evaluation (pH)

As demonstrated in Table 4, the performance measures of the Random Forest (RF) and Support Vector Regression (SVR) models reveal notable variations in their prediction accuracy and fitness. The RF model demonstrates superior performance, with an MSE of 0.0275, an MAE of 0.0231, and an R-squared (R^2) value of 0. 9501.

These findings demonstrate the RF model's outstanding fit, with 95.01% of the variance in the dependent variable explained by the model and its remarkable precision, as seen by the extremely low MSE and MAE values.

In contrast, the SVR model shows a higher MSE of 0.2867 and MAE of 0.2674, suggesting less precise predictions. Additionally, the R^2 value of 0.4799 for the SVR model indicates a moderate fit, with only 47.99% of the variance explained.

Overall, the RF model is markedly more effective than the SVR model in capturing the underlying patterns in the data, making it a more reliable choice for predictive tasks in this context.

Table 4. pH Performance Metrics for RF and SVR

Metrics	pH	
	RF	SVR
MSE	0.0275	0.2867
MAE	0.0231	0.2674
R^2	0.9501	0.4799

4.3. Predictive Analysis of Temperature

The RF model (Figure 8a) shows high accuracy, with predicted values closely following actual values and effectively capturing trends and fluctuations.

It demonstrates strong predictive ability throughout most of the timeframe, though it struggles with sharp spikes around September 11th and 15th. Overall, the RF model effectively predicts temperature levels, accurately capturing the overall trend and small fluctuations.

The SVR model (Figure 8b) also aligns well with actual temperature values but has slightly more deviations than the RF model. It performs well during stable periods but faces challenges during the same sharp fluctuations observed in the RF model.

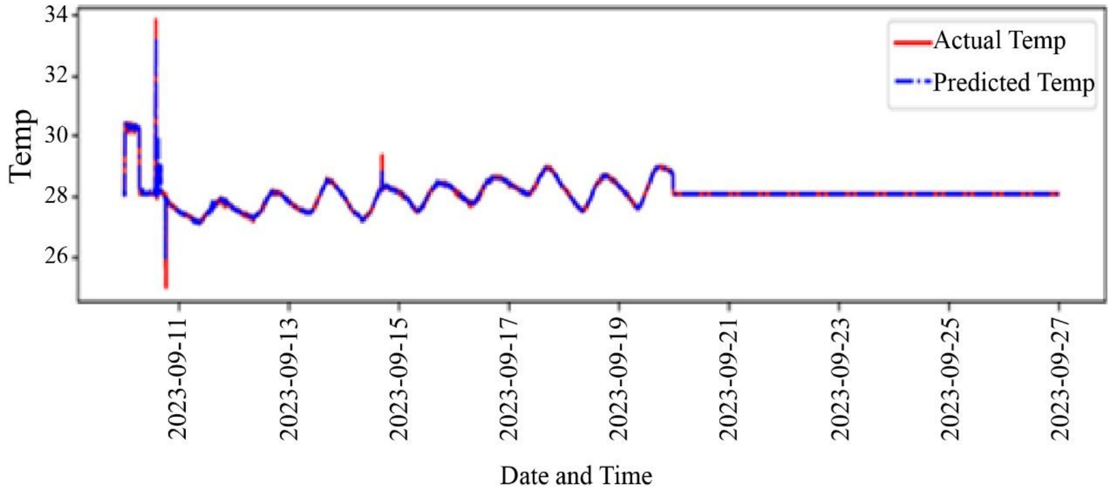
These deviations suggest that while the SVR model performs adequately, it is slightly less precise than the RF model in predicting temperature values. Overall, the RF model is more reliable for predicting temperature levels.

