Turk. J. Fish.& Aquat. Sci. 26(1), TRJFAS27622
https://doi.org/10.4194/TRJFAS27622

PROOF
RESEARCH PAPER

Turkish Journal of FISHERIES and AQUATIC SCIENCES

Using Different Machine-Learning Algorithms to Predict Dissolved Oxygen Concentration in Rainbow Trout Farms

Boran Karataş[1,*], Cihan Çakmakçı[2], Elif Sena Yücel[3], Muhammet Demir[4], Fazıl Şen[5]

[1]Van Yüzüncü Yıl University, Faculty of Fisheries, Department of Aquaculture, Van, Türkiye.
[2]Van Yüzüncü Yıl University, Faculty of Agriculture, Department of Agricultural Biotechnology, Animal Biotechnology Section, Van, Türkiye.
[3]Van Yüzüncü Yıl University, Institute of Science, Van, Türkiye.
[4]Fisheries and Aquaculture Branch Directorate, Van Provincial Directorate of Agriculture and Forestry, Van, Türkiye.
[5]Van Yüzüncü Yıl University, Faculty of Fisheries, Department of Basic Sciences, Van, Türkiye.

How to Cite

Karataş, B., Çakmakçı, C., Yücel, E.S., Demir, M., Şen, F. (2026). Using Different Machine-Learning Algorithms to Predict Dissolved Oxygen Concentration in Rainbow Trout Farms. *Turkish Journal of Fisheries and Aquatic Sciences*, *26(1)*, *TRJFAS27622*. https://doi.org/10.4194/TRJFAS27622

Article History
Received 25 December 2024
Accepted 11 June 2025
First Online 01 July 2025

Corresponding Author
E-mail: borankaratas@yyu.edu.tr

Keywords
Aquaculture
Machine learning
Rainbow trout
Water quality

Abstract

Dissolved oxygen (DO) is a vital parameter in intensive rainbow trout aquaculture, directly influencing fish growth, health, and survival. As such, accurate monitoring and prediction of DO levels are crucial for ensuring sustainable and efficient aquaculture practices. This study assessed and compared the predictive performance of four machine learning algorithms Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Averaged Neural Networks (avNNet) in estimating DO concentrations based on a range of water quality parameters. A total of 120 samples were collected from 12 rainbow trout farms across Türkiye. The input variables included suspended solids, electrical conductivity, turbidity, nitrate, nitrite, ammonia, ammonium, orthophosphate, pH, water temperature, and total phosphorus. DO levels ranged between 8 and 15 mg/L. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE), and the coefficient of determination (R^2). All models demonstrated strong predictive accuracy, with XGBoost achieving the best overall performance (MAE: 0.44, RMSE: 0.58, MAPE: 0.04, R^2: 0.78), followed by RF, avNNet, and MARS. These findings highlight XGBoost as a robust predictor of dissolved oxygen levels in aquaculture systems, which may contribute to improving water quality management and increasing productivity in rainbow trout aquaculture.

Introduction

Rainbow trout (*Oncorhynchus mykiss*) is one of the most widely farmed cold-water fish species globally. In 2019, global trout production was estimated at approximately 940,000 tons, with rainbow trout accounting for 97% of this total (FAO, 2022). Countries such as Chile, Türkiye, and Iran have experienced substantial growth in intensive rainbow trout aquaculture in recent years (FAO, 2020), a trend expected to continue due to the species' high commercial value and strong market demand (D'Agaro et al., 2022). Among the various environmental factors affecting trout farming, water quality, particularly dissolved oxygen (DO) concentration, plays a central role. DO is a key indicator of water quality and a critical determinant of fish metabolism, growth, immune response, and overall survival. Inadequate DO levels can impair feeding behavior, suppress growth, and increase susceptibility to disease and mortality (Welker et al., 2019). Conversely, excessively high DO concentrations may lead to gas bubble disease, a potentially fatal condition (Arabacı et al., 2020). Beyond its biological significance, DO management also represents a major operational cost in trout farming, second only to feed and labor (Royer et al., 2021). Therefore, effective control and accurate prediction of DO levels are essential for maintaining fish health and optimizing aquaculture profitability (Jiang et al., 2021).

DO concentrations in aquaculture systems are influenced by a complex interplay of biotic and abiotic factors, including water temperature, pH, ammonia, nitrite, nitrate, orthophosphate, turbidity, electrical conductivity, and suspended solids (Jiang & Yan, 2022; Moon et al., 2022). These variables often interact in dynamic and nonlinear ways, which makes modeling DO behavior particularly challenging (Sun et al., 2021; Yin et al., 2021; Guo et al., 2022). Traditional statistical or linear regression models generally fall short in capturing these complex relationships. In contrast, machine learning (ML) algorithms, especially those capable of modeling nonlinear interactions and high-dimensional datasets, have emerged as powerful alternatives for environmental and water quality prediction tasks (Li et al., 2018; Cao et al., 2019; Li et al., 2022). ML models are increasingly applied in hydrological, ecological, and aquaculture contexts, including DO estimation.

Numerous studies have demonstrated the predictive capabilities of ML approaches such as support vector machines (SVM), decision trees, neural networks, and ensemble models for DO modeling. However, most of these applications have focused on natural aquatic systems, such as rivers, lakes, and reservoirs (Ahmed et al., 2019; Ross & Stock, 2019; Nasir et al., 2022). In comparison, studies addressing DO prediction in aquaculture environments are relatively scarce. This gap is critical, as aquaculture systems often exhibit greater variability due to factors like feeding schedules, stocking densities, and site-specific management practices. Additionally, DO requirements are species-specific. For example, cold-water fish such as rainbow trout require a minimum of 6 mg/L DO to support normal physiological functions, tissue repair, growth, and reproduction (Svobodova et al., 1993; Mallya, 2007; Abdel-Tawwab et al., 2019). As such, predictive models for DO should be tailored to the specific physiological and environmental needs of the cultured species.

While recent ML-based studies have addressed DO prediction for species such as sea bream (Chatziantoniou et al., 2022), sea bass (Guo et al., 2022), shrimp (Jasmin et al., 2022), and crabs (Chen et al., 2018; Li et al., 2021), no published research to date has focused on developing ML models specifically for DO prediction in intensive rainbow trout farming. Given the species' sensitivity to DO fluctuations, the economic significance of trout aquaculture, and the complexity of intensive farming systems, a targeted modeling approach is clearly warranted.

