

## PAPER

# A Smart Decision Support System for Floating Net Cage Site Selection Based on Water Quality and Machine Learning

Dedy Hermanto<sup>1,2</sup> ,  
Deris Stiawan<sup>2</sup> ,  
Bhakti Yudho Suprpto<sup>2</sup> ,  
Mohd. Yazid Idris<sup>3</sup> ,  
Rahmat Budiarto<sup>4</sup> 

<sup>1</sup>Universitas Multi  
Data Palembang,  
Palembang, Indonesia

<sup>2</sup>Universitas Sriwijaya,  
Palembang, Indonesia

<sup>3</sup>Universiti Teknologi Malaysia,  
Johor Bahru, Malaysia

<sup>4</sup>Al-Baha University, Al-Aqiq,  
Saudi Arabia

[dedy@mdp.ac.id](mailto:dedy@mdp.ac.id)

## ABSTRACT

Fish is a vital food source with high nutritional value and a growing global demand. However, natural fishing can no longer meet consumption needs, prompting a shift toward sustainable aquaculture such as floating net cages. The success of these systems depends greatly on water quality, as sudden changes may cause mass fish mortality and economic loss. This study applies five machine-learning algorithms: 1) support vector machine (SVM), 2) K-nearest neighbor (KNN), 3) Naïve Bayes, 4) AdaBoost, and 5) random forest—to classify water quality using three features: 1) dissolved oxygen (DO), 2) pH, and 3) temperature. The dataset, publicly available at <https://doi.org/10.5281/zenodo.15600660>, shows DO as the most influential feature (48.15%), followed by pH (40.23%) and temperature (11.62%). Random forest achieved the highest accuracy (99.96%) with the lowest errors (MSE = 0.0004, RMSE = 0.0196, R<sup>2</sup> = 0.9981). The trained model was embedded in an IoT device coded in C for real-time water-quality monitoring and site recommendations. The results confirm that combining IoT and machine learning offers an intelligent and efficient solution for adaptive, sustainable fish-farming systems.

## KEYWORDS

fish, floating net cage, Internet of Things (IoT), location-based, machine learning, sensors, water quality index

## 1 INTRODUCTION

The public's demand for fish consumption continues to rise as awareness of the importance of high-quality nutritional intake grows. Fish is known as one of the most nutritious and relatively accessible sources of animal protein, making it a highly strategic food commodity. A report by the Food and Agriculture Organization (FAO) [1] notes that the global demand for fishery products continues to grow year after year. Specifically, the Asian region contributes over 70% of global production.

Hermanto, D., Stiawan, D., Suprpto, B. Y., Idris, M. Y., Budiarto, R. (2025). A Smart Decision Support System for Floating Net Cage Site Selection Based on Water Quality and Machine Learning. *International Journal of Online and Biomedical Engineering (iJOE)*, 21(14), pp. 58–75. <https://doi.org/10.3991/ijoe.v21i14.59301>

Article submitted 2025-09-02. Revision uploaded 2025-10-14. Final acceptance 2025-10-15.

© 2025 by the authors of this article. Published under CC-BY.

In Indonesia, fish is not only a staple food for the population but also a crucial pillar in maintaining national food security.

To meet these needs, various countries have issued regulations and standard guidelines regarding water quality, both for drinking water and fisheries [2], [3], [4]. This underscores that water quality is a fundamental factor that cannot be ignored in the success of fish farming. Indonesia also places great importance on this issue through various government policies. [5], including the establishment of water quality standards by the Ministry of Marine Affairs and Fisheries. Good water quality is crucial for production success; conversely, drastic changes in water conditions, such as temperature and salinity fluctuations, as well as dissolved oxygen (DO) levels, can cause stress in fish, leading to mass mortality. This condition undoubtedly causes significant economic losses for fishermen and poses a threat to the availability of fish-based food.

In practice, traditional fishermen still heavily rely on catching fish directly from nature. However, this method often fails to meet the high demand, notably when it lacks sustainable management. Therefore, fish farming becomes the main alternative. One of the most widely used cultivation methods is the floating net cage. [6], [7], [8], which is a cultivation system that allows fish to be raised in open water using a box-shaped container lined with netting and placed in rivers or the sea. Although this method has proven effective in increasing production, significant challenges remain, particularly regarding the monitoring of water quality indices, which significantly affect the productivity and sustainability of floating net cages.

Changes in water quality, whether due to natural or anthropogenic factors, can disrupt fish growth in floating net cages. Therefore, real-time monitoring of water quality indices has become an urgent need. Technological advancements offer solutions through the Internet of Things (IoT) [9], [10], [11], enabling continuous monitoring of water conditions with directly integrated data. Furthermore, the application of machine learning to IoT systems can provide predictive analysis and intelligent recommendations for fishermen. By incorporating knowledge of water quality indices into IoT devices, this system can help determine the optimal location for floating cage placement and provide early warnings of potential water quality decline. [12], [13]. The integration of IoT and machine learning thus has the potential to be a strategic approach in supporting more adaptive, efficient, and sustainable fish farming.

This study will discuss floating net cages integrated with IoT devices, incorporating machine learning with the best model resulting from training. This is useful for obtaining water condition monitoring results at the floating net cage placement site, based on the best location determined by sensors installed on-site. The best approach will be taken with the hope of helping fishermen monitor in real time, without distance limitations, and provide recommendations for the best floating net cage location placement based on water quality index predictions using machine learning.

The structure of this paper is as follows: Section 1 discusses the general problems and the specific issues addressed in this study; Section 2 covers the literature related to the study being conducted; Section 3 discusses the methodology developed to carry out this study; Section 4 presents the experimental process, results, and evaluation of the model generated from the research; Section 5 discusses the implementation of the model used in embedded devices and communication integration at various locations; and Section 6 provides the conclusion of the research.

## 2 RELATED WORK

The use of technology in the fisheries sector is becoming increasingly widespread, particularly to help fishermen simplify and improve the effectiveness of

aquaculture processes. This transformation is inextricably linked to innovation in modern cultivation methods, particularly the use of floating net cages [8], [14], [15]. As time goes on, technology is not only serving as an aid but is also becoming an integral component in the management of floating net cages. Innovation is being implemented in both the physical design aspects of the cages and the integration of sensor-based monitoring devices and intelligent systems that support aquaculture activities.

Every research and development effort contributes uniquely to the advancement of floating net cage methods. For example, some studies focus on strengthening the physical structure of the cages to make them more resilient to environmental conditions. In contrast, others focus on integrating an IoT-based water quality monitoring system. This development demonstrates that floating net cages are no longer just a cultivation container, but an integral part of an evolving technological ecosystem designed to enhance productivity and sustainability.

This study aims to contribute to the ongoing advancement of intelligent aquaculture systems, with a specific focus on water quality monitoring as a key factor determining the success of floating net cage operations. Through an IoT-based framework enhanced with machine learning models, this study proposes a real-time solution for site selection and continuous monitoring of the floating net cage. The proposed approach not only strengthens fish farming practices in the field but also supports global efforts to ensure the sustainable and adaptive availability of fish-based food resources. By leveraging artificial intelligence technologies, this work contributes to the technological development and operational efficiency of floating net cage systems. Furthermore, the proposed methodology builds upon insights from several previous studies, summarized in a comparative analysis. Table 1 provides an overview of related research focusing on water quality index assessment and machine learning-based classification, including implementations that integrate IoT and machine learning to determine water quality in floating net cages using binary classification methods. In addition, this study addresses existing limitations in earlier studies, such as the lack of real-time integration and limited sensor utilization, by developing a fully connected embedded system capable of on-site decision support. The results of this development are expected to enhance prediction accuracy, reduce manual monitoring errors, and improve aquaculture sustainability through intelligent automation.

