Scientific Journal of Informatics
Vol. 13, No. 2, May 2026
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji e-ISSN 2460-0040

Stacking Ensemble with Hybrid Balancing for Aquaculture Water Quality Prediction

Ari Nugroho Putro[1]*, Much Aziz Muslim[2]
[1,2] Department of Computer Science, Universitas Negeri Semarang, Indonesia

Abstract.
Purpose: This study aims to improve the accuracy of water quality classification models by addressing class imbalance and data noise. Aquaculture water quality monitoring is essential to support sustainable aquaculture production and maintain aquatic organism health. As aquaculture systems become more intensive, accurate predictive models are needed for effective monitoring and decision-making. Therefore, this study proposes a stacking ensemble model optimized using SMOTEENN hybrid balancing to improve classification performance on aquaculture water quality datasets.
Methods: The proposed approach combines SMOTEENN hybrid balancing with a meta stacking ensemble framework. First, Synthetic Minority Oversampling Technique (SMOTE) was applied to balance minority classes by generating synthetic samples. Next, Edited Nearest Neighbor (ENN) was used to remove noisy data from both original and synthetic datasets. After preprocessing, the classification process employed a meta stacking ensemble model consisting of Extra Trees (ET) and Random Forest (RF) as base learners, while Logistic Regression (LR) served as the meta learner. The model was evaluated using the Aquaculture Water Quality (AWQ) dataset.
Result: Experimental results show that the proposed model achieves the highest performance under the SMOTEENN scenario, reaching an accuracy of 99.88%, outperforming SMOTE 99.20%, ENN 99.78%, and the original imbalanced data 99.18%. The results indicate that combining class balancing and noise reduction significantly improves classification performance.
Novelty: This study presents a novel integration of SMOTEENN hybrid balancing and meta stacking ensemble learning, offering an effective solution for handling imbalanced and noisy environmental datasets in water quality classification.

Keywords: Stacking Ensemble, Hybrid Balancing, Data Noise, Classification, Accuracy

Received May 2026 / **Revised** May 2026 / **Accepted** May 2026

INTRODUCTION

Water quality is one of the crucial aspects that directly affects human health [1] and various social and economic activities such as industry and agriculture [2]. In the context of aquaculture, water quality plays a vital role in maintaining aquatic organism health, supporting sustainable fish farming practices, and ensuring ecosystem stability. Water quality degradation due to chemical, biological, and physical pollution is becoming an increasingly complex global challenge as urbanization, industrialization, climate change, and the intensification of aquaculture systems increase [3]. To overcome these problems, a fast, accurate, and reliable water quality detection and evaluation system is needed [4].

In recent years, the use of machine learning techniques in water quality classification has experienced rapid development [5]. These predictive models can assist in the decision-making process and accelerate the assessment of water conditions based on available parameters [6]. However, the success of the model is greatly influenced by the quality of the dataset used [7]. In aquaculture water quality datasets, common problems such as class imbalance and the presence of data noise can cause a decrease in accuracy and prevent the model from recognizing important patterns in the data [8], [9].

Several studies have attempted various approaches to improve the accuracy of classification models, including performance enhancement techniques such as ensemble learning [2], [10] and resampling methods [11], [12], [13]. Although class balancing and ensemble learning techniques have been widely applied [1],

[1]*Corresponding author.
Email addresses: arinugrohoputro@students.unnes.ac.id (Putro), a212muslim@mail.unnes.ac.id (Muslim)
DOI: 10.15294/sji.v13i1.49383

[2], their combined use, particularly integrating hybrid resampling techniques such as SMOTEENN within a stacking ensemble framework for water quality classification, remains insufficiently explored.

Based on this gap, this study proposes a meta stacking ensemble model optimized with SMOTEENN hybrid balancing. The proposed framework combines two base learners, Extra Trees (ET) and Random Forest (RF), whose outputs are integrated using Logistic Regression (LR) as a meta learner [10], [14]. In addition, SMOTEENN is employed to address class imbalance by generating synthetic samples and removing noisy data simultaneously [15], [16].

METHODS

A. Data Description

This study uses the AWQ dataset, taken from the Mendeley Data repository https://data.mendeley.com/datasets/y78ty2g293/1, containing 4,300 fish pond water samples with 14 features: Temperature, Turbidity, Dissolved Oxygen, Biochemical Oxygen Demand (BOD), CO2, pH, Alkalinity, Hardness, Calcium, Ammonia, Nitrite, Phosphorus, H2S, and Plankton. Water quality is categorized into three classes: Very Good (0), Good (1), and Poor (2), with distributions of 1,400, 1,400, and 1,500 samples, respectively. Details of the data description can be found in the Table 1.