Accordingly, the objective of this study is to evaluate and compare the performance of four machine learning algorithms, Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Averaged Neural Networks (avNNet) for predicting dissolved oxygen concentration in rainbow trout aquaculture systems. The models utilize multiple water quality parameters, including temperature, pH, turbidity, and conductivity, as input features. The analysis is based on a dataset comprising 120 observations collected from 12 commercial rainbow trout farms located across Türkiye. Ultimately, the study aims to identify the most suitable ML model for accurate and practical DO prediction, thereby supporting improved water quality monitoring and enhancing both fish health and farm productivity in intensive aquaculture.

Materials and Methods

Study Area and Data Collection

This study was conducted in 12 land-based rainbow trout farms of different capacities (10-250 tons/year) operating in Gürpınar and Çatak districts of Van province in Türkiye (Figure 1).

While selecting the enterprises, districts in Van province that intensively produce rainbow trout were selected. At the same time, different factors such as different water quality and climate characteristics in these districts, time taken to reach table weight, and use of spring or stream water were taken into consideration.

Data were collected using a purposive sampling method, aiming to capture diverse production conditions across the study region. Water quality data were obtained from surface water resources utilized by these facilities and are based on a previously published dataset (Demir, 2019). The study utilized water quality measurements and dissolved oxygen (DO) concentrations collected over a five-month period from March to July 2016, covering the peak growing season for rainbow trout in the region. The dataset comprised 120 observations across 12 variables: suspended solids, electrical conductivity, turbidity, nitrate, nitrite, ammonia, ammonium, orthophosphate, pH, water temperature, dissolved oxygen, and total phosphorus.

Water temperature, DO, pH and EI were determined by on-site measurements at the sampling points determined in the study. For measurements of AKM, turbidity, nitrate, nitrite, ammonia, ammonium, orthophosphate, and phosphorus, water samples were taken with 1-1.5 liter PET bottles and measurements were made in Van Yüzüncü Yıl University, Faculty of Fisheries Water Quality Laboratory and Van Provincial Directorate of Agriculture and Forestry Mobile Laboratory. Detailed measurement methods of water samples collected in the study are listed in Table 1. In addition, the average values of the data regarding the water quality parameters used in the study are given in Table 2.

Machine Learning Estimation Models

The dataset was divided into training (80%) and testing (20%). The training data numerical predictors were then preprocessed, which included feature filtering, normalization, centering, and scaling. Four different machine learning methods were used to estimate dissolved oxygen using the other variables as

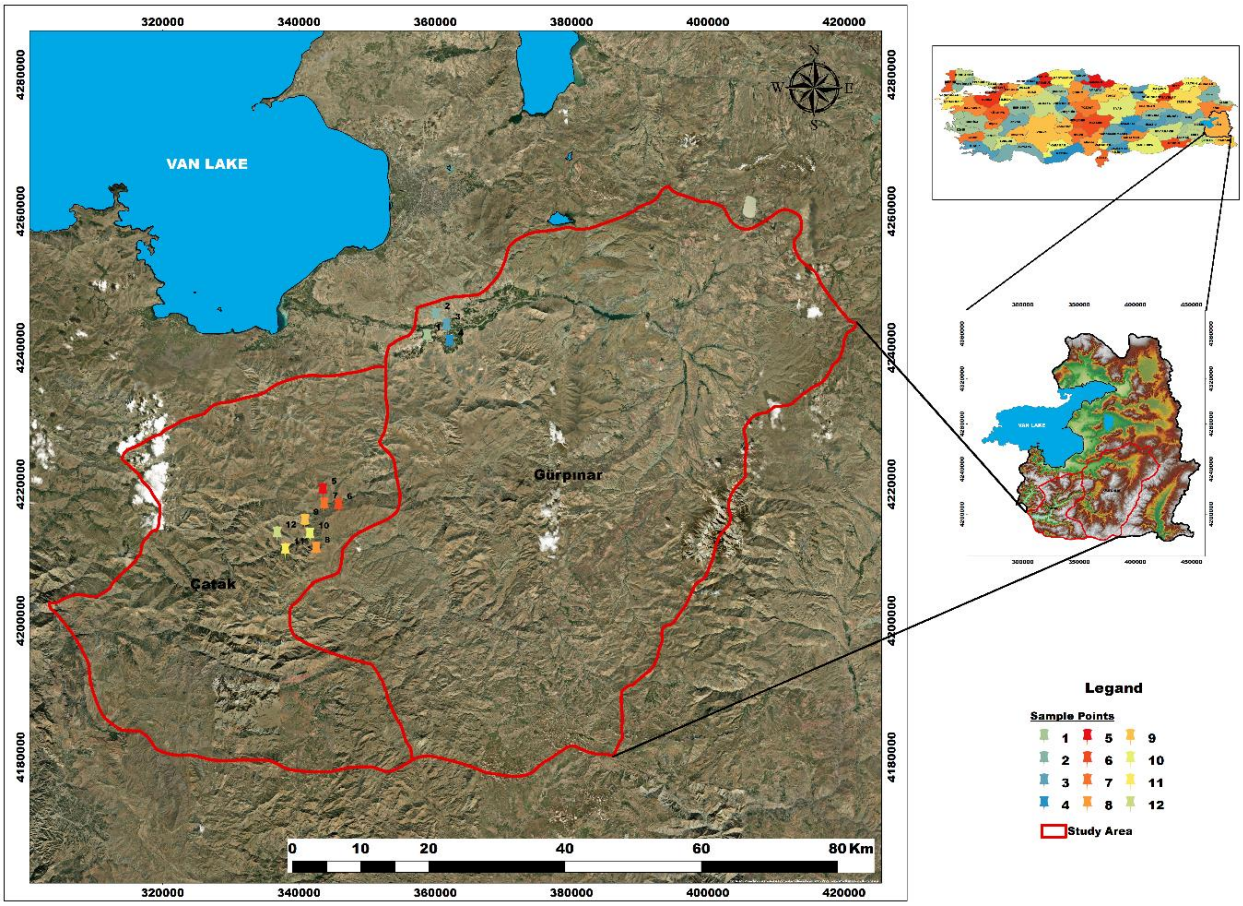

Figure 1. Location map of the study area.