**Table 1.** Summary of ML-based on water quality for fishery

Ref	Algorithm	Classification	Model Selection
[16]	LR, DT, RF, KNN, SVM, NB	Binary	Split Dataset
[17]	AB, RF, DT, KNN, SVM	Multi-Class	Split Dataset
[18]	RF, MLR	Binary	Split Dataset
[19]	RF, KNN, DT	Binary	Split Dataset
[20]	CNN	Multi-Class	K-Folds
[21]	SVM, MLP, RF, XGBOOST, LSTM	Binary	Split Dataset
[22]	LR, RF, SVM, KNN	Binary	Split Dataset
[23]	RF, SVM	Binary	Split Dataset
[24]	BPNN, RBFNN, SVM, LSSVM	Multi-Class	K-Folds
[25]	PCA	Binary	K-Folds

### 3 METHODOLOGY

This section outlines the initial phase of the research, which began with the design and development of an embedded device serving as the primary instrument for collecting the dataset. The device was specifically engineered to measure key water quality parameters relevant to determining the optimal placement of floating net cages. The development of this embedded system represented a critical stage in the research process, as the accuracy, stability, and reliability of the data collected directly influenced the validity of subsequent analyses and model development.

#### 3.1 Embedded devices

The initial phase of this study began with the design and construction of an embedded device functioning as a dataset acquisition system. The device employs an ESP32 processing module as the central control unit [26], which is integrated with three primary sensors for monitoring water quality: DO [27], pH [28], and water temperature [29]. The ESP32 was selected due to its efficient data processing capability and built-in wireless connectivity, which enables real-time data transmission. These three sensors are directly interfaced with the ESP32 microcontroller, allowing simultaneous acquisition and synchronization of water quality parameters. An illustration of the relationships, communication, and operation of the devices built in this study is presented in Figure 1.

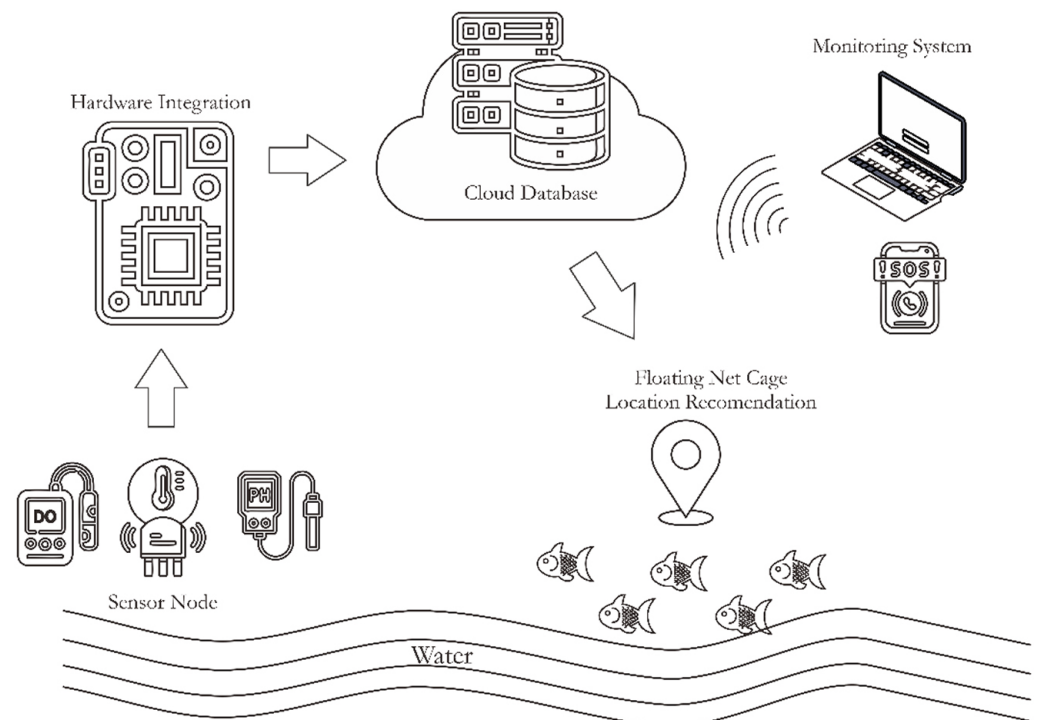


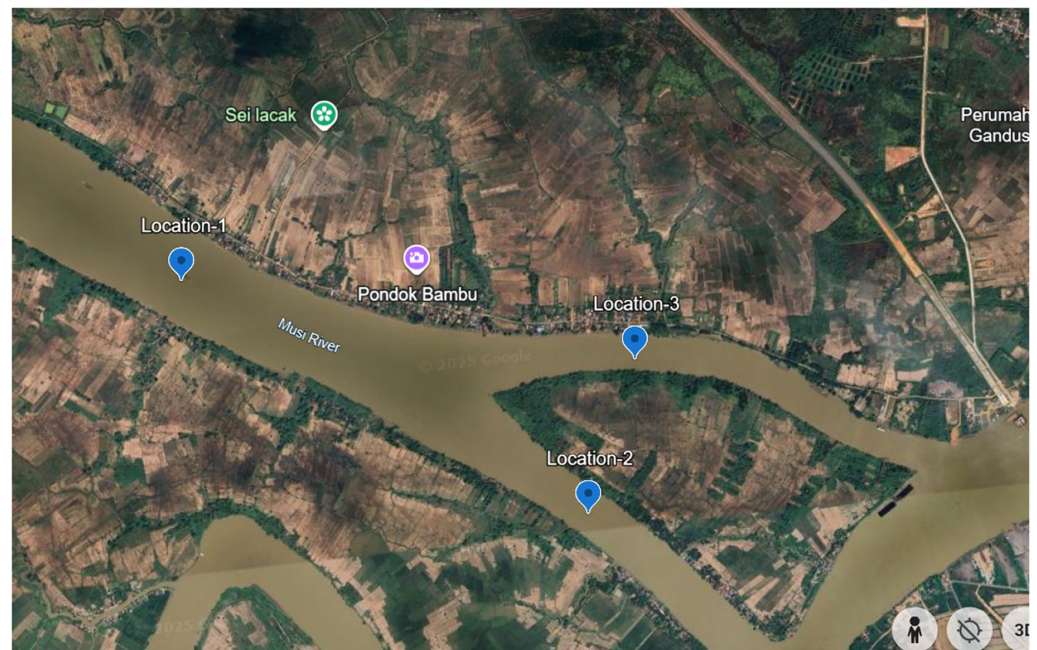
Fig. 1. Controller and system operation

After all device components were successfully integrated and their connectivity verified, the next stage involved preparing the device for deployment at the designated field site for dataset collection. This preparation included refining the physical

design of the device to ensure durability under aquatic environmental conditions and verifying that it could operate reliably throughout the entire data acquisition period. Additional adjustments were made to enhance waterproofing, power stability, and sensor protection, thereby ensuring continuous and accurate data recording during field operation. The dataset used in this study was collected as part of a previous work, where the field data acquisition process was comprehensively described and validated. The current study focuses on the analysis, modelling, and system implementation stages using the previously collected dataset [26].

