ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 14 (2024) No. 1, pp. 301-321
DOI:10.33168/JSMS.2024.0118

Optimizing Predictive Modeling for Water Quality in Farms with Blended Artificial Neural Network

Sumitra Nuanmeesri [1], Lap Poomhiran [2], Preedawon Kadmateekarun [1],
Chaisri Tarasawatpipat [1]

[1] Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok, Thailand
[2] Faculty of Information Technology and Digital Innovation, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand

sumitra.nu@ssru.ac.th; lap.p@email.kmutnb.ac.th; preedawan.ka@ssru.ac.th; chaisri.th@ssru.ac.th

Abstract. Coastal water quality directly impacts coastal industries like aquaculture and salt production. This research aimed to optimize predictive models for coastal water quality using blended machine learning techniques. Specifically, Multi-Layer Perceptron Neural Networks were combined with Support Vector Machine classifiers. This blended model outperformed Softmax activation and One-vs-All SVM classifiers. Additionally, Multiple Generative Adversarial Imputation Networks were used to impute missing data. This made the dataset more complete and improved model performance. The blended model combining Neural Networks, One-vs-One SVMs, and imputation had the highest accuracy for water quality classification. Therefore, blended artificial neural networks show promise for optimizing solutions for water quality modeling in coastal environments.

Keywords: data imputation, multi-layer perceptron neural networks, multiple generative adversarial imputation network, support vector machine, water quality

1. Introduction

Water is the world's most natural resource, accounting for 71% of the Earth's surface area (Abuzir & Abuzir, 2022). Of all this water, 97.5% is seawater (Mishra, 2023). Water quality is essential for many farming activities, such as crop, fish, and salt farms. This is because water is the main element for plants or vegetation to grow and is a habitat for aquatic animals. In addition, seawater is also the primary raw material for salt production. Thus, water quality monitoring and management to ensure water quality suitable for such farms is inevitable. Evaluating water quality contributes to care, control, and reduce impact by ensuring the water is safe (Islam et al., 2021). In Samut Songkhram Province, Thailand, there are many areas where canals and the river connect to the sea. Different farms within the same area may have different water quality requirements. For example, the salt farm requires a salinity of around 27 to 30 parts per thousand (ppt) (Kim et al., 2020), while the fish farm, the Asian seabass (Lates Calcarifer) requires a salinity of 20 ppt for growth (Hassan et al., 2024). In contrast, the crop farm does not require salinity in the water. However, since water sources are from the same source or transport route, it is difficult to determine whether the water quality harmed any farms in the same area at that time. Therefore, developing models to classify water quality as suitable for various farms is necessary and challenging, as the same water quality may or may not be suitable for various farms.

Most of the water quality data are in the form of quantitative data, which may include the water temperature (WT), the potential of hydrogen (pH), Dissolved Oxygen (DO), electrical conductivity (EC), biochemical oxygen demand (BOD), chemical oxygen demand (COD), turbidity, ammonia (NH3), lead (Pb), copper (Cu), cadmium (Cd), mercury (Hg), and arsenic (As). Some of these parameters have a significant direct effect on water quality. For example, The rise in sea temperature by only 2°Celsius resulted in a decrease in the amount of water or shortages (Paltán et al., 2021), but in turn, increased the concentration of fish waste in the cages, which is harmful to the fish (Li et al., 2023). In addition, if the amount of DO decreases, it will affect the growth of the fish because it requires more energy to exchange oxygen (Jasmin et al., 2022). Imbalanced salinity and mineral affect the fish's feed conversation rate (Harpaz et al., 2005). Moreover, the salinity causes some horticultural crops to be dry and produce less fruit (de Luna Souto et al., 2023). Most studies collect water quality data from one or multiple water sources but a binary class or a few classes between three and five classes. However, there is very little research on using water sources in the same area and then classifying them for different contexts, such as various farms. These water quality classes can be susceptible to different farming practices. Currently, real-time water quality data is collected by Internet of Things (IoT) devices, including probes or sensors that transmit analog data, process it, and then automatically send it to a database server via internet and network communication. However, sometimes there may be cases where the probe or sensor does not work correctly, resulting in multiple missing values in the records. More than one attribute may be missing within the same record. Several data imputation techniques exist, such as mean, regression, RF, Expectation-Maximization (EM), time series, and multiple imputations that can be solved these issues. Especially the Multiple Generative Adversarial Imputation Network (MGAIN), which is based on Generative Adversarial Imputation Networks (GAIN) (Yoon et al., 2018), was designed to solve multiple missing data problems (Zhao et al., 2022). These techniques apply the generator and discriminator with a weighted sum of the losses. Thus, data preprocessing and data imputation need to be applied to help improve model accuracy.