Table 1. Measurement methods for each water quality parameter

Data	Measurement Methods
Water temperature	
Dissolved oxygen	Multimeter device (HACH HQ 40d ve YSI Pro 20)
pH	
Electrical conductivity	
Turbidity	Turbiditimeter (HACH 2100 Q)
Suspended solids	Photometric method (HACH DR 2010 ve 5000 model spectrophotometer)
Nitrate	Cadmium Reduction method (HACH DR 5000 UV/VIS model spectrophotometer)
Nitrite	Diazotization method (HACH DR 5000 UV/VIS model spectrophotometer)
Ammonium	Nessler method (HACH DR LANGE 5000 UV/VIS model spectrophotometer)
Ammonia	
Total phosphorus	PhosVer 3 (Ascorbic Acid) method (HACH DR 5000 UV/VIS model spectrophotometer)
Orthophosphate	

Table 2. Descriptive statistics of the water quality data

Water quality variable	Minimum	Maximum	Avg±sd
Dissolved oxygen (mg/L)	8.00	15.00	10.91±1.04
Water temperature (°C)	8.20	19.30	11.64±3.47
pH	7.73	8.90	8.28±0.26
Electrical conductivity (µS/cm)	231.40	294.80	266.62±15.50
Ammonia (mg/L)	0.12	0.40	0.20±0.05
Ammonium (mg/L)	0.13	0.43	0.21±0.05
Nitrite (mg/L)	0.01	0.04	0.02±0.01
Nitrate (mg/L)	0.13	1.10	0.47±0.19
Total phosphorus (mg/L)	0.09	0.49	0.24±0.09
Orthophosphate (mg/L)	0.03	0.19	0.089±0.03
Turbidity (NTU)	0.27	2.52	1.19±0.45
Suspended solids (mg/L)	13.75	27.50	17.03±2.65

predictors. The first model was Multivariate Adaptive Regression Splines (MARS), which is a non-parametric regression technique that builds a flexible model by partitioning the input space into regions and fitting a linear regression model to each region (Friedman, 1991). It is often used in situations where there are complex, non-linear relationships between the predictors and the target variable and where traditional linear regression models may not be appropriate. They can be particularly useful when there are interactions between the predictor variables that are difficult to model using traditional linear or polynomial regression techniques (Kuhn and Johnson, 2013). The second model was Random Forest (RF), which is an ensemble learning method for classification and regression tasks. It works by building multiple decision trees on different subsets of the input data and combining their predictions to make a final prediction (Breiman, 2001). The third model was Extreme Gradient Boosting (XGBoost), which is a machine learning algorithm that is used for supervised learning tasks such as classification, regression, and ranking. It is an extension of the gradient boosting algorithm that is designed to improve the speed and accuracy of the model (Chen and Guestrin, 2016). The fourth model was model-averaged neural networks (avNNet) which is a type of ensemble method for building predictive models using neural networks. The algorithm combines multiple neural network models by averaging their predictions, thus reducing the variance of the final model and improving its generalization ability (Kuhn and Johnson, 2013). Highly correlated variables can introduce multicollinearity, which can cause the model to overestimate the importance of certain variables or produce unstable estimates of the model coefficients (Chan et al., 2022). Therefore, before the application of machine learning (ML) models, highly correlated variables were eliminated in order to improve the performance of the models. The correlation between predictor variables (Figure 2) were extracted using ggstatsplot R package (Patil, 2021). Models were then implemented using the caret R package (Kuhn, 2022). A 10-folds and repeated 10-times cross-validation method and the tuneGrid argument were used to specify the hyperparameter tuning grid for the models. Variable importance is a measure used to determine the impact of input variables on the output variable in a machine learning model. It can be used to understand the significance of each feature in the model and help with feature selection, model interpretation, and improving model performance (Boehmke and Greenwell, 2019). The VIP R package (Greenwell et al., 2020) was used to extract variable importance for each variable. Models' performance were evaluated by Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percent Error (MAPE), which were generated using the Metrics R package (Hamner and Frasco, 2018). Analysis was performed using the R programming language, version 4.2.2 (RCoreTeam, 2022).

Results

Table 3 presents the performance metrics of four machine learning algorithms, MARS, avNNet, RF, and XGBoost, used to predict dissolved oxygen (DO) concentration in intensive rainbow trout culture based on multiple water quality parameters. All models produced reasonably accurate estimates, but XGBoost achieved the best overall performance, with the lowest Mean Absolute Error (MAE=0.44) and the highest coefficient of determination (R^2=0.78). In comparison, RF achieved an MAE of 0.54 and R^2 of 0.65; avNNet, an MAE of 0.69 and R^2 of 0.45; and MARS, an MAE of 0.72 and R^2 of 0.52. These results indicate that ensemble-based and nonlinear learning approaches, particularly XGBoost, were more effective in modeling the complex interactions among water quality variables.