4.3.1. Statistical Evaluation (Temp)

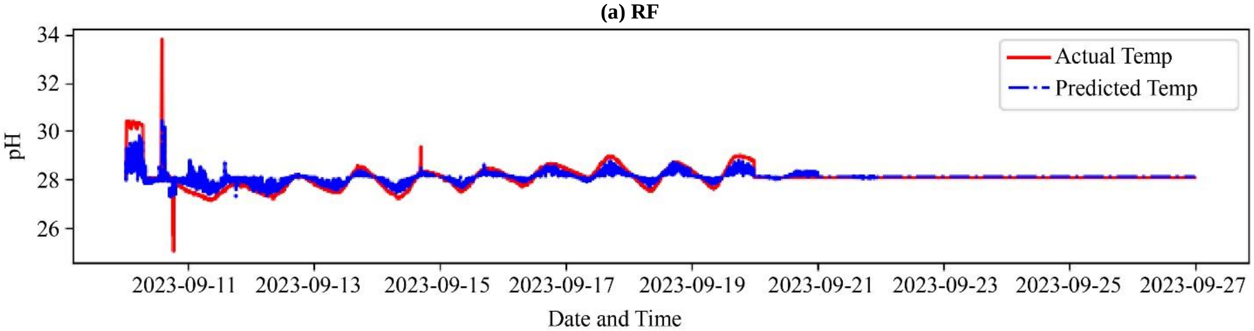
As Table 5 illustrates, the performance indicators of the Random Forest (RF) and Support Vector Regression (SVR) models reveal notable variations in their prediction accuracy and fitness. The RF model demonstrates superior performance, with an MSE of 0.0034, MAE of 0.0101, and an R-squared (R^2) value of 0.9831. These results highlight the RF model's exceptional accuracy, as indicated by the very low MSE and MAE values and its excellent fit, with 98.31% of the variance in the dependent variable explained by the model.

In contrast, the SVR model shows a higher MSE of 0.0882 and MAE of 0.1642, suggesting less precise predictions. Additionally, the R^2 value of 0.5634 for the SVR model indicates a moderate fit, with only 56.34% of the variance explained.

Overall, the RF model is markedly more effective than the SVR model in capturing the underlying patterns in the data, making it a more reliable choice for predictive tasks in this context.



(a) RF



(b) SVR

Fig. 8(a) Actual vs. Predicted Temp Values using Random Forest (RF) (b) Actual vs. Predicted Temp Values using Support Vector Regression (SVR)

Table 5. Temp Performance Metrics for RF and SVR

Metrics	**Temperature**	
	RF	**SVR**
MSE	0.0034	0.0882
MAE	0.0101	0.1642
R^2	0.9831	0.5634

5. Conclusion

The study evaluated the performance of two ML models, Random Forest Regression (RF) and Support Vector Regression (SVR), in predicting water quality parameters—pH, DO, and temperature—in Mahseer fish farming. The results demonstrated that the RF model consistently outperforms the SVR model in terms of predictive accuracy and fit.

The RF model exhibited lower MSE and MAE values across all parameters, indicating its predictions are closer to the actual observed values, thus minimizing prediction errors. Specifically, the RF model achieved R-squared values of 99% for DO, 95% for pH, and 98% for temperature, showcasing its high level of accuracy in capturing the variance in these parameters. These high R-squared values suggest that the RF model can effectively explain most of the variability in water quality parameters, making it a strong tool for monitoring and managing water quality in Mahseer fish farming.

In contrast, while still performing reasonably well, the SVR model showed higher MSE and MAE values and lower R-squared values of 61% for DO, 48% for pH, and 56% for temperature. These results indicate that the SVR model's predictions are less accurate and reliable compared to the RF model.

Given its superior performance, the Random Forest model is recommended for predicting water quality parameters in Mahseer fish farming.

The RF model's ability to closely follow actual water quality trends and accurately capture fluctuations makes it an essential tool for ensuring optimal water conditions, which are critical for the health and growth of Mahseer fish.

Implementing the RF model can enhance water quality management practices, leading to better decision-making and potentially improving overall aquaculture productivity and sustainability.