### 3.2 Dataset

The dataset was collected directly from the research site, located at  $3^{\circ} 01' 58.1''$  S,  $104^{\circ} 39' 17.6''$  E, as illustrated in Figure 2. Field data were acquired through continuous monitoring and stored on local storage media for subsequent preprocessing. The preprocessing stage involved data cleaning, validation, and formatting to ensure consistency and accuracy before analysis. Following the cleaning phase, the refined dataset was used as input for machine learning models to generate recommendations for water quality conditions. The dataset comprises three primary features: DO, pH, and water temperature, as well as one output label representing the water condition category, classified as either good or not good.



**Fig. 2.** Dataset acquisition location

As a contribution to scientific transparency and open research practices, the dataset generated in this study has been made publicly available and can be accessed via the following DOI link: <https://doi.org/10.5281/zenodo.1560066>. The dataset distribution is presented in Figures 3, 4, and 5, which illustrate the value distributions of the three main features used in this study: DO, pH, and water temperature. These visualizations provide an overview of the data characteristics and ensure reproducibility for future research based on the same dataset.

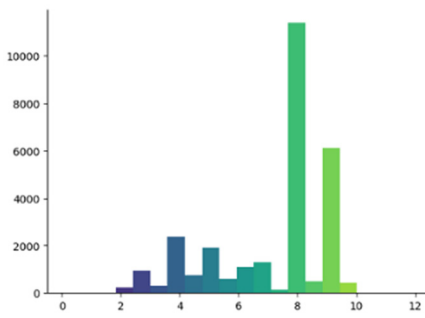


Fig. 3. DO data

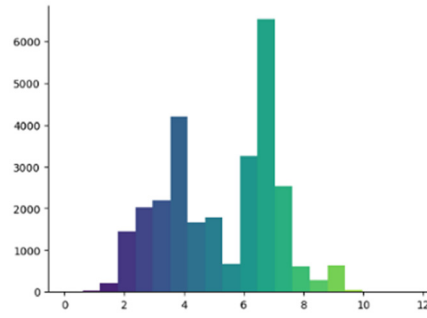


Fig. 4. pH data

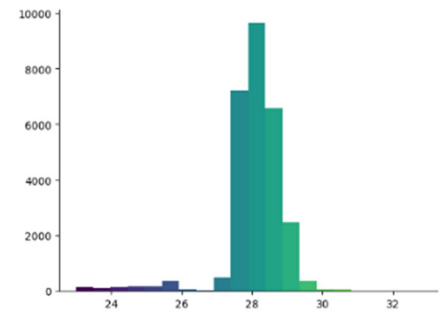


Fig. 5. Temp data

### 3.3 Feature

Feature importance analysis in this study was conducted to identify the variables that have the most significant influence on model performance [30]. The random forest algorithm was selected for this process because, in addition to its strong classification capability, it provides an inherent mechanism for computing feature importance values. This makes it highly suitable for machine learning-based studies that aim to interpret feature contributions within the model. The overall process of feature importance analysis is illustrated in Figure 6. The results indicate that the DO parameter exhibits the highest importance level compared to the pH and temperature parameters. Consequently, DO can be considered the dominant factor in determining water quality conditions within floating net cage systems.

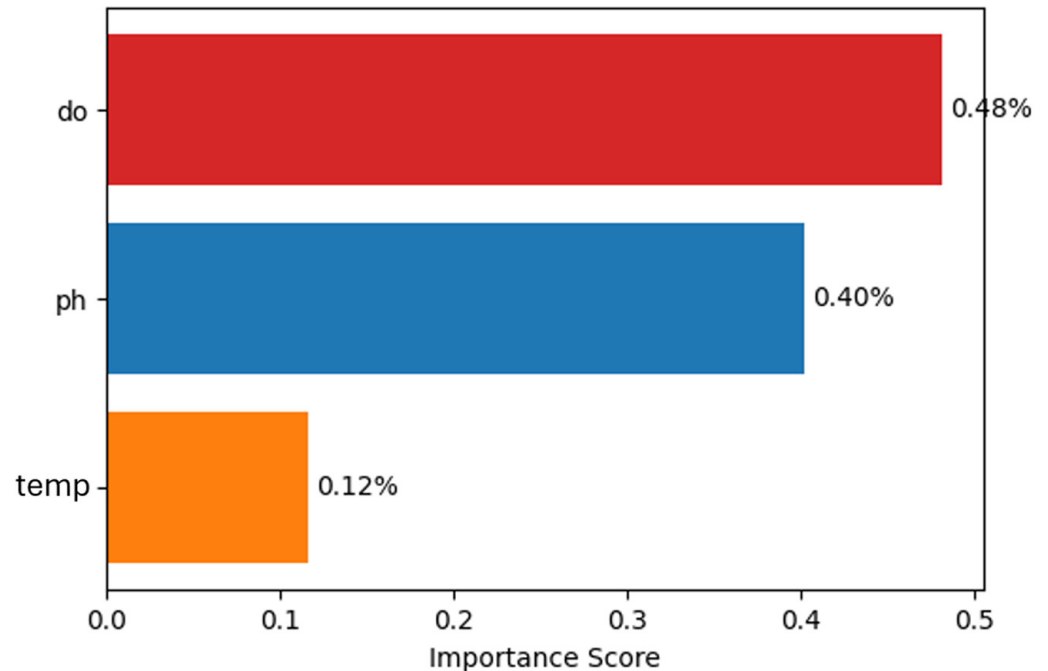


Fig. 6. Feature importance

After completing the feature importance level identification stage, the next step is to examine the data type characteristics of the features used in this study. The three features analyzed, DO, pH, and water temperature, have uniform data types

and consistent data volume. No missing values were found in any of the variables, and the total number of data points collected for each feature was 28109. The results of the statistical description for each feature are presented in Table 2.

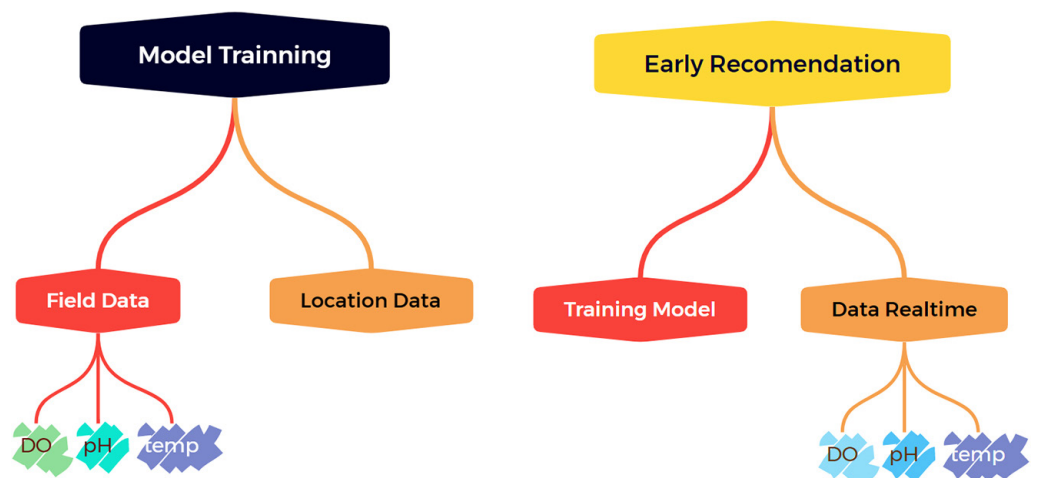
**Table 2.** Feature description

Feature	Data Type	Null	Count
Do	Float64	No Null	28109
Ph	Float64	No Null	28109
Temp	Float64	No Null	28109
Target	Float64	No Null	28109

### 3.4 Classification

Classification is a crucial stage in implementing machine learning methods, as the input features used significantly impact the quality of the results obtained [31], [32]. In the context of this study, classification is used to determine water quality-related to the suitability of floating net cages, with the target output being the condition categories of good or poor. The initial stage of the classification process is featuring extraction, which aims to identify the features that contribute most significantly to the classification results. Next, various machine learning algorithms were applied, including support vector machine (SVM), K-nearest neighbor (KNN), Naïve Bayes, AdaBoost, and random forest, to evaluate the performance of the classification process.