Table 1. Data description

Features	Median	Skewness
Temp	25.04	2.53
Turbidity	30.21	0.70
DO	5.00	1.11
BOD	2.24	2.21
CO2	6.60	-0.16
pH	7.74	-0.38
Alkalinity	67.56	1.18
Hardness	111.06	0.91
Calcium	62.85	1.34
Ammonia	0.03	5.60
Nitrite	0.10	1.91
Phosphorus	0.98	0.58
H2S	0.02	3.23
Plankton	3729.40	0.22

B. Data Preprocessing

The preprocessing stage was carried out to produce a cleaner and more representative dataset. Skewness in the dataset was reduced using Power Transformer with the Yeo–Johnson method, which empirically reduced the skewness value of the features and produced a data distribution that was closer to symmetrical than the initial condition. The Yeo–Johnson transformation is defined in (1):

$$y_i = \begin{cases} \frac{(x_i+1)^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ \ln(x_i+1), & \lambda = 0 \end{cases} \quad (1)$$

Where x_i represents the original value and y_i represents the transformed value. After this stage, the data is ready for resampling and noise handling.

C. SMOTE

Handle class imbalance by adding synthetic records from the minority class. If x_i is a minority class record and x_{zi} is one of its closest neighbors, then the synthetic record is generated according to (2):

$$x_{\text{new}} = x_i + \lambda(x_{zi} - x_i), \lambda \in (0,1) \quad (2)$$

This step allows for the evaluation of the effect of class balancing on model performance without removing noise.

D. ENN

Delete records whose labels differ from the majority of their nearest neighbors (k-NN). Deleted records are considered noise. Formally, a sample is removed if the condition in (3) is satisfied:

$$x_i \text{ is removed if } y_i \neq \text{mode}\{\, y_j \mid x_j \in \mathcal{N}_k(x_i) \,\} \tag{3}$$

This stage evaluates the effect of noise cleaning without adding synthetic records. In this study, ENN was applied with $k = 3$, which is the default value to balance noise removal sensitivity and class distribution.

E. SMOTEENN

SMOTEENN is a hybrid resampling approach that combines SMOTE and ENN to generate a balanced and noise-reduced dataset. The procedure is expressed in (4):

$$D' = \text{SMOTE}(D), D'' = \text{ENN}(D') \tag{4}$$

SMOTE was first used to balance class distribution, then ENN was applied to remove records that were potentially noise in the oversampling dataset. This scenario provides a comprehensive analysis of the effects of class balancing and noise removal simultaneously. Details of the data distribution for all resampling schemes are shown in Table 2.

Table 2. Data distribution

Resampling	Class 0	Class 1	Class 2	Total Sampel
Raw Data	1400	1400	1500	4300
SMOTE	1500	1500	1500	4500
ENN	1400	642	471	2513
SMOTEENN	1031	762	471	2264

F. Classification Model

To evaluate the proposed approach, this study employs two classification schemes, namely a single classifier scheme and a stacking ensemble scheme. In the single classifier scheme, Random Forest (RF) and Extra Trees (ET) are independently trained using datasets generated from each resampling strategy (i.e., without resampling, SMOTE, ENN, and SMOTEENN). This scheme aims to analyze the direct impact of resampling strategies on the performance of individual models. In the stacking ensemble scheme, RF and ET are used as base learners. Each base learner generates a class probability vector for each sample $x_i$as defined in (5):

$$\mathrm{p}_m(x_i) = [p_{m1}(x_i), p_{m2}(x_i), \ldots, p_{mC}(x_i)] \tag{5}$$

where Cdenotes the number of classes and $m$denotes the base learner index. The outputs of all base learners are then concatenated to form the second-level feature representation as expressed in (6):

$$\mathrm{z}_i = [\mathrm{p}_1(x_i), \mathrm{p}_2(x_i), \ldots, \mathrm{p}_M(x_i)] \tag{6}$$

The probability vectors used to construct the second-level features are obtained using out-of-fold predictions through 10-fold cross-validation on the training data to avoid information leakage. At the meta level, Logistic Regression (LR) is employed as the meta-learner to generate the final prediction as shown in (7):

$$\hat{y}_i = \arg\max_c \frac{\exp(\mathrm{w}_c^{\top}\mathrm{z}_i + b_c)}{\sum_{k=1}^{C} \exp(\mathrm{w}_k^{\top}\mathrm{z}_i + b_k)} \tag{7}$$

where w_c and b_c are the weight vector and bias term associated with class c, respectively. The complete pipeline of the proposed method is illustrated in Figure 1, covering load dataset, preprocessing, resampling, classification model, and model evaluation.