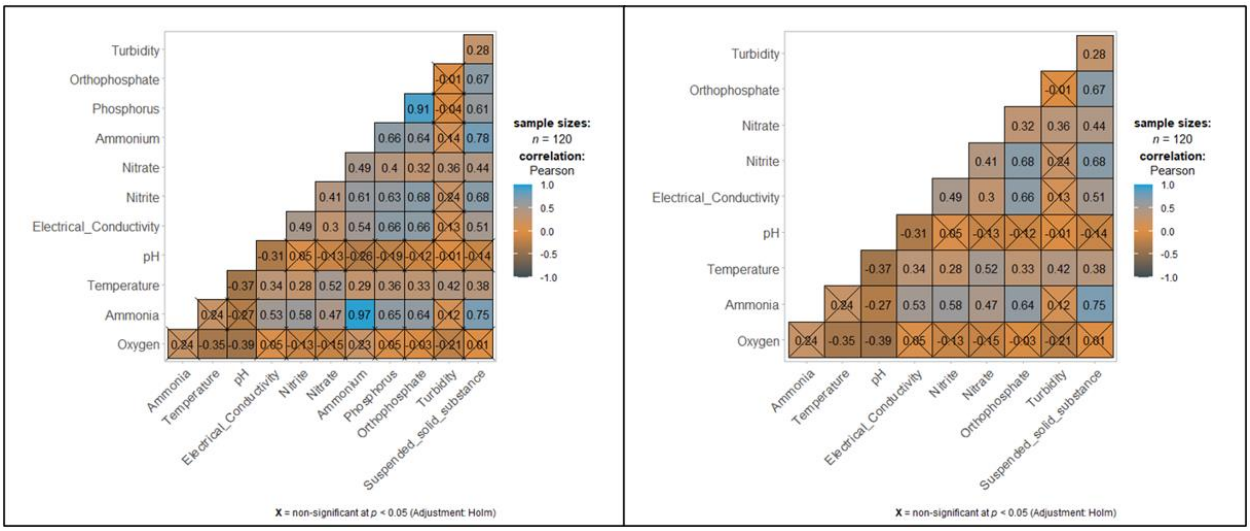

Figure 2. Correlation coefficient between variables, for (A) the entire dataset, and (B) the correlations after removing highly correlated variables.

Table 4 further supports this finding, showing statistically significant differences in MAE between XGBoost and the other models. Specifically, XGBoost outperformed MARS (P=0.017), avNNet (P<0.001), and RF (P=0.001), confirming its superior predictive accuracy in this context.

Figure 3 illustrates the variable importance scores assigned by each model. Temperature and pH consistently emerged as the most influential variables across all four algorithms, underscoring their critical role in determining DO concentration. Electrical conductivity was also identified as a key predictor, ranking third in three of the models and fourth in XGBoost. Meanwhile, the importance rankings for suspended solids, turbidity, ammonia, and orthophosphate varied across models, suggesting algorithm-specific sensitivities. Nitrite and nitrate consistently ranked lowest in importance, indicating minimal predictive contribution to DO estimation in this dataset. Overall, the rankings showed broad consistency, highlighting temperature, pH, and electrical conductivity as dominant predictors across models.

Figure 4 compares the predicted DO values with observed measurements. While the mean predicted values for all models were slightly lower than the observed average (10,93 mg/L), the differences were not statistically significant (P= 0.92). This indicates that the models were able to approximate the actual DO values with a high degree of accuracy. However, it is important to note that although average prediction errors were low, the models may still be affected by occasional outliers or non-systematic deviations. For example, in some individual cases, the prediction error exceeded ±1 mg/L. These outliers, though infrequent, could be significant in real-world aquaculture operations where precise DO control is critical. Additionally, the absence of seasonal variation in the dataset may limit the models' generalizability to other production periods or environmental conditions. These limitations suggest that future studies should explore error distribution, sensitivity analyses, and potentially incorporate temporal and environmental variability to enhance model robustness and practical applicability.

Discussion and Conclusion

Machine learning algorithms have shown promising results for predicting water quality parameters in aquaculture (Chen et al., 2018; Chatziantoniou et al., 2022; Jiang and Yan 2022). In this study, we compared the performance of four machine learning algorithms for predicting dissolved oxygen concentration in rainbow trout farming. The results of this our study demonstrate the potential of machine learning algorithms in predicting dissolved oxygen concentration in rainbow trout culture based on water quality parameters. The results showed that all algorithms provided accurate estimates of dissolved oxygen concentration. The XGBoost algorithm, in particular, showed the highest accuracy predictive performance compared to other algorithms. This could be attributed to its ability to handle complex interactions between variables and its ability to handle missing data (Polikar., 2012; Uddin et al., 2023). In addition, the incorporation of multiple water quality parameters into the models enhanced the accuracy of the predictions. After XGBoost, the algorithms with the highest performance were RF, avNNet, and MARS. The findings of this study suggest that the XGBoost algorithm could be a useful instrument for predicting the concentration of dissolved oxygen in rainbow trout culture and enhancing fish health and production.

Our findings are consistent with other studies that demonstrate XGBoost's superior performance in predicting water quality parameters compared to alternative models. For instance, Khoi et al. (2022) evaluated twelve ML models and found extreme gradient boosting (XGBoost) to have the highest accuracy (R^2=0.989 and RMSE=0.107) in predicting the Water Quality Index. Similarly, in previous studies conducted to predict water quality using classification methods, it was found that the XGBoost model outperformed other models in terms of performance. Yusri et al. (2022) reported that XGBoost outperformed Support Vector Machine (SVM) in water quality prediction, with an accuracy of 94% compared to SVM's 67%. XGBoost also exhibited a lower misclassification

Table 3. Performance matrics of the ML models

Model	R^2	MAE	RMSE	MAPE
MARS	0.52	0.72	0.87	0.07
avNNet	0.45	0.69	0.84	0.06
RF	0.65	0.54	0.70	0.05
XGBoost	0.78	0.44	0.58	0.04

R^2:Coefficient of Determination, MAE: Mean Absolute Error, RMSE:Root Mean Squared Error, MAPE:Mean Absolute Percentage Error.

Table 4. Compare performance of the models based on MAE (p-values)

Model	MARS	avNNet	RF	XGBoost
MARS	***	2.766e-09	0.06547	0.01713
avNNet	2.766e-09	***	< 2.2e-16	3.418e-09
RF	0.06547	< 2.2e-16	***	5.047e-10
XGBoost	0.01713	3.418e-09	5.047e-10	***