The training process is conducted using a labelled dataset to obtain relevant patterns and relationships between variables. The trained model is then used to classify water conditions on new data. Further implementation aims to integrate this classification model into embedded devices, enabling the system to provide real-time recommendations for the placement of floating net cages. With this approach, the research not only produces computationally tested models but also presents practical solutions that fishermen can directly utilize. Figure 7 presents an illustration of the machine learning process flow and model implementation in this study.



**Fig. 7.** Machine learning model

The classification process begins with the utilization of input data obtained from the sensors used in the research. This data is used not only in the model training process to build a supervised learning-based prediction model but also in conjunction with location data logging, which serves as the basis for recommendations on the placement of floating net cages. The model generated during the training stage is then integrated into the embedded system, enabling it to serve as an early detector of the water quality index. Thus, the system can provide real-time recommendations for the placement of floating net cages based on direct sensor readings from the field.

The primary advantage of this approach over manual monitoring is its ability to quickly and accurately detect changes in water conditions. Manual monitoring is often inconsistent, relies on the experience of fishermen, and risks being late in responding to changes in water quality. Conversely, integrating machine learning models into embedded devices allows data analysis to occur automatically, continuously, and adaptively to environmental conditions. This provides significant practical value, as it can help fishermen reduce the risk of crop failure, improve operational efficiency, and support the sustainability of fish farming using innovative technology. Figure 8 illustrates the contribution of improvement technology using the four stages of location-based recommendation for the water quality index ML model and the aquaculture system. The four main layers are: 1) data acquisition, 2) preprocessing, 3) machine learning model, and 4) alert recommendation.

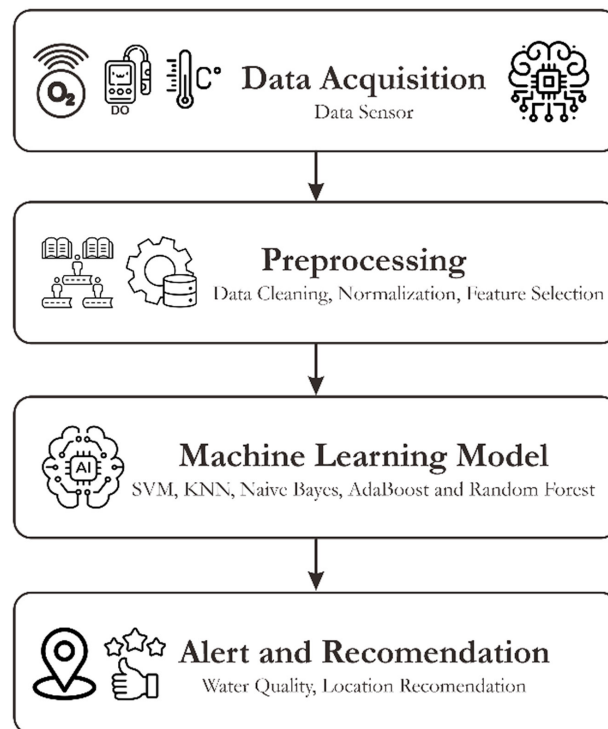


Fig. 8. Architecture of ML-based location recommendation

### 3.5 Proposed model

The proposed model developed in this study is illustrated in Figure 7. The dataset was collected directly from the field using embedded sensing devices, stored in a database, and subsequently extracted into CSV format to facilitate further data processing. The target output is defined as a binary label: good (1) and not good (0),

representing the water quality condition at the location of the floating net cages. Before the training phase, the dataset undergoes a preprocessing stage involving data cleaning and validation. Missing values are handled using mean imputation to maintain data integrity and ensure reliable input for classification analysis.

Following the data preparation, the classification model is constructed using the Python programming language executed on the Google Colab platform, which enables scalable data processing with cloud-based computational support. The processed dataset is then used to train and evaluate several machine learning models, resulting in a final classification model that can generate real-time recommendations for determining the optimal placement of floating net cages based on water quality conditions (refer to Algorithm 1).

The model's performance evaluation in this study was conducted using a confusion matrix, as shown in Table 3 and Figure 9. The confusion matrix is used to measure classification performance by comparing the model's prediction results against the actual labels. In this study, an evaluation was conducted on two classes: a positive class representing good water conditions and a negative class representing poor water conditions. Through a confusion matrix, information was obtained regarding the number of correct and incorrect predictions for each class.

**Algorithm 1: Finding the values in the "target" label from the dataset**

```

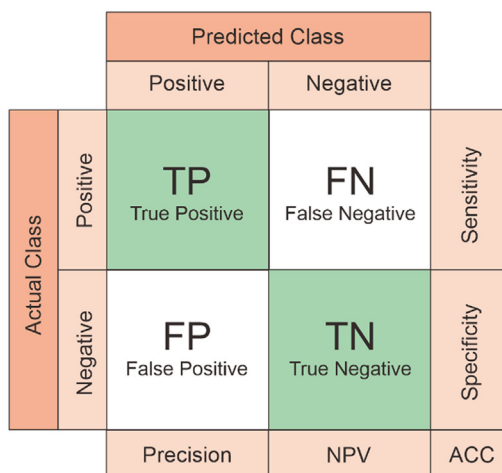
1. import pandas as pd
2. from sklearn.model_selection import train_test_split
3. from sklearn.ensemble import RandomForestClassifier
4. from sklearn.metrics import accuracy_score, classification_report
5. dataset = pd.read_csv('dataset.csv')
6. #Water Condition depends on sensor data
   target_condition = (3.0<=dfx.do) & (dfx.do<=7.0) & (6.0<=dfx.ph) & (dfx.ph<=9.0) & (25.0<=dfx.temp)
   & (dfx.temp<=32)
7. #Water Good Condition target = 1 and Bad target = 0
   dfx.loc[permen_condition, "target"] = 1
   dfx.loc[permen_condition, "target"] = 0
8. Append the Data into the dataset
9. X = dataset.drop('label', axis=1)
10. y = dataset['label']
11. X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
12. classifier = RandomForestClassifier(n_estimators=100, random_state=42)
13. classifier.fit(X_train, y_train)
14. y_pred = classifier.predict(X_test)
15. accuracy = accuracy_score(y_test, y_pred)
16. report = classification_report(y_test, y_pred)
17. print ("Classification Report:")

```

Next, the confusion matrix was used as the foundation for calculating several evaluation metrics, including accuracy, precision, recall, and F1-score, which collectively provide a comprehensive representation of the classification model's performance [33], [34]. Additionally, the receiver operating characteristic (ROC) curve and the area under the curve (AUC) values were utilized to evaluate further the model's discriminative capability in distinguishing between positive and negative classes [35]. The combination of these metrics provides a comprehensive evaluation of model reliability, encompassing both overall accuracy and the balance of sensitivity and specificity in detecting water quality conditions. Through this multi-layered evaluation approach, the model's performance can be assessed more objectively and comprehensively, ensuring the robustness of its predictive capability in real-world scenarios.