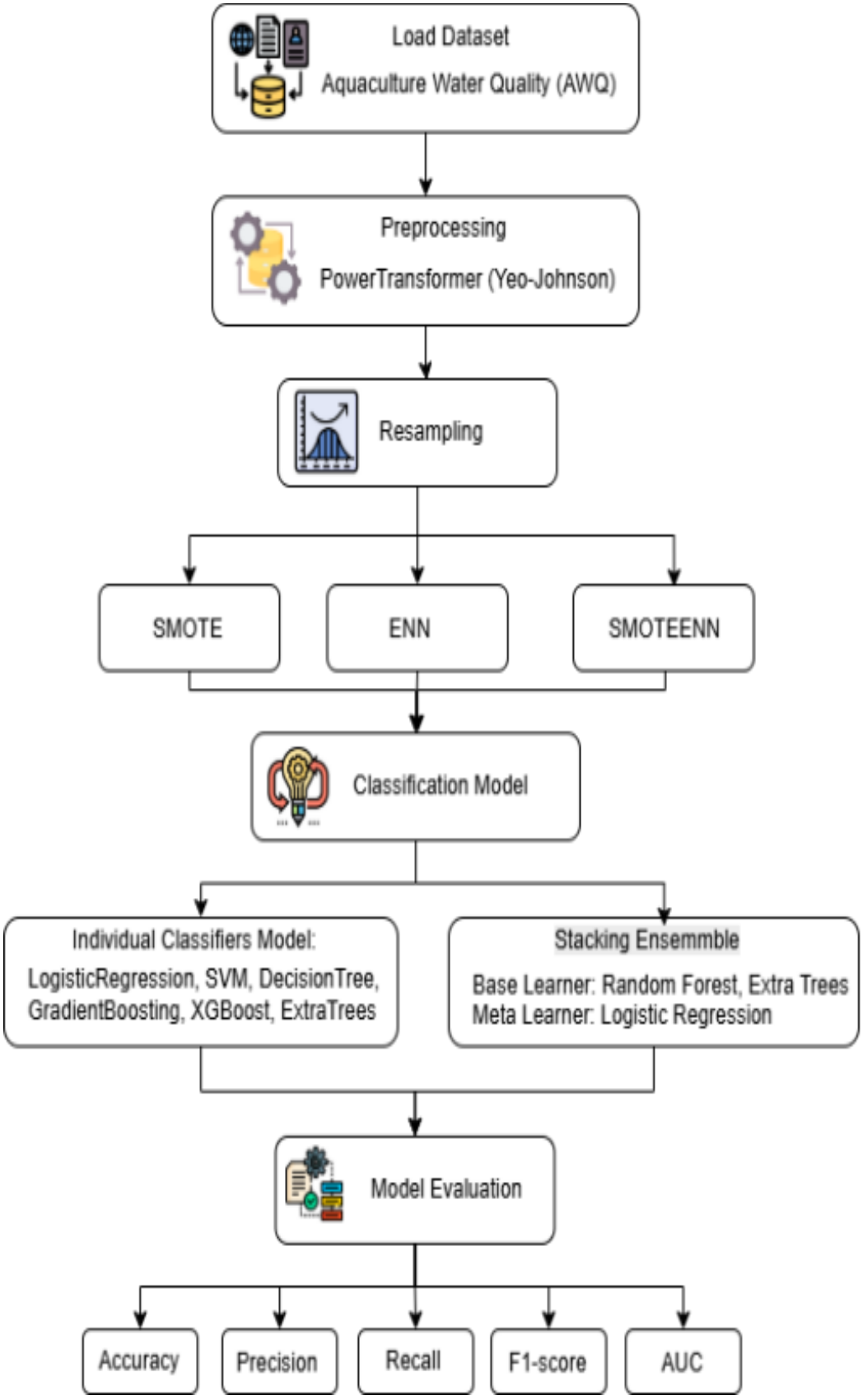

Figure 1. Proposed method

G. Model Evaluation

To ensure model performance consistency, all experiments were evaluated using stratified 10-fold cross-validation. Classification performance was analyzed based on a confusion matrix that included true positive (TP), true negative (TN), false positive (FP), and false negative (FN). [17], [18]. The confusion matrix presents the relationship between model predictions and actual data conditions, thereby forming the basis for calculating various evaluation metrics. [19], [20]. Based on the confusion matrix, several evaluation metrics are used, namely accuracy, precision, recall, F1-score, and AUC.