Nowadays, machine learning and deep neural networks play an essential role in predicting or classifying outcomes based on learning from historical data. The descriptive data can be trained on a neural network model, such as Multi-Layer Perceptron Neural Networks (MLPNN), which includes several hidden deep layers. Some studies have found that the MLPNN model is more effective than some machine learning models such as Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Support Vector Machine (SVM), and Naïve Bayes (NB) (Nuanmeesri et al., 2022). Much research is related to water quality classification by applying neural networks or machine learning for water quality index prediction (Chawla et al., 2021; Kaur et al., 2023; Nasir et al., 2022). Ladjal et al. (2023), who developed a water quality classification model by fusion between MLPNN and Principal Component Analysis (PCA) and SVM and PCA, tested that the MLPNN model had higher efficiency than SVM for four water quality parameters. However, Zhu et al. (2022) have shown that SVM provides better feedback accuracy than neural networks, possibly due to instability. MLPNN might be suitable for datasets with few features or dimensionality. On the other hand, if the data set has more features or is high dimensional, using SVM as a classifier will produce better results. When constructing a classification model, it is challenging to decide between MLPNN and SVM. In addition, some research applied the Softmax activation function in the output layer of MLPNN. In addition, the MLPNN allows assigning machine learning classifiers such as multi-class SVM (MCSVM) instead of the Softmax activation function. However, it is hard to say that there are not a few research that combines MLPNN and MCSVM or other machine learning for water quality classification.

According to mentioned above, the problem is that the same water source is used in many farming types where different levels of water quality must be considered. This results in many classes, which is a great challenge to classify the appropriate water quality to notify farmers and find solutions. Conventional machine learning, such as SVM (Hong et al., 2022), can help classify water quality, which requires sufficient features to make decisions under the assumption that a multiclass dataset that extracts the features using MLPNN and then classifies with MCSVM can help optimize the classify water quality for different farms. Therefore, this work aims to optimize the MLPNN-based model blended with MCSVM and then train the model with the dataset that passed the MGAIN data imputation technique to increase the classification accuracy and efficiency for various farms in the same sub-district area connected to the coastal sea.