5.1. Suggestions for Future Work

To further improve water quality management in Mahseer fish farming, several recommendations can be made based on the findings of this study and additional considerations for comprehensive monitoring. Collecting data over extended periods is crucial for capturing seasonal variations and long-term trends in water quality, providing insights into how different times of the year affect fish health and growth. This long-term data collection helps identify patterns and trends that might not be apparent in short-term studies, allowing for better anticipation of changes that could impact the farming environment.

Integrating additional sensors, such as those measuring ammonia, turbidity, and conductivity, will offer a more comprehensive view of water quality. Ammonia sensors are essential because high ammonia levels can be toxic to fish, and early detection allows for timely interventions. Turbidity sensors help monitor the clarity of the water, which affects fish respiration and overall water quality, ensuring a suitable living environment. Conductivity sensors measure the water's ability to conduct electricity, indicating the concentration of dissolved salts, which can reflect changes due to runoff, pollution, or other factors impacting water quality.

Moreover, creating a user-friendly mobile application for farmers to access real-time water quality data, alerts, and predictions is necessary. Such an application would empower farmers with timely information, enabling them to make informed decisions and take proactive measures to ensure optimal water conditions. The app could provide instant notifications about critical changes in water quality parameters, offer historical data analysis, and suggest best practices for maintaining a healthy farming environment. This technological integration would significantly enhance farmers' ability to effectively manage and sustain Mahseer fish farming.

Funding Statement

SIMAD University funded this publication under the Centre for Research and Development office (CRD). Its contents are solely the authors' responsibility and do not necessarily represent the official views of SIMAD University or the Centre and Research Development Office.

References

[1] Haochen Hou et al., "An Environmental Impact Assessment of Largemouth Bass (Micropterus Salmoides) Aquaculture in Hangzhou, China," *Sustainability (Switzerland)*, vol. 15, no. 16, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[2] Lahsen Ababouch et al., "Value Chains and Market Access for Aquaculture Products," *Journal of the World Aquaculture Society*, vol. 54, no. 2, pp. 527-553, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] I. Mulyani et al., "The Ecological Analysis of the Habitat of Semah Fish (Tor Tambroides, Bleeker 1854) in the Kampar Kanan Hulu River Koto Kampar Hulu District, Kampar Regency, Riau Province," *IOP Conference Series: Earth and Environmental Science, 11th International and National Seminar on Fisheries and Marine Science*, Pekanbaru, Indonesia, vol. 1118, pp. 1-8, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Peda Gopi Arepalli, and Jairam Naik Khetavath, "An IoT Framework for Quality Analysis of Aquatic Water Data Using Time-Series Convolutional Neural Network," *Environmental Science and Pollution Research*, vol. 30, pp. 125275-125294, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Harsh Dadhaneeya, Prabhat K. Nema, and Vinkel Kumar Arora, "Internet of Things in Food Processing and its Potential in Industry 4.0 Era: A Review," *Trends in Food Science and Technology*, vol. 139, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Abu Taher Tamim et al., "Development of IoT Based Fish Monitoring System for Aquaculture," *Intelligent Automation & Soft Computing*, vol. 32, no. 1, pp. 55-71, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Andrew B. Gallemit, "Water Monitoring and Analysis System: Validating an IoT-Enabled Prototype towards Sustainable Aquaculture," *International Journal of Advanced Multidisciplinary Studies*, vol. 3, no. 6, pp. 496-517, 2023. [Google Scholar] [Publisher Link]