**Table 3.** Confusion matrix

Predicted	Actual	
	Good	Not Good
Good	TP	FP
Not Good	FN	TN



**Fig. 9.** Binary confusion matrix

The classification process in this study employs a binary classification approach, emphasizing the calculation of the sensitivity (recall) metric. Sensitivity measures the model’s ability to correctly identify water conditions that are genuinely suitable for the placement of floating net cages. This metric is crucial for ensuring that the system does not fail to recognize optimal water conditions, thereby enhancing the accuracy and practical relevance of the generated recommendations for fishermen. The mathematical formulation for calculating sensitivity is presented in Equation (1).

$$Sensitivity = \frac{(TP)}{(TP + FN)} \tag{1}$$

To obtain the true negative rate (TNR), also referred to as specificity, the formula presented in Equation (2) is applied. This metric measures the model’s capability to correctly identify negative instances, corresponding to water conditions categorized as poor. In this context, specificity ensures that the system does not misclassify unsuitable water conditions as favorable. Accurate calculation of this metric is crucial to minimizing potential errors in recommending floating net cage placement locations, thereby reducing the risk of aquaculture failure and ensuring more reliable decision-making for sustainable fish farming operations.

$$Specificity = \frac{(TN)}{(TN + FP)} \tag{2}$$

The testing plan for validating the error metrics obtained from the training and validation phases of the machine learning model is structured as follows. The evaluation was conducted using three primary performance metrics: mean squared error (MSE), root mean squared error (RMSE), and the coefficient of determination (R<sup>2</sup>). These metrics were selected to quantitatively assess the model’s predictive accuracy and consistency between the predicted and actual values. The mathematical

formulations for MSE, RMSE, and  $R^2$  are presented in Equations (3), (4), and (5), respectively.

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i)^2 \tag{3}$$

$$RMSE = \sqrt{MSE} \tag{4}$$

$$R^2 = 1 - \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2} \tag{5}$$

This methodology is structured into a logically presented sequence of development, as shown in Figure 10.

### 4 RESULT AND ANALYSIS

The model training was conducted on a computer equipped with an Intel Core i7-13620H CPU @ 2.4 GHz, 16 GB RAM, and an NVIDIA GeForce RTX 3050 GPU, operating under the Windows 11 environment. The dataset used for model development was processed using the Python programming language, with the scikit-learn library serving as the primary module. Scikit-learn was chosen due to its high stability, extensive support for standardized machine learning algorithms, and a comprehensive suite of evaluation tools for classification analysis. These advantages make scikit-learn one of the most widely adopted libraries in both academic research and practical machine learning implementations.

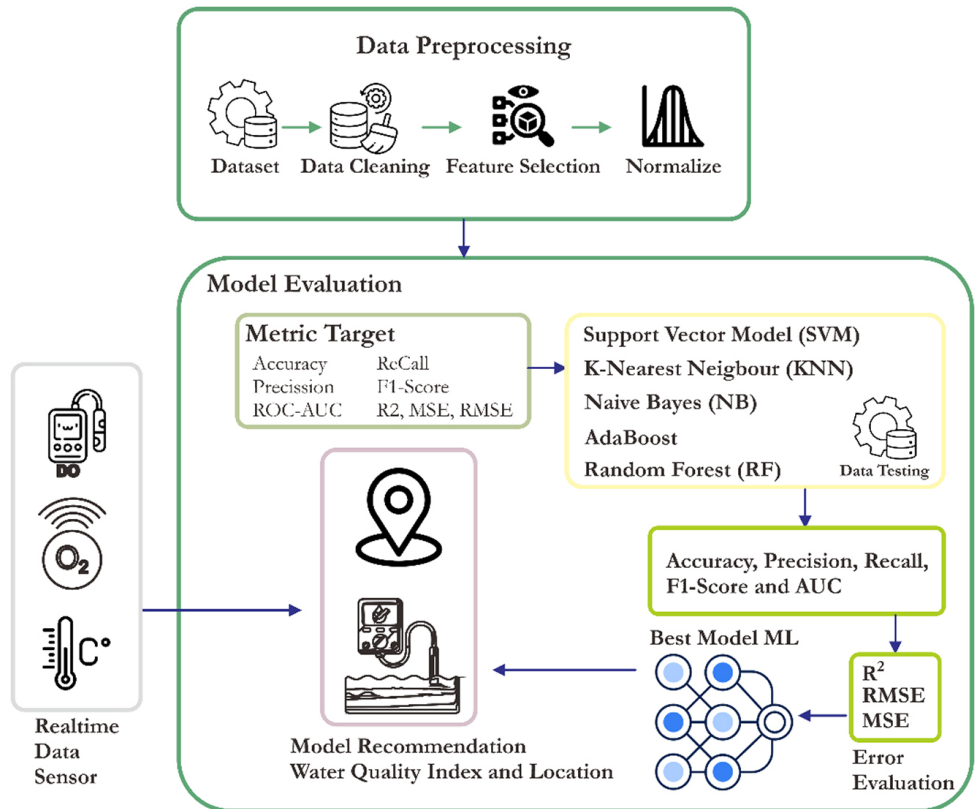


Fig. 10. Methodology diagram

The performance evaluation of the trained model was conducted using a binary confusion matrix, as illustrated in Figure 11. The confusion matrix provides an in-depth representation of the model's classification capability by mapping predicted and actual outcomes across two classes: good and not good water quality conditions. Furthermore, it serves as the foundation for computing other essential evaluation metrics, including accuracy, precision, recall, and F1-score, which collectively describe the model's overall performance and reliability in classifying water quality for floating net cage applications.