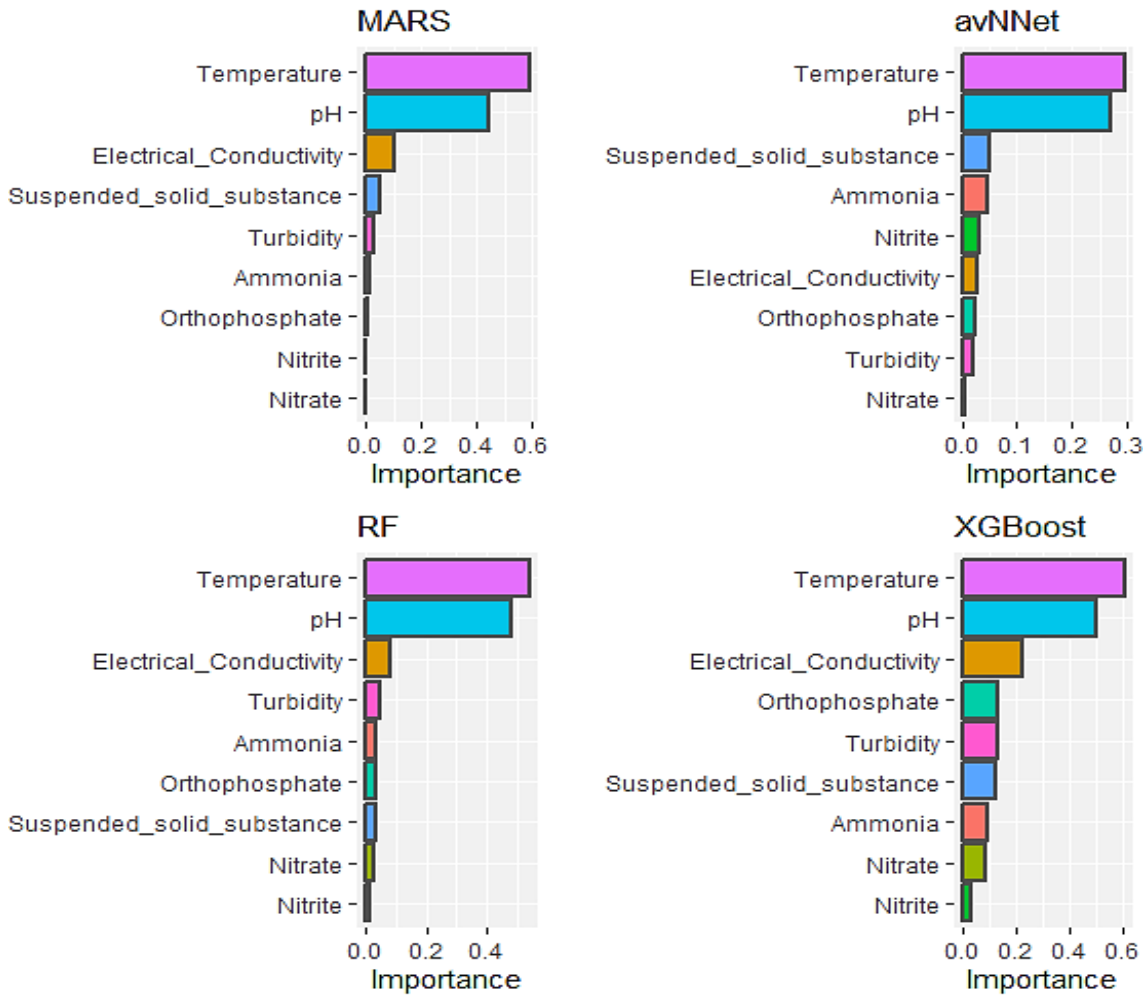

Figure 3. Variable importance for the machine learning models.

error (6% vs. 33%) and superior results in 5-fold validation. Another study by Uddin et al. (2023) evaluated six commonly used algorithms (XGBoost, SVM, ANN, NB, KNN and DT) for water quality prediction, with XGBoost and KNN achieving the highest accuracy (99.9% and 100% correct, respectively).

Furthermore, some studies have reported that hybrid models combining the XGBoost algorithm with other algorithms also show good performance. In a study by Wu et al. (2021), a novel hybrid model (XGBoost-ISSA-LSTM) was proposed to predict dissolved oxygen (DO) levels in river crab pond culture. The hybrid model demonstrated remarkable accuracy, with a 1-hour prediction having an RMSE of 0.5571, MAE of 0.2572, and R2 of 0.9276. Additionally, the 24-hour prediction maintained strong performance, with an RMSE of 0.6310, MAE of 0.4562, and R2 of 0.9082, highlighting the effectiveness of the XGBoost-ISSA-LSTM hybrid model in accurately predicting DO levels in river crab pond culture.

Lu and Ma (2020) introduced two hybrid decision tree-based models, CEEMDAN-XGBoost and CEEMDAN-RF, with the smallest average MAPEs (3.90% and 3.71%), indicating superior overall prediction performance for predicting six water quality indicators, including water temperature, dissolved oxygen, pH value, specific conductance, turbidity, and fluorescent dissolved organic matter. Similarly, Tiyasha et al., (2021) suggested four tree-based predictive models: RF, Ranger, cForest, and XGBoost, compared with algorithms MARS and Boruta-GA. Additionally, four feature selector techniques (GA, Boruta, XGBoost, and MARS) were employed to determine the optimal independent variables for DO forecasting. The study showed good performance for all predictive approaches based on features selected by the algorithms MARS and XGBoost. Moreover, the XGBoost predictive technique exhibited the best performance when combined with MARS and XGBoost algorithms in terms of various statistical criteria.

In our study, MARS had the worst performance, followed by avNNet and RF. Yet, other research has shown these algorithms to be effective in predicting dissolved oxygen concentration. For instance, In a 2018 study by Heddam and Kisi, the MARS algorithm was found to outperform both Least Squares Support Vector Machine (LSSVM) and M5 Model Tree (M5T) in predicting DO concentration from water quality parameters, achieving approximately 8.22% to 11.05% lower RMSE and 23.50% to 25.05% lower MAE values.