2. Literature Review

Several studies have been undertaken to develop models for water quality classification using machine learning (Emamgholizadeh et al., 2013; Haghiabi et al., 2018). Most research uses machine learning methods to classify water quality: RF, NB, Artificial Neural Network (ANN), and SVM. These models were used to compare their efficiency with the different classifiers (Bui et al., 2020; Hassan et al., 2021; Ho et al., 2019; Nasir et al., 2022; Radhakrishnan & Pillai, 2020; Shan et al., 2018) . Nevertheless, further studies are still being conducted on using these machine learnings. RF uses a method to combine the results from each decision tree and then vote for the final output (Sakaa et al., 2022). Several studies used RF to classify water quality (Gakii & Jepkoech, 2019; Lu & Ma, 2020; Vijay & Kamaraj, 2019). Swetha et al. (2023) developed the water quality prediction model based on RF for aquaculture. The NB classification technique relies on the separateness of characteristic conditions, and while defining the target value characteristics, the Bayes theorem is taken into account as a conditional independent (Aldhyani et al., 2020). The NB has been widely used for water quality prediction (Malek et al., 2022). The NB and RF perform better than the C5.0 algorithm (Vijay & Kamaraj, 2019). The NB model with the four water quality parameters, pH, BOD, DO, and EC, has an accuracy of 95.17% higher than the SVM model. However, the correctly classified by the SVM model is higher than the NB model (Radhakrishnan & Pillai, 2020). Many works have successfully used ANN to classify water quality (Ahmad et al., 2016; Gaafar et al., 2020; Al-Adhaileh & Alsaade, 2021). Because due to the learning rules, it continuously adjusts the weight until the desired output is reached (Manoj & Ananda, 2022). Some studies use ANN as time series to predict water quality (Aldhyani et al., 2020; Ömer Faruk, 2010; Wang et al., 2013). The MLPNN is the ANN widely used for classification (Ahmed et al., 2019). It is also capable of regression (Günther & Fritsch, 2010). Rajaee et al. (Rajaee et al., 2020) applied the ANN model for river water quality classification. They agree that the ANN can provide realistic results for water quality classification. At the same time, SVM is also used as a classifier based on the decision hyperplane and maximum margin distance in binary classification (Zhang et al., 2020). The water quality in the Chao Phraya River was classified by attribute realization and SVM. The result shows that the model has an accuracy of up to 95% for water quality classification (Sillberg et al., 2021).

Muhammad et al. (2015) confirmed that SVM is suitable for water quality classification. In comparing the model performance between SVM and RF to predict nitrate in groundwater, the RF was better than the SVM model (Rahmati et al., 2019). However, the SVM models are the best in some studies comparing water quality classification models (Radhakrishnan & Pillai, 2020). For example, the SVM and DT are the best models compared to ANN, NB, and K-Nearest Neighbor (KNN) (Sengorur et al., 2015). The SVM model has an accuracy that performs better than the MLPNN model (Giri & Singh, 2014; Haghiabi et al., 2018; Tan et al., 2012) and the KNN or NB algorithm (Aldhyani et al., 2020). Meanwhile, the MLPNN model is more accurate than the NB, KNN, DT, RF, and SVM models. It had an accuracy of 85.07% when the MLPNN model was set to the configuration of (3,7) for four water quality parameters (Ahmed et al., 2019). Ladjal et al. (2023), who developed a water quality classification model by fusion between PCA-MLPNN and then compared it with PCA-SVM, tested that the PCA-MLPNN model had higher efficiency than PCA-SVM. This study compared the efficiency of the MLPNN model with between one and three hidden layers. It was found that three hidden layers gave better results. It consisted of 4-8-12 neurons for each hidden layer with eight input data attributes, with a test accuracy of 86.61% when fusion with PCA of feature selection yielded a test accuracy of 99.13%. However, there are only three classes for the dataset of four attributes (EC, WT, pH, and turbidity).

Most research has defined a small number of classes in water quality classification, for example, three classes consisting of good, poor, and unsuitable (Hassan et al., 2021). In the case of an increase in the number of classes, this may decrease the accuracy rate of the model. Specifically, SVM applications must be considered, and support multi-class classification must be selected appropriately to make the model more reliable. MCSVM generally has two approaches: the One-against-One (OvO) and One-against-All (OvA) or One-against-Rest (OvR) (Silva & Villela, 2020). In the OvO approach, a separate binary SVM is trained for each pair of classes in the dataset. During the prediction phase, each SVM gives a vote for the class it belongs to, and the class with the most votes is selected as the final prediction. OvO is generally more computationally expensive than OvA because it requires training multiple SVMs. In the OvA approach, a binary SVM is trained for each class, considering it as the positive class and all the other classes as the negative class. During prediction, the test sample is evaluated against all the SVMs, and the class with the highest decision value or probability is assigned as the predicted class. OvA is computationally more efficient than OvO since it requires training only N binary SVM, where N is the number of classes. These two methods, OvO and OvA, are the most commonly used strategies to extend SVMs for multi-class classification problems. The choice between them depends on factors such as the size of the dataset, computational resources, and the specific requirements of the problem at hand. The SVM OvA was used to classify in many studies, such as rainfall (Belghit et al., 2023) and object detection (Yumang et al., 2020). However, it is rare to find studies comparing SVM OvO and SVM OvA in water quality classification.