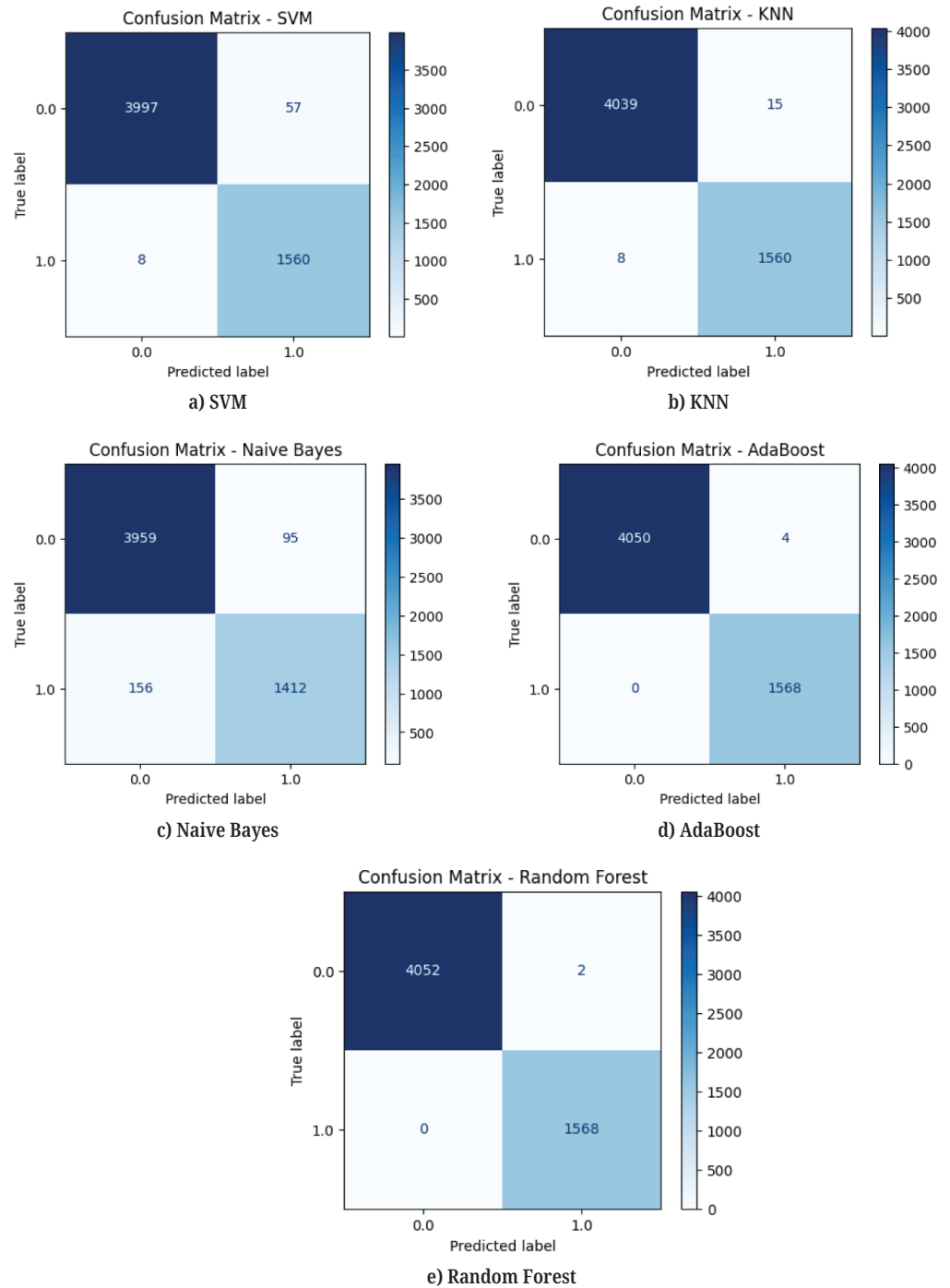


Fig. 11. Binary confusion matrix result

The evaluation results, using a binary confusion matrix, show that the Random Forest algorithm performs very well, with a true positive (TP) value of 4052, false negatives (FN) of 2, false positives (FP) of 0, and true negatives (TN) of 1568. This value indicates that the model can classify almost all data correctly, with only two prediction errors in the positive class and no errors found in the negative class. A detailed overview of the evaluation results is presented in Table 4.

**Table 4.** Comparison of binary confusion matrix

Algorithm	TP	FN	FP	TN
SVM	3997	57	8	1560
KNN	4039	15	8	1560
Naive Bayes	3959	95	156	1412
AdaBoost	4050	4	0	1568
Random Forest	4052	2	0	1568

Further testing was performed to compare the performance of the classification models by calculating several evaluation metrics, namely precision, recall, F1-score, and accuracy. These four metrics were used to provide a comprehensive assessment of the model's performance, capturing not only its overall accuracy but also its ability to classify both positive and negative classes correctly. Such an evaluation ensures that the model's predictive capabilities are well-balanced across all categories, thereby improving its reliability in real-world applications. The detailed results of the model performance evaluation are presented in Table 5.

**Table 5.** Performance comparison of algorithm

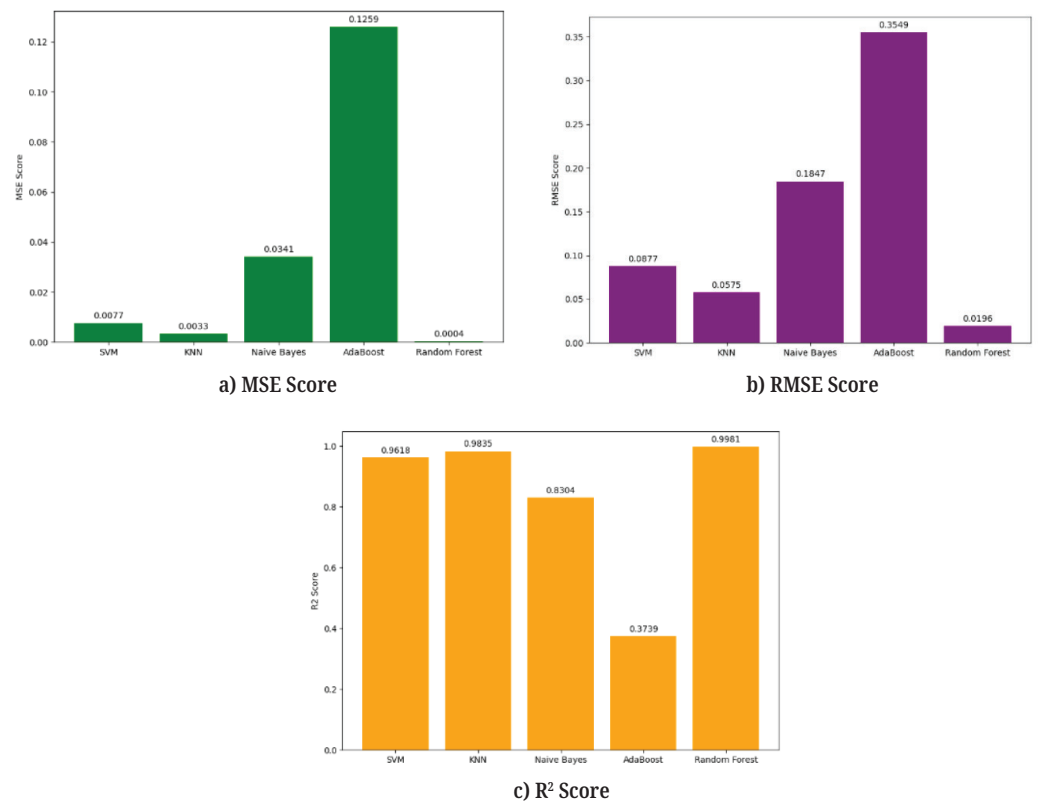
Algorithm	Precision	F1-Score	Accuracy	AUC
SVM	0.99	0.9795	0.9884	0.9997
KNN	1.0	0.9926	0.9959	0.9989
Naive Bayes	0.96	0.9183	0.9553	0.9889
AdaBoost	1.0	0.9987	0.9992	1.0
Random Forest	1.0	0.9993	0.9996	1.0

The subsequent evaluation was conducted by calculating the error values obtained from the training and validation phases of the machine learning model. This analysis utilized three primary error metrics: MSE, RMSE, and the coefficient of determination ( $R^2$ ). The MSE and RMSE metrics were employed to quantify the magnitude of deviation between the model's predicted values and the actual observations. At the same time,  $R^2$  was used to measure how well the model explains the variability of the dataset. Together, these three metrics provide a comprehensive assessment of the model's predictive accuracy and generalization capability. The detailed results of these evaluations are presented in Table 6.