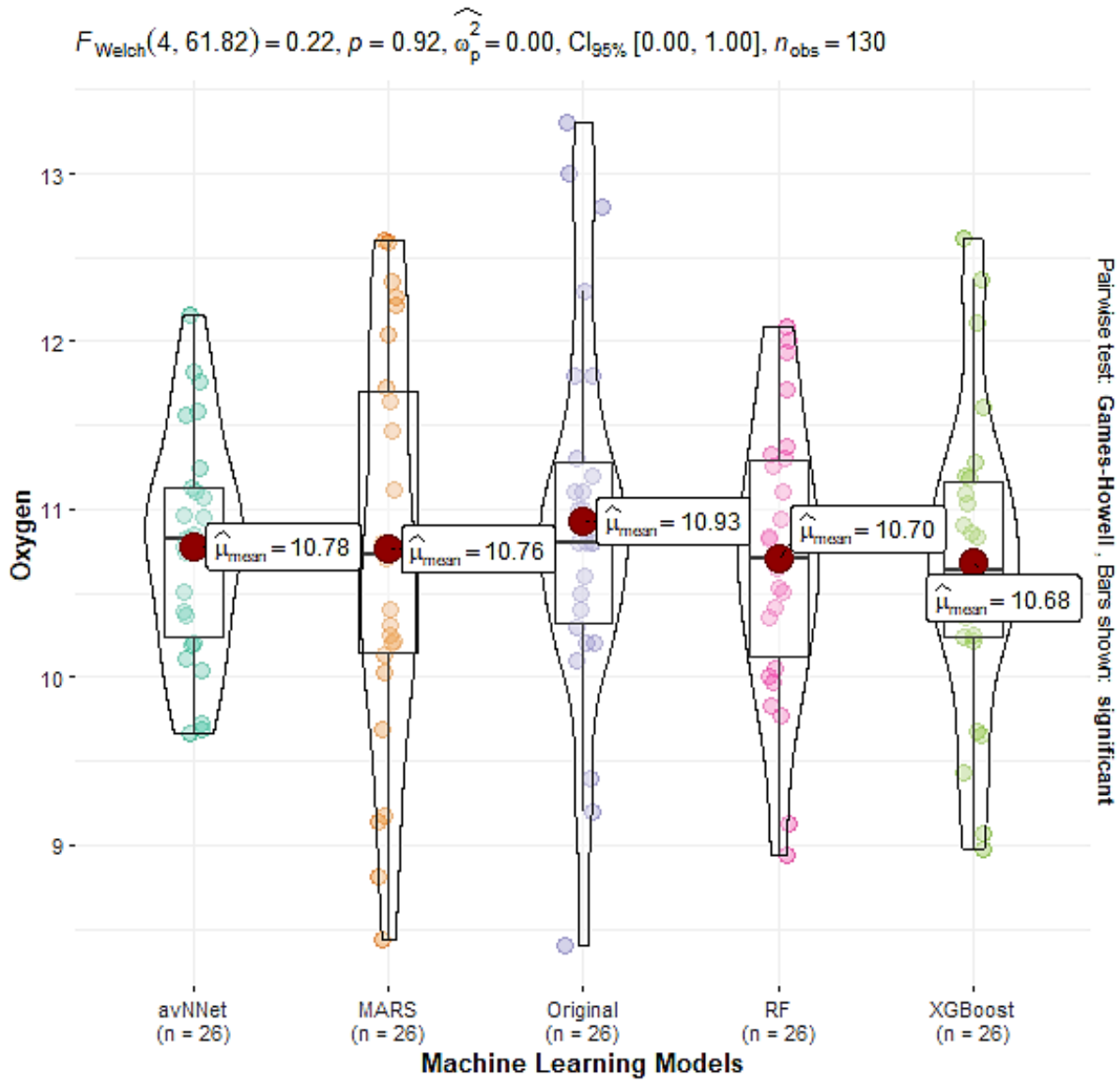

Figure 4. Differences between original (observed) dissolved oxygen concentration andthe predicted dissolved oxygen concentration obtained from different machine learning models.

Similarly, Jasmin et al. (2022) predicted DO and biofloc amounts in shrimp culture using Random Forest, Adaboost, and Deep Neural Networks. They found that the RF outperformed the others, with R^2=0.7381, MAE=0.187, and RMSE=0.235.

The consistent identification of temperature and pH as the most important predictors of dissolved oxygen (DO) concentration across all four models (MARS, avNNET, RF, and XGBoost) highlights their critical role in regulating aquatic oxygen dynamics. Temperature significantly influences the solubility of gases in water; as temperature increases, the solubility of oxygen decreases, thereby lowering DO availability (Wetzel, 2001; Boyd & Tucker, 2012). Additionally, higher temperatures intensify microbial metabolism and biological oxygen demand, accelerating DO depletion (Chapra, 2008). pH, on the other hand, can regulate microbial processes and chemical equilibria, such as the activity of nitrifying bacteria and the balance between ammonium and ammonia, thereby indirectly increasing or decreasing DO levels (El-Gohary et al., 1995; Tchobanoglous et al., 2003). Although the machine learning models used in this study are data-driven and non-mechanistic, they are capable of capturing and reflecting complex interactions among input variables. Ensemble tree-based algorithms such as Random Forest and XGBoost are particularly well-known for their ability to learn nonlinear and high-order interactions between features (Hastie et al., 2009; Chen & Guestrin, 2016). The high importance scores assigned to both pH and temperature suggest that these models may implicitly account not only for the individual effects of these variables but also for their joint influence on DO dynamics. Such synergistic or antagonistic interactions are especially relevant in aquaculture systems, where environmental variables often fluctuate simultaneously.

To further interpret these interactions, model-agnostic tools such as Partial Dependence Plots (PDPs), Accumulated Local Effect (ALE) plots, or SHAP (SHapley Additive Explanations) values can be employed to quantify and visualize the effects of individual predictors and their combinations on model outcomes (Lundberg & Lee, 2017; Christoph, 2020). These tools can enhance the interpretability of complex models and support more informed decision-making in aquaculture management.

In conclusion, the present study examined the performance of four machine learning algorithms in predicting dissolved oxygen concentrations in intensive rainbow trout culture based on water quality parameters. The results indicated that all four algorithms produced accurate estimates of dissolved oxygen levels, with the XGBoost algorithm demonstrating the highest performance. These findings suggest that machine learning algorithms hold great promise as valuable tools for predicting dissolved oxygen levels in rainbow trout culture, which in turn could have significant implications for improving fish health and production in aquaculture systems. The study's results also align with previous research in the field, which has demonstrated the superiority of XGBoost and other machine learning algorithms in predicting water quality parameters. The ability of XGBoost to handle complex interactions between variables and address missing data may explain its superior performance in this context. Additionally, the incorporation of multiple water quality parameters into the models further enhanced the accuracy of the predictions. Moreover, the results highlighted the potential of hybrid models combining XGBoost with other algorithms to achieve even better performance in certain applications. It is important to note that while XGBoost exhibited the highest performance in this study, other algorithms, such as MARS, and Random Forest, showed promising results as well, in line with previous research findings. Future research should explore the applicability of these machine learning algorithms in diverse aquaculture systems and investigate the potential of additional algorithms for predicting dissolved oxygen levels in aquaculture. Overall, this study contributes to the growing body of literature supporting the use of machine learning in aquaculture and provides valuable insights into optimizing dissolved oxygen prediction models for sustainable fish farming practices.