According to the literature gathered above, it can be seen that a model between ANN and conventional machine learning is possible to classify water quality. However, MLPNN provides better model performance in cases where the dataset has a small volume of features. In contrast, SVMs produce better results as feature volume increases. Further, when compared between SVM OvO and SVM OvA, both have a similar multiclass label classification ability depending on the class and dataset volumes implemented in the model. In particular, the combination between MLPNN and SVM transfers the feature extraction results from MLPNN to SVM for further water quality classification. Sometimes, collecting water quality values may be incomplete during the recording process; for example, some attributes are lost. Thus, this research aims to solve the problem of missing water quality data for the modeling process for predicting water quality in coastal agricultural and fishery areas. Usually, the water quality will be randomly checked by experts and IoT sensor devices for automatic water quality data collection. Nevertheless, it was found that some data was missing by the system or technical glitches, including duplicate records appearing in the database. Problematic data should be corrected before processing. Then, blended ANN with MCSVM and data imputation are expected to be the techniques that improve model efficiency for classifying salinity-based water quality and other parameters for better water quality predicting models in various farms.

3. Research Methods

This work proposed a method for optimizing predictive modeling for water quality using MLPNN blended with the MCSVM and MGAIN data imputation techniques can be illustrated in the research framework as shown in Figure 1. The details of the research method are as follows.

Fig. 1: The proposed research framework

3.1. Water Quality Data Collection

This research's collection water quality area is in three sub-districts: Don Manora, Phraek Nam Daeng, and Bang Kaeo. These three sub-districts are located in Samut Songkhram Province, where a long canal connects to the Mae Klong River that meets the sea in the Gulf of Thailand. The waters in the three sampling areas might change over time, qualifying them as freshwater, brackish water, or saltwater, depending on their distance from the coastal area and the opening and closing time of the sluice gates. The three sub-districts have various farms, consisting of crop, salt, and fish farms which include shrimp, shellfish, and other aquatic animals for food. Two methods for collecting water quality data are expert water sampling and IoT sensor devices. Expert water sampling suits all water areas, even those without facilities. However, there is a limitation in that it is a manual action, so the water quality values cannot always be collected. In contrast, IoT sensor devices were installed in the canal for each sub-district. It has probes or sensors to collect the water quality data in real-time with schedule and collect data automatically. The second method has disadvantages because it requires a budget and fixed equipment installation area. This device cannot be moved easily to collect data in other areas, unlike using experts who can collect data in any area. Moreover, there is a high risk that the devices can be stolen. One set of IoT sensor devices will be installed in each canal per sub-district, totaling three areas for collecting water data automatically. The IoT sensor devices will be installed at the station near the bank of the canal, except for the sensors or probes located under the pontoon, which are submerged in the water and can be adjusted automatically according to the tide conditions. This station uses electricity from solar panels. The water data collection area is a canal that distributes water to different farms in each sub-district so that it can be used to represent the data for each farm type in each sub-district. The IoT system is configured to measure water quality and send data to a database server every hour from February 1st to April 10th, 2023. Further, once a week, the experts will collect water quality data for each location to compare and verify the accuracy of IoT sensor devices. These collected water quality parameters include WT (°Celsius), pH, DO (mg/l), EC (µS/cm), salinity (ppt), turbidity (ppt), and a date-time stamp. The collected water quality data totaled 4,997 records. The process of collecting the water quality data using IoT sensor devices and experts is shown in Figure 2.