**Table 6.** Performance comparison of MSE, RMSE, and R<sup>2</sup>

Algorithm	MSE	RMSE	R <sup>2</sup>
SVM	0.0077	0.0877	0.9618
KNN	0.0033	0.0575	0.9835
Naive Bayes	0.0341	0.1847	0.8304
AdaBoost	0.1259	0.3549	0.3739
Random Forest	0.0004	0.0196	0.9981

The visualization of the performance comparison results for each machine learning algorithm was carried out using three evaluation metrics: MSE, RMSE, and R<sup>2</sup>. The comparative graph of MSE values is presented in Figure 12(a), the RMSE comparison in Figure 12(b), and the R<sup>2</sup> comparison in Figure 12(c). These visualizations were designed to provide a clearer understanding of the performance variations among the algorithms, thereby facilitating the analysis and identification of the most suitable model for generating recommendations for floating net cage placement. Through this visual comparison, the strengths and weaknesses of each algorithm can be more easily interpreted, supporting data-driven decision-making in model selection.



**Fig. 12.** MSE score, RMSE score, and R<sup>2</sup> score

## 5 CONCLUSION AND FUTURE WORKS

The results of this study on determining water quality for floating net cage placement were obtained using a primary dataset consisting of three input features:

DO, pH, and temperature, and one output label. Based on the feature importance analysis, the DO sensor demonstrated the highest influence on model performance, contributing 48.15%, followed by pH at 40.23% and temperature at 11.62%. This finding suggests that DO is the primary parameter in evaluating water quality for aquaculture environments.

Among the five machine learning algorithms evaluated, the Random Forest model achieved the best performance, with an overall accuracy of 99.96%, followed by AdaBoost (99.92%), KNN (99.59%), SVM (98.84%), and Naïve Bayes (95.53%). The evaluation of model errors using three primary metrics, MSE, RMSE, and  $R^2$ , produced the results summarized in Table 6. The random forest algorithm yielded the best outcomes across all three metrics, with  $MSE = 0.0004$ ,  $RMSE = 0.0196$ , and  $R^2 = 0.9981$ , demonstrating its superior predictive accuracy and robustness for water quality classification and site recommendation in floating net cage applications.

Overall, the results of this study demonstrate that integrating IoT-based data acquisition with machine learning models provides a practical and reliable approach for assessing water quality in floating net cage systems. The high accuracy achieved by the random forest algorithm, supported by low error values and strong coefficient determination, confirms its suitability for real-time classification and decision support in aquaculture site selection. These findings underscore the significant role of DO in determining water suitability, reaffirming its crucial importance in maintaining fish health and preventing mass mortality. Moreover, the developed model offers practical benefits for fishermen and aquaculture managers, enabling them to make informed decisions regarding cage placement based on objective and data-driven insights. By leveraging artificial intelligence and IoT technologies, this study contributes to the development of innovative and sustainable aquaculture practices that can enhance productivity while preserving aquatic ecosystems.




## 6 REFERENCES




- [1] Food and Agriculture Organization of the United Nations, *The State of World Fisheries and Aquaculture – Blue Transformation in Action*, 2024. [Online]. Available: [https://www.fao.org/3/ca9229en/online/ca9229en.html#chapter-1\\_1](https://www.fao.org/3/ca9229en/online/ca9229en.html#chapter-1_1)
- [2] C. Nithin, N. Reddy, S. Gautham, P. A. V, and P. Kumar, “Water quality monitoring system based on IoT,” *International Journal of Novel Research and Development*, vol. 8, no. 4, pp. 69–73, 2023.
- [3] W. Eric, E. Wa, U. Eric, N. Pb-, and G. Eric, “Water quality & standards,” *Water*, pp. 76–86, 2017.
- [4] W. Khan, A. Vahab, A. Masood, and N. Hasan, “Water quality requirements and management strategies for fish farming a case study of ponds around gurgaon canal NUH Palwal,” *Int. J. Trend Sci. Res. Dev.*, vol. 2, no. 1, pp. 388–393, 2017. <https://doi.org/10.31142/ijtsrd5914>
- [5] K. S. Negara, “Appendix VI concerning national water quality standards – PP Number 22 of 2021 concerning the implementation of environmental protection and management,” vol. 1, no. 078487A, p. 483, 2021. [Online]. Available: <http://www.jdih.setjen.kemendagri.go.id/>
- [6] D. Hermanto, D. Stiawan, B. Y. Suprpto, E. Permata, and E. P. Widiyanto, “New approach monitoring system with ESP32 and MQTT for the best position of the floating net cage,” in *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, 2023, pp. 139–144. <https://doi.org/10.1109/EECSI59885.2023.10295802>
- [7] J. S. Fidari and M. Bisri, “Intensive pond design model planning for probolinggo regency”.




- [8] B. Di, W. Plta, K. Panjang, and K. Riau, "Sustainable floating net cage development strategy in koto panjang hydroelectric reservoir kampar riau," *Jurnal Perikanan dan Kelautan (JPK)*, vol. 2, pp. 145–160, 2010.
- [9] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Comput. Networks*, vol. 54, no. 15, pp. 2787–2805, 2025. <https://doi.org/10.1016/j.comnet.2010.05.010>
- [10] P. Malhotra *et al.*, "Internet of things: Evolution, concerns and security challenges," *Sensors*, vol. 21, no. 5, p. 1809, 2021. <https://doi.org/10.3390/s21051809>
- [11] G. Gao, K. Xiao, and M. Chen, "An intelligent IoT-based control and traceability system to forecast and maintain water quality in freshwater fish farms," *Comput. Electron. Agric.*, vol. 166, 2019. <https://doi.org/10.1016/j.compag.2019.105013>
- [12] Muti'ah, M. S, A. Majid, and C. Seftarita, "What determines aquaculture fish production? empirical evidence from South Aceh Regency, Indonesia," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 674, 2021. <https://doi.org/10.1088/1755-1315/674/1/012040>
- [13] W. Kusumawijaya and A. Aula, "Water quality monitoring and control system in fish cultivation pond using arduino cloud," *Comput. Electr. Eng. J.*, vol. 1, no. 3, pp. 333–340, 2024. <https://doi.org/10.26418/telectrical.v1i3.73567>
- [14] J. S. Fidari and M. Bisri, "Intensive pond pond design model planning," *Tek. Pengair.*, vol. 8, pp. 252–261, 2017. <https://doi.org/10.21776/ub.pengairan.2017.008.02.11>
- [15] F. Simanjuntak and I. C. B. H dan Muhamad, "Floating net cage management strategy," in *Semin. Nas. Edusainstek FMIPA UNIMUS 2018*, 2018, pp. 38–46.
- [16] K. Abirami, P. C. Radhakrishna, and M. A. Venkatesan, "Water quality analysis and prediction using machine learning," in *Proc. – 2023 12th IEEE Int. Conf. Commun. Syst. Netw. Technol. CSNT 2023*, vol. 10, 2023, pp. 241–245. <https://doi.org/10.1109/CSNT57126.2023.10134661>
- [17] B. Shrimali, S. Surati, A. Patel, and R. Kansagara, "Predictive modeling of water quality in Indian rivers: A machine learning approach for sustainable resource management," in *Proceedings of the 1st International Conference on Cognitive & Cloud Computing (IC3Com 2024)*, vol. 1, 2025, pp. 165–174. <https://doi.org/10.5220/0013303800004646>
- [18] G. A. Rudrani *et al.*, "Machine learning approaches for predicting water quality towards climate-resilient groundwater management in southern India," *Hydrol. Res.*, vol. 56, no. 8, pp. 754–773, 2025. <https://doi.org/10.2166/nh.2025.042>
- [19] A. Sharma, R. Sharma, R. Rana, and A. Kalia, "Water quality prediction using machine learning models," *E3S Web Conf.*, vol. 596, pp. 1–12, 2024. <https://doi.org/10.1051/e3sconf/202459601025>
- [20] M. Vijay Anand, C. Sohitha, G. N. Saraswathi, and G. V. Lavanya, "Water quality prediction using CNN," *J. Phys. Conf. Ser.*, vol. 2484, no. 1, 2023. <https://doi.org/10.1088/1742-6596/2484/1/012051>
- [21] X. Wang, Y. Li, Q. Qiao, A. Tavares, and Y. Liang, "Water quality prediction based on machine learning and comprehensive weighting methods," *Entropy*, vol. 25, no. 8, pp. 1–20, 2023. <https://doi.org/10.3390/e25081186>
- [22] S. Sen, S. Maiti, S. Manna, B. Roy, A. Gosh, "Smart prediction of water quality system for aquaculture using machine learning algorithms," *J. Curr. Trends Comput. Sci. Res.*, vol. 2, no. 3, pp. 212–225, 2023. <https://doi.org/10.33140/JCTCSR.02.03.01>
- [23] J. Julian, A. B. Dewantara, and F. Wahyuni, "Design of machine learning-based water quality prediction system with recursive feature elimination cross-validation," *J. Infotel*, vol. 15, no. 3, pp. 249–255, 2023. <https://doi.org/10.20895/infotel.v15i3.977>
- [24] C. L. Li Tingting, L. Jian, W. Jun, and Z. Zhenhua, "Predicting aquaculture water quality using machine learning approaches," *Educ. Psychol.*, vol. 40, no. 5, pp. 533–534, 2020. <https://doi.org/10.1080/01443410.2020.1755503>