Ethical Statement

Not applicable.

Funding Information

The authors did not receive support from any organization for the submitted work.

Author Contribution

Boran Karataş: Conceptualization, Data curation, Formal analysis, Methodology, Resources, Validation, Visualization, Writing- original draft, Writing- review & editing. Cihan Çakmakçı: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing- review & editing. Elif Sena YÜCEL: Conceptualization, Data curation, Investigation, Methodology, Resources, Validation, Visualization, Writing- review & editing. Muhammet Demir: Data acquisition, Resources, Writing- review & editing. Fazıl Şen: Data acquisition, Resources, Writing- review & editing. All authors contributed to the article and approved the submitted version.

Conflict of Interest

The author(s) declare that they have no known competing financial or non-financial, professional, or personal conflicts that could have appeared to influence the work reported in this paper.

References

Abdel-Tawwab, M., Monier, M. N., Hoseinifar, S. H., & Faggio, C. (2019). Fish response to hypoxia stress: growth, physiological, and immunological biomarkers. *Fish physiology and biochemistry*, *45*, 997-1013. https://doi.org/10.1007/s10695-019-00614-9

Ahmed, A. N., Othman, F. B., Afan, H. A., Ibrahim, R. K., Fai, C. M., Hossain, M. S., & Elshafie, A. (2019). Machine learning methods for better water quality prediction. *Journal of Hydrology*, *578*, 124084. https://doi.org/10.1016/j.jhydrol.2019.124084

Arabacı, M., Karataş, B., & Akkuş, M. (2020). Gökkuşağı Alabalığı Anaç ve Larvalarda Görülen Gaz Kabarcığı Hastalığı. *Journal of Anatolian Environmental and Animal Sciences*, *5*(3), 304-308. https://doi.org/10.35229/jaes.737359

Boehmke, B., & Greenwell, B. M. (2019). *Hands-on machine learning with R*. Chapman and Hall/CRC. https://doi.org/10.1201/9780367816377

Boyd, C. E., & Tucker, C. S. (2012). *Pond aquaculture water quality management*. Springer Science & Business Media.

Breiman, L., (2001). *Random Forests. Machine Learning* 45(1):5-32. https://doi.org/10.1023/A:1010933404324

Cao, W., Huan, J., Liu, C., Qin, Y., & Wu, F. (2019). A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multi-scale feature extraction. *Aquacultural Engineering*, *84*, 50-59. https://doi.org/10.1016/j.aquaeng.2018.12.003

Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics*, *10*(8), 1283. https://doi.org/10.3390/math10081283

Chapra, S. C. (2008). *Surface water-quality modeling*. Waveland press.

Chatziantoniou, A., Spondylidis, S. C., Stavrakidis-Zachou, O., Papandroulakis, N., & Topouzelis, K. (2022). Dissolved oxygen estimation in aquaculture sites using remote sensing and machine learning. *Remote Sensing Applications: Society and Environment*, *28*, 100865. https://doi.org/10.1016/j.rsase.2022.100865

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794). https://doi.org/10.1145/2939672.2939785

Chen, Y., Yu, H., Cheng, Y., Cheng, Q., & Li, D. (2018). A hybrid intelligent method for three-dimensional short-term prediction of dissolved oxygen content in aquaculture. *PLoS One*, *13*(2), e0192456. https://doi.org/10.1371/journal.pone.0192456

Christoph, M. (2020). Interpretable machine learning: A guide for making black box models explainable.

D'Agaro, E., Gibertoni, P., & Esposito, S. (2022). Recent trends and economic aspects in the rainbow trout (*Oncorhynchus mykiss*) sector. *Applied Sciences*, *12*(17), 8773. https://doi.org/10.3390/app12178773

Demir, M. (2019). *Investigation of the inflow and outflow water properties of trout facilities with different capacities installed on land in Van province*. M.Sc. Thesis, Van Yüzüncü Yıl University, Institute of Natural and Applied Sciences, Department of Aquaculture Engineering. Van Turkey.

El-Gohary, F., El-Hawarry, S., Badr, S., & Rashed, Y. (1995). Wastewater treatment and reuse for aquaculture. *Water Science and Technology*, *32*(11), 127-136. https://doi.org/10.1016/0273-1223(96)00126-6

FAO., (2020). The State of World Fisheries and Aquaculture 2020; FAO: Rome, Italy; ISBN 978-92-5-132692.

FAO., (2022). The State of World Fisheries and Aquaculture. Towards Blue Transformation. 2022. Available online: www.fao.org (accessed 4 February 2023).

Friedman, J.H. (1991). Multivariate Adaptive Regression Splines. The Annals of Statistics 19(1):1-67. https://doi.org/10.1214/aos/1176347963

Greenwell, B. M., Boehmke, B. C., & Gray, B. (2020). Variable Importance Plots-An Introduction to the vip Package. *R J.*, *12*(1), 343. https://doi.org /10.32614/rj-2020-013

Guo, J., Dong, J., Zhou, B., Zhao, X., Liu, S., Han, Q., & Hassan, S. G. (2022). A hybrid model for the prediction of dissolved oxygen in seabass farming. *Computers and Electronics in Agriculture*, *198*, 106971. https://doi.org/10.1016/j.compag.2022.106971

Hamner, B., & Frasco. M. (2018). Metrics: Evaluation Metrics for Machine Learning. R package version 0.1.4. https://CRAN.R-project.org/package=Metrics.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction 2nd edn (Springer Science & Business Media). https://doi.org/10.1007/978-0-387-21606-5

Heddam, S., & Kisi, O. (2018). Modelling daily dissolved oxygen concentration using least square support vector machine, multivariate adaptive regression splines and M5 model tree. *Journal of Hydrology*, *559*, 499-509. https://doi.org/10.1016/j.jhydrol.2018.02.061

Jasmin, S. A., Ramesh, P., & Tanveer, M. (2022). An intelligent framework for prediction and forecasting of dissolved oxygen level and biofloc amount in a shrimp culture system using machine learning techniques. *Expert Systems with Applications*, *199*, 117160. https://doi.org/10.1016/j.eswa.2022.117160