[8] Danial Mohammad Ghazali et al., "Smart IoT Based Monitoring System for Fish Breeding," *Journal of Advanced Research in Applied Mechanics*, vol. 104, no. 1, pp. 1-11, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Illa Iza Suhana Shamsuddin, Zalinda Othman, and Nor Samsiah Sani, "Water Quality Index Classification Based on ML: A Case from the Langat River Basin Model," *Water*, vol. 14, no. 19, pp. 1-20, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[10] Shili Zhao et al., "Application of ML in Intelligent Fish Aquaculture: A Review," *Aquaculture*, vol. 540, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[11] Jianlong Xu et al., "An Alternative to Laboratory Testing: Random Forest-Based Water Quality Prediction Framework for Inland and Nearshore Water Bodies," *Water*, vol. 13, no. 22, pp. 1-19, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[12] Melinda Mei Lin Lau et al., "A Review on the Emerging Asian Aquaculture Fish, the Malaysian Mahseer (Tor tambroides): Current Status and the Way Forward," *Proceedings of the Zoological Society*, vol. 74, no. 2, pp. 227-237, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[13] K.C. Ahmed Jala et al., "Phototrophic Purple Bacteria as Feed Supplement on the Growth, Feed Utilization and Body Compositions of Malaysian Mahseer, *Tor Tambroides* Juveniles," *Malaysian Science*, vol. 45, no. 1, pp. 135-140, 2016. [Google Scholar] [Publisher Link]

[14] Nur Syuhada Iskandar et al., "Elevated Carbon Dioxide and its Impact on Growth, Blood Properties, and Vertebral Column of Freshwater Fish Mahseer, Tor tambroides Juveniles," *Fishes*, vol. 8, no. 6, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Mohammod Kamruzzaman Hossain et al., "Growth Performance, Fatty Acid Profile, Gut, and Muscle Histo-Morphology of Malaysian Mahseer, Tor Tambroides Post Larvae Fed Short-Term Host Associated Probiotics," *Aquaculture and Fisheries*, vol. 9, no. 1, pp. 35-45, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Aidil Ikhwan Redhwan et al., "Mahseer in Malaysia: A Review of Feed for Cultured Tor Tambroides and Tor Tambra," *Bioscience Research*, vol. 19, no. SI-1, pp. 349-359, 2022. [Google Scholar] [Publisher Link]

[17] P. Kirankumar et al., "Smart Monitoring and Water Quality Management in Aquaculture using lOT and ML," *IEEE International Conference on Intelligent Systems, Smart and Green Technologies*, Visakhapatnam, India, pp. 32-36, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Azimbek Khudoyberdiev et al., "Enhanced Water Quality Control Based on Predictive Optimization for Smart Fish Farming," *Computers, Materials and Continua*, vol. 75, no. 3, pp. 5471-5499, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Munesh Singh, Kshira Sagar Sahoo, and Anand Nayyar, "Sustainable IoT Solution for Freshwater Aquaculture Management," *IEEE Sensors Journal*, vol. 22, no. 16, pp. 16563-16572, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[20] Zeenat Shareef, and S.R.N. Reddy, "Design and Development of IoT-Based Framework for Indian Aquaculture," *Intelligent Communication, Control and Devices, Advances in Intelligent Systems and Computing*, vol. 989, pp. 195-201, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Xuxiang Ta, and Yaoguang Wei, "Research on a Dissolved Oxygen Prediction Method for Recirculating Aquaculture Systems Based on a Convolution Neural Network," *Computers and Electronics in Agriculture*, vol. 145, pp. 302-310, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[22] Maxime Lafont et al., "Back to the Future: IoT to Improve Aquaculture: Real-Time Monitoring and Algorithmic Prediction of Water Parameters for Aquaculture Needs," *Global IoT Summit*, Aarhus, Denmark, pp. 1-6, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[23] Marzia Ahmed et al., "Analyzing the Quality of Water and Predicting the Suitability for Fish Farming Based on IoT in the Context of Bangladesh," *International Conference on Sustainable Technologies for Industry 4.0*, Dhaka, Bangladesh, pp. 1-5, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[24] Guandong Gao, Ke Xiao, and Miaomiao Chen, "An Intelligent Iot-Based Control and Traceability System to Forecast and Maintain Water Quality in Freshwater Fish Farms," *Computers and Electronics in Agriculture*, vol. 166, 2019. [CrossRef] [Google Scholar] [Publisher Link]