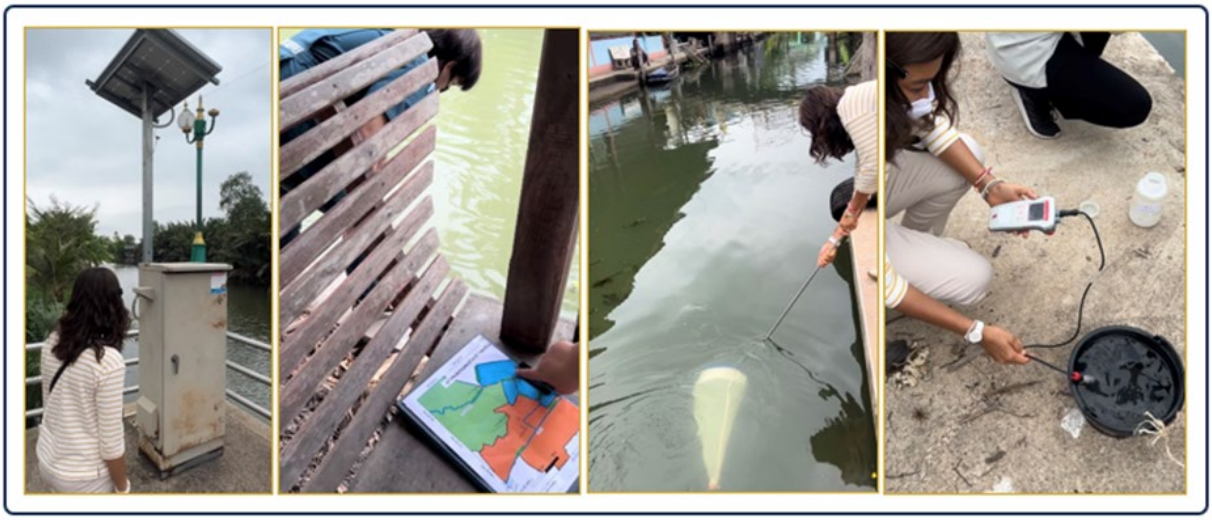


Fig. 2: Water quality data collection in the sample area by IoT sensor devices and experts

3.2. Data Cleansing

Some of the collected data contained redundancy and incompleteness of water quality values. This may be caused by the reason that IoT devices send duplicate data or some sensors do not work correctly or sends the data lacking to the controller, resulting in missing some water quality values. Thus, these data must be cleaned, imputation, and labeled the classes. In this process, the authors directly check the redundancy and validity of the data. Any duplicates will be removed based on the date-time stamp on the list as a judging criterion. The records will also be removed for incomplete entries over the threshold value (ratio of missing values of 50%) or three missing attributes. If the record with missing values is not over the threshold value, it will be processed by data imputation in the next step. After completing this process, the cleaned data has 4,765 records.

3.3. Data Imputation

The pre-cleaned dataset is then filled with the missing parameters as closely or optimally as possible according to the data imputation approach. The data 101 records that need to be fixed include missing values between one and three attributes, which amounted to 67 records, 26 records, and 8 records, respectively. The most common missing value is EC. The following missing value is turbidity. There are a few records where salinity and DO are missing, as shown in Table 1.

Table 1: The missing attributes in the dataset

Number of missing attributes	**Total (records)**	**The missing attribute (records)**				**Remark**
		EC	Turbidity	DO	Salinity	
1	67	38	20	6	3	-
2	26	20	17	10	5	EC or Turbidity+DO (or Salinity), EC+Turbidity
3	8	8	8	6	2	EC+Turbidity+DO (or Salinity)

The MGAIN (Zhao et al., 2022) was applied for the multiple missing value calculation and fill-in to the records with the missing value not over three parameters. The MGAIN technique is based on the GAIN structure (Yoon et al., 2018) for reducing the dimension of attributes in the dataset, as shown in Figure 3.