Accuracy is defined in (8) which shows the proportion of correct predictions against all records [21]. Recall is defined in (9) which measures the model's ability to correctly detect all positive records [22]. Precision is defined in (10) which indicates the level of reliability of the positive predictions generated by the model [22]. The F1-score is formulated in (11) which represents a balance between precision and recall [23]. AUC (Area Under the Curve) is calculated using (12) which measures the model's ability to distinguish between positive and negative records [23].

$$\frac{TP+TN}{TP+TN+FP+FN} \tag{8}$$

$$\frac{TP}{TP+FN} \tag{9}$$

$$\frac{TP}{TP+FP} \tag{10}$$

$$\frac{2 \cdot Precision \cdot Recall}{Precision+Recall} \tag{11}$$

$$\frac{1}{N_+ N_-} \sum_{i:y_i=1} \sum_{j:y_j=0} I\left(\hat{p_i} > \hat{p_j}\right) \tag{12}$$

RESULT AND DISCUSSION

This study evaluates the effect of resampling strategies on the performance of water quality classification in datasets with different distribution characteristics, as well as their impact when combined with stacking ensemble. The accuracy performance of each model can be seen in Table 3.

Table 3. The accuracy performance of each model

Dataset	Model	Raw Data	SMOTE	ENN	SMOTEENN
AWQ	LogisticRegression	82.13%	83.13%	92.12%	92.41%
AWQ	SVM	94.51%	94.68%	99.69%	99.71%
AWQ	DecisionTree	99.53%	99.53%	99.85%	99.85%
AWQ	GradientBoosting	99.27%	99.20%	99.72%	99.67%
AWQ	RandomForest	99.62%	99.64%	99.81%	99.76%
AWQ	XGBoost	99.23%	99.20%	99.84%	99.76%
AWQ	ExtraTrees	95.37%	95.48%	99.63%	99.62%
AWQ	**Stacking (RF, ET, LR)**	**99.18%**	**99.20%**	**99.78%**	**99.88%**

At baseline, the stacking ensemble achieved an accuracy of 99.18%. The highest increase in accuracy was obtained when the model was optimized using the hybrid balancing method SMOTEENN, with accuracy increasing to 99.88%. This improvement was higher than using SMOTE or ENN individually. SMOTE increased accuracy to 99.20%, while ENN produced 99.78%. These results show that the hybrid approach provides more consistent performance improvement compared to single resampling techniques. The overall performance of SMOTEENN in each model can be seen in Table 4.

Table 4. Model performance in the SMOTEENN scenario

Model	Accuracy	Precision	Recall	F1	AUC
LogisticRegression	92.41%	92.29%	92.41%	92.09%	96.02%
ExtraTrees	99.62%	99.62%	99.62%	99.61%	99.99%
GradientBoosting	99.67%	99.68%	99.67%	99.67%	99.99%
SVM	99.70%	99.71%	99.70%	99.70%	99.97%
XGBoost	99.76%	99.76%	99.76%	99.76%	100%
RandomForest	99.76%	99.76%	99.76%	99.76%	99.99%
DecisionTree	99.85%	99.85%	99.85%	99.85%	99.85%
Stacking (RF, ET, LR)	**99.88%**	**99.88%**	**99.88%**	**99.88%**	**99.99%**

SMOTE and ENN operate in fundamentally different ways to improve model performance. SMOTE increases minority class representation by generating synthetic samples, resulting in an addition of 200 synthetic samples, increasing the dataset size from 4,300 to 4,500. In contrast, ENN removes noisy samples based on nearest neighbor's disagreement, eliminating 1,787 samples from the dataset. When combined, SMOTEENN first balances the class distribution and then applies noise filtering, resulting in a final dataset of 2,264 samples. Compared to previous studies, the SMOTEENN–Stacking ensemble shows consistent performance improvements. Comparison of accuracy with similar studies can be seen in Table 5.