Jiang, X., Dong, S., Liu, R., Huang, M., Dong, K., Ge, J., & Zhou, Y. (2021). Effects of temperature, dissolved oxygen, and their interaction on the growth performance and condition of rainbow trout (*Oncorhynchus mykiss*). *Journal of Thermal Biology*, *98*, 102928. https://doi.org/10.1016/j.jtherbio.2021.102928

Jiang, Y., & Yan, F. (2022). Aquaculture Prediction Model Based on Improved Water Quality Parameter Data Prediction Algorithm under the Background of Big Data. *Journal of Applied Mathematics*, *2022*(1), 2071360. https://doi.org/10.1155/2022/2071360

Khoi, D. N., Quan, N. T., Linh, D. Q., Nhi, P. T. T., & Thuy, N. T. D. (2022). Using machine learning models for predicting the water quality index in the La Buong River, Vietnam. *Water*, *14*(10), 1552. https://doi.org/10.3390/w14101552

Kuhn, M. (2022). caret: Classification and Regression Training_. R package version 6.0-93, https://CRAN.R-project.org/package=caret.

Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling. Springer.

Li, C., Li, Z., Wu, J., Zhu, L., & Yue, J. (2018). A hybrid model for dissolved oxygen prediction in aquaculture based on multi-scale features. *Information processing in agriculture*, *5*(1), 11-20. https://doi.org/10.1016/j.inpa.2017.11.002

Li, T., Lu, J., Wu, J., Zhang, Z., & Chen, L. (2022). Predicting aquaculture water quality using machine learning approaches. *Water*, *14*(18), 2836. https://doi.org/10.3390/w14182836

Li, W., Wu, H., Zhu, N., Jiang, Y., Tan, J., & Guo, Y. (2021). Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Information Processing in Agriculture*, *8*(1), 185-193. https://doi.org/10.1016/j.inpa.2020.02.002

Lu, H., & Ma, X. (2020). Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere*, *249*, 126169. https://doi.org/10.1016/j.chemosphere.2020.126169

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Mallya, Y.J. (2007). The effects of dissolved oxygen on fish growth in aquaculture. The United Nations University Fisheries Training Programme, Final Project.

Moon, S., Kim, H., You, Y. H., Kim, C. H., & Hwang, I. (2022). Deep neural network for beam and blockage prediction in 3GPP-based indoor hotspot environments. *Wireless Personal Communications*, *124*(4), 3287-3306. https://doi.org/10.1007/s11277-022-09513-4

Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A., & Al-Shamma'a, A. (2022). Water quality classification using machine learning algorithms. *Journal of Water Process Engineering*, *48*, 102920. https://doi.org/10.1016/j.jwpe.2022.102920

Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach. Journal of Open Source Software 6(61):3167. https://doi.org/10.21105/joss.03167

Polikar, R. (2012). Ensemble Learning; in C. Zhang and Y. Ma (eds.), Ensemble Machine Learning: Methods and Applications, © Springer Science+Business Media, LLC Springer, New York, NY. https://doi.org/10.1007/978-1-4419-9326-7_1

RCoreTeam., (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ross, A.C., & Stock, C.A. (2019). An assessment of the predictability of column minimum dissolved oxygen concentrations in Chesapeake Bay using a machine learning model. Estuarine, Coastal and Shelf Science, 221, 53-65. https://doi.org/10.1016/j.ecss.2019.03.007

Royer, E., Faccenda, F., & Pastres, R. (2021). Estimating oxygen consumption of rainbow trout (*Oncorhynchus mykiss*) in a raceway: A Precision Fish Farming approach. *Aquacultural Engineering,* 92, 102141. https://doi.org/10.1016/j.aquaeng.2020.102141

Sun, L., Wu, Y., Li, D., Wang, B., Sun, X., & Luo, B. (2021). Prediction model for the number of crucian carp hypoxia based on the fusion of fish behavior and water environment factors. *Computers and Electronics in Agriculture*, 189, 106386. https://doi.org/10.1016/j.compag.2021.106386

Svobodova, Z., Richard, L., Jana, M., & Blanka, V. (1993). Water quality and fish health. EIFAC technical paper (54) Food & Agriculture Org.

Tchobanoglus, G., Burton, F., & Stensel, H. D. (2003). Wastewater engineering: treatment and reuse. *American Water Works Association. Journal*, *95*(5), 201.

Tiyasha, T., Tung, T.M., Bhagat, S.K., Tan, M.L., Jawad, A.H., Mohtar, W.H.M.W., & Yaseen, Z.M. (2021). Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models. *Marine pollution bulletin,* 170, 112639. https://doi.org/10.1016/j.marpolbul.2021.112639

Uddin, M.G., Nash, S., Rahman, A., & Olbert, A.I. (2023). Performance analysis of the water quality index model for predicting water state using machine learning techniques. *Process Safety and Environmental Protection,* 169, 808-828. https://doi.org/10.1016/j.psep.2022.11.073

Welker, T.L., Overturf, K., & Abernathy, J. (2019). Effect of aeration and oxygenation on growth and survival of rainbow trout in a commercial serial-pass, flow-through raceway system. *Aquaculture Reports,* 14, 100194. https://doi.org/10.1016/j.aqrep.2019.100194

Wetzel, R. G. (2001). *Limnology: lake and river ecosystems*. gulf professional publishing. 3rd ed., Academic Press Elsevier.

Wu, Y., Sun, L., Sun, X., & Wang, B. (2021). A hybrid XGBoost-ISSA-LSTM model for accurate short-term and long-term dissolved oxygen prediction in ponds. *Environmental Science and Pollution Research,* 1-18. https://doi.org/10.1007/s11356-021-17020-5

Yin, L., Fu, L., Wu, H., Xia, Q., Jiang, Y., Tan, J., & Guo, Y. (2021). Modeling dissolved oxygen in a crab pond. *Ecological Modelling*, 440, 109385.c https://doi.org/10.1016/j.ecolmodel.2020.109385

Yusri, H.I.H., Hassan, S.L.M., Halim, I.S.A., & Abdullah, N.E. (2022, July). *Water Quality Classification Using SVM And XGBoost Method. In 2022 IEEE* 13th Control and System Graduate Research Colloquium (ICSGRC) (pp. 231-236). IEEE. https://doi.org/10.1109/ICSGRC55096.2022.9845143