- [25] M. Manoj, R. Rajan, W. Quality, and R. Pi, "Precision aquaculture using Iot & machine learning," *AGPE R. Gondwana Res. J. Hist. Sci. Econ. Polit. Soc. Sci.*, vol. 2, no. 1, pp. 105–111, 2021.
- [26] D. Hermanto, D. Stiawan, B. Y. Suprpto, E. Permata, and E. P. Widiyanto, "New approach monitoring system with ESP32 and MQTT for the best position of the floating net cage," in *2023 10th Int. Conf. Electr. Eng. Comput. Sci. Informatics*, pp. 139–144, 2023. <https://doi.org/10.1109/EECSI59885.2023.10295802>
- [27] K. Phani Rama Krishna *et al.*, "Dissolved oxygen measurement in water," *J. Theor. Appl. Inf. Technol.*, vol. 103, no. 4, pp. 1426–1431, 2025.
- [28] R. Pramana, B. Y. Suprpto, and Z. Nawawi, "Application of fuzzy logic in decision-making process for relocation of floating net cages in river fish farming," *Glob. J. Environ. Sci. Manag.*, vol. 10, no. 1, pp. 117–132, 2024. <https://doi.org/10.22034/gjesm.2024.01.09>
- [29] P. G. Arepalli, K. Jairam Naik, and J. K. Rout, "Aquaculture water quality classification with sparse attention transformers: Leveraging water and environmental parameters," in *ACM Int. Conf. Proceeding Ser.*, 2024, pp. 318–325. <https://doi.org/10.1145/3651781.3651829>
- [30] H. Liu, Y. Pei, Q. Bei, and L. Deng, "Improved DeepSORT algorithm based on multi-feature fusion," *Appl. Syst. Innov.*, vol. 5, no. 3, p. 55, 2022. <https://doi.org/10.3390/asi5030055>
- [31] M. Ragab, "Internet of things assisted solid biofuel classification using sailfish optimizer hybrid deep learning model for smart cities," *Sustain.*, vol. 15, no. 16, 2023. <https://doi.org/10.3390/su151612523>
- [32] B. Sliwa, N. Piatkowski, and C. Wietfeld, "LIMITS: Lightweight machine learning for IoT systems with resource limitations," in *IEEE International Conference on Communications*, 2020, pp. 1–7. <https://doi.org/10.1109/ICC40277.2020.9149180>
- [33] K. P. Rasheed Abdul Haq and V. P. Harigovindan, "Water quality prediction for smart aquaculture using hybrid deep learning models," *IEEE Access*, vol. 10, pp. 60078–60098, 2022. <https://doi.org/10.1109/ACCESS.2022.3180482>
- [34] I. Cvitić, D. Peraković, M. Periša, and B. Gupta, "Ensemble machine learning approach for classification of IoT devices in smart home," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 11, pp. 3179–3202, 2021. <https://doi.org/10.1007/s13042-020-01241-0>
- [35] Z. R. Saputra Elsi *et al.*, "Feature selection using chi square to improve attack detection classification in IoT network: Work in progress," in *Int. Conf. Electr. Eng. Comput. Sci. Informatics*, vol. 2022, 2022, pp. 226–232. <https://doi.org/10.23919/EECSI56542.2022.9946621>



## 7 AUTHORS




**Dedy Hermanto**    received a master's degree in information technology from the University of Indonesia, Depok, West Java, Indonesia. Currently, he is a PhD candidate at the Faculty of Engineering, Sriwijaya University. He now works as a Senior Lecturer at the Faculty of Computer Science and Engineering, Multi Data Palembang University, Palembang, Indonesia. My research interests include the Internet of Things, computer networks, geographic information systems, and robotics (E-mail: [dedy@mdp.ac.id](mailto:dedy@mdp.ac.id)).

**Deris Stiawan**    received the PhD degree in Computer Engineering from Universiti Teknologi Malaysia, Malaysia. He is currently serving as a Professor at the Department of Computer Engineering, Faculty of Computer Science, Universitas Sriwijaya. His research interests include computer networks, Intrusion Detection/Prevention Systems, and heterogeneous networks (E-mail: [deris@unsri.ac.id](mailto:deris@unsri.ac.id)).

**Bhakti Yudho Suprpto**    is an Associate Professor in the Electrical department at the Faculty of Engineering, Universitas Sriwijaya, Indonesia. He obtained

a Doctor's Degree in Electrical Engineering from Universitas Indonesia. His professional profile has led to a specialization in Robotics and Control, with a focus on fuzzy logic and neural networks. His research interests include control systems (E-mail: [bhakti@ft.unsri.ac.id](mailto:bhakti@ft.unsri.ac.id)).

**Mohd. Yazid Idris**   is an Associate Professor at the Faculty of Computing, Universiti Teknologi Malaysia. He obtained his M.Sc. and Ph.D. in the area of Software Engineering and Information Technology (IT) Security in 1998 and 2008, respectively. In software engineering, he focuses on researching the design and development of mobile and telecommunication software. His main research activity in IT security is in the area of Intrusion Prevention and Detection (IPD) (E-mail: [yazid@utm.my](mailto:yazid@utm.my)).

**Rahmat Budiarto**    received B.Sc. degree from Bandung Institute of Technology in 1986, and M.Eng. and Dr.Eng. degrees in Computer Science from Nagoya Institute of Technology in 1995 and 1998, respectively. Currently, he is a full Professor at the College of Computer Science and IT, Albaha University, Saudi Arabia. His research interests include intelligent systems, brain modeling, IPv6, network security, wireless sensor networks, and MANETs (E-mail: [rahmat@bu.edu.sa](mailto:rahmat@bu.edu.sa)).