Table 5. Comparison of accuracy with similar studies

Author	Model	Dataset	Accuracy
Hridoy et al. (2025) [24]	LightGBM and XGBoost	AWQ	99.6%
Arabelli et al. (2025) [25]	DNN with Adam optimizer	AWQ	97.41%
Naim et al. (2025) [26]	HBA with XGBoost classifier	AWQ	98.05%
Proposed Model	**SMOTEENN-Stacking Ensemble**	**AWQ**	**99.88%**

In the AWQ dataset, the best previously reported performance was achieved by Hridoy et al. (2025) [24] using a LightGBM and XGBoost-based approach, with an accuracy of 99.60%. This is slightly higher than other prior methods such as Naim et al. (2025) [26], which reported 98.05%, and Arabelli et al. (2025) [25], which achieved 97.41% using a deep neural network. In this study, the proposed SMOTEENN–Stacking ensemble achieves an accuracy of 99.88%, outperforming all compared methods and confirming that the integration of SMOTE-based oversampling, ENN-based noise reduction, and stacking ensemble learning significantly improves water quality classification performance compared to previous approaches.

CONCLUSION

This study evaluates the impact of SMOTE, ENN, and SMOTEENN resampling strategies on water quality classification using multiple machine learning models and a stacking ensemble on the AWQ dataset. The results show that resampling significantly improves performance compared to imbalanced data, with the hybrid SMOTEENN approach providing the most consistent gains across all models by combining oversampling and noise reduction. The stacking ensemble achieves the best overall performance, reaching 99.88% accuracy under SMOTEENN, outperforming all individual classifiers and baseline methods. These findings confirm that the integration of SMOTEENN with ensemble learning is an effective and robust approach for improving water quality classification performance.

REFERENCES

[1] A. Juna *et al.*, "Water Quality Prediction Using KNN Imputer and Multilayer Perceptron," *Water (Switzerland)*, vol. 14, no. 17, Sep. 2022, doi: 10.3390/w14172592.

[2] C. T. Doan and H. Du Nguyen, "Robust water quality prediction across multiple indicator formulations using an explainable ensemble learning model," *Water Resour. Ind.*, vol. 34, Dec. 2025, doi: 10.1016/j.wri.2025.100329.

[3] N. A. Misman, M. F. Sharif, A. J. K. Chowdhury, and N. H. Azizan, "Water pollution and the assessment of water quality parameters: a review," *Desalination Water Treat.*, vol. 294, pp. 79–88, May 2023, doi: 10.5004/dwt.2023.29433.

[4] S. Palabıyık and T. Akkan, "Evaluation of water quality based on artificial intelligence: performance of multilayer perceptron neural networks and multiple linear regression versus water quality indexes," *Environ. Dev. Sustain.*, 2024, doi: 10.1007/s10668-024-05075-6.

[5] M. A. Helaly *et al.*, "Advancements in water quality prediction: a practical review of machine learning and deep learning approaches," Oct. 01, 2025, *Springer*. doi: 10.1007/s10586-025-05221-3.
[6] P. Singh *et al.*, "An ensemble-driven machine learning framework for enhanced water quality classification," *Discover Sustainability*, vol. 6, no. 1, Dec. 2025, doi: 10.1007/s43621-025-01467-4.
[7] M. He, Q. Qian, X. Liu, J. Zhang, and J. Curry, "Recent Progress on Surface Water Quality Models Utilizing Machine Learning Techniques," Dec. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/w16243616.
[8] R. Xu, S. Hu, H. Wan, Y. Xie, Y. Cai, and J. Wen, "A unified deep learning framework for water quality prediction based on time-frequency feature extraction and data feature enhancement," *J. Environ. Manage.*, vol. 351, Feb. 2024, doi: 10.1016/j.jenvman.2023.119894.
[9] N. Nasaruddin, N. Masseran, W. M. R. Idris, and A. Z. Ul-Saufie, "A SMOTE PCA HDBSCAN approach for enhancing water quality classification in imbalanced datasets," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-97248-0.
[10] M. H. Ganjgah, P. Azhdari, and K. F. Vajargah, "A hybrid stacking ensemble approach combining fuzzy logistic regression and fuzzy neural networks for enhanced COVID-19 diagnostic accuracy," *Discover Applied Sciences*, vol. 7, no. 10, Oct. 2025, doi: 10.1007/s42452-025-07456-6.
[11] K. Hwang, W. Kang, and Y. Jung, "Application of the class-balancing strategies with bootstrapping for fitting logistic regression models for post-fire tree mortality in South Korea," *Environ. Ecol. Stat.*, vol. 30, no. 3, pp. 575–598, Sep. 2023, doi: 10.1007/s10651-023-00573-8.
[12] H. Ahmad, B. Kasasbeh, B. Aldabaybah, and E. Rawashdeh, "Class balancing framework for credit card fraud detection based on clustering and similarity-based selection (SBS)," *International Journal of Information Technology (Singapore)*, vol. 15, no. 1, pp. 325–333, Jan. 2023, doi: 10.1007/s41870-022-00987-w.
[13] S. Matharaarachchi, M. Domaratzki, and S. Muthukumarana, "Enhancing SMOTE for imbalanced data with abnormal minority instances," *Machine Learning with Applications*, vol. 18, p. 100597, Dec. 2024, doi: 10.1016/j.mlwa.2024.100597.
[14] E. Öz, O. Bulut, Z. F. Cellat, and H. Yürekli, "Stacking: An ensemble learning approach to predict student performance in PISA 2022," *Educ. Inf. Technol. (Dordr).*, vol. 30, no. 6, pp. 7753–7779, Apr. 2025, doi: 10.1007/s10639-024-13110-2.
[15] B. Zhu, X. Jing, L. Qiu, and R. Li, "An Imbalanced Data Classification Method Based on Hybrid Resampling and Fine Cost Sensitive Support Vector Machine," *Computers, Materials and Continua*, vol. 79, no. 3, pp. 3977–3999, 2024, doi: 10.32604/cmc.2024.048062.
[16] P. M. Vieira and F. Rodrigues, "An automated approach for binary classification on imbalanced data," *Knowl. Inf. Syst.*, vol. 66, no. 5, pp. 2747–2767, May 2024, doi: 10.1007/s10115-023-02046-7.
[17] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," *J. Big Data*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00657-5.
[18] D. K. Dake, E. Nwiah, G. S. Klogo, and W. X. Ativi, "Instructor-assisted question classification system using machine learning algorithms with N-gram and weighting schemes," *Discover Artificial Intelligence*, vol. 3, no. 1, Dec. 2023, doi: 10.1007/s44163-023-00073-5.
[19] A. Kord, A. Aboelfetouh, and S. M. Shohieb, "Academic course planning recommendation and students' performance prediction multi-modal based on educational data mining techniques," *J. Comput. High. Educ.*, 2025, doi: 10.1007/s12528-024-09426-0.
[20] X. Yang *et al.*, "Coal burst spatio-temporal prediction method based on bidirectional long short-term memory network," *Int. J. Coal Sci. Technol.*, vol. 12, no. 1, Dec. 2025, doi: 10.1007/s40789-025-00759-4.
[21] D. Ananda Agustina Pertiwi and M. Aziz Muslim, "Improvement accuracy of gradient boosting in app rating prediction on google playstore," *Journal of Numerical Optimization and Technology Management*, vol. 1, no. 2, p. 2023, doi: 10.00000/jnotm.0000.00.00.000.
[22] M. Jebbari, B. Cherradi, S. Hamida, and A. Raihani, "Identifying learning styles in MOOCs environment through machine learning predictive modeling," *Educ. Inf. Technol. (Dordr).*, vol. 29, no. 16, pp. 20977–21014, Nov. 2024, doi: 10.1007/s10639-024-12637-8.
[23] M. Ismail, S. Alrabaee, K. K. R. Choo, L. Ali, and S. Harous, "A Comprehensive Evaluation of Machine Learning Algorithms for Web Application Attack Detection with Knowledge Graph Integration," *Mobile Networks and Applications*, vol. 29, no. 3, pp. 1008–1037, Jun. 2024, doi: 10.1007/s11036-024-02367-z.

[24] M. A. A. M. Hridoy *et al.*, "Advanced machine learning models for accurate water quality classification and WQI prediction: Implications for aquatic disease risk management," *Science of the Total Environment*, vol. 1008, Dec. 2025, doi: 10.1016/j.scitotenv.2025.180965.
[25] R. Arabelli, T. Bernatin, and V. Veeramsetty, "Water quality assessment for aquaculture using deep neural network," *Desalination Water Treat.*, vol. 321, Jan. 2025, doi: 10.1016/j.dwt.2025.101016.
[26] S. M. Naim, P. Das, J. J. Tiang, and A. Al Nahid, "Aquaculture Water Quality Classification Using XGBoost Classifier Model Optimized by the Honey Badger Algorithm with SHAP and DiCE-Based Explanations," *Water (Switzerland)*, vol. 17, no. 20, Oct. 2025, doi: 10.3390/w17202993.