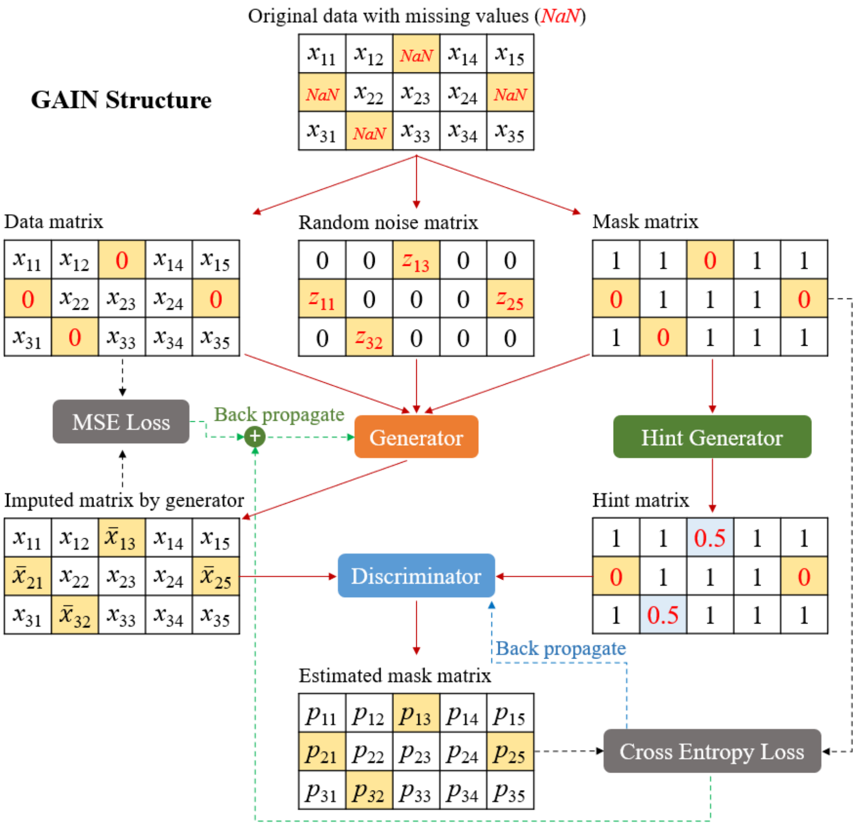


Fig. 3: The GAIN structure

The pre-cleaned dataset D has m records (D_1, D_2, D_3, …, D_m) with n features or attribute (F_1, F_2, F_3, …, F_n) in this work. Let x is an attribute; thus, the dataset D has the attribute x_{11}, x_{12}, x_{13}, …, x_{mn}. It then constructs a subset of the probability of the attributes grouping two or more attributes without picking up duplicates from dataset D, for example, $\{F_1, F_2\}$, $\{F_1, F_3\}$, $\{F_2, F_3\}$, …, $\{F_1, F_2, F_3\}$, …, $\{F_1, F_2, F_3, …, F_n\}$. Resulting in the total number of multiple attribute subsets being j subsets for a record that calculates as in (1). calculates as in (1). calculates as in (1). calculates as in (1). calculates as in (1).

$$j = 2^n - n - 1 \tag{1}$$

These subsets were trained and tested while the accuracy (a) was measured for each subset (a1, a2, a3, …, aj). Subsets with high accuracy were selected, and subsets with low accuracy were ignored or deleted. Thus, the selected subsets will have k subsets where k<j. These selected subsets were re-trained through the GAIN process to calculate their accuracy and then calculate the y predicted missing value, which should be the sum of the average weight of the imputed missing value, to complete the dataset as shown in (2) (Zhao et al., 2022). The MGAIN structure design for predicting the multiple missing values is shown in Figure 4.

$$y = \sum_{i=1}^{k} \left(y_i \frac{a_i}{\sum_{l=1}^{k} a_l} \right) \tag{2}$$

where y_i refers to the imputed missing value for each subset, and $y \in \{y_{11}, y_{12}, y_{12}, …, y_{mn}\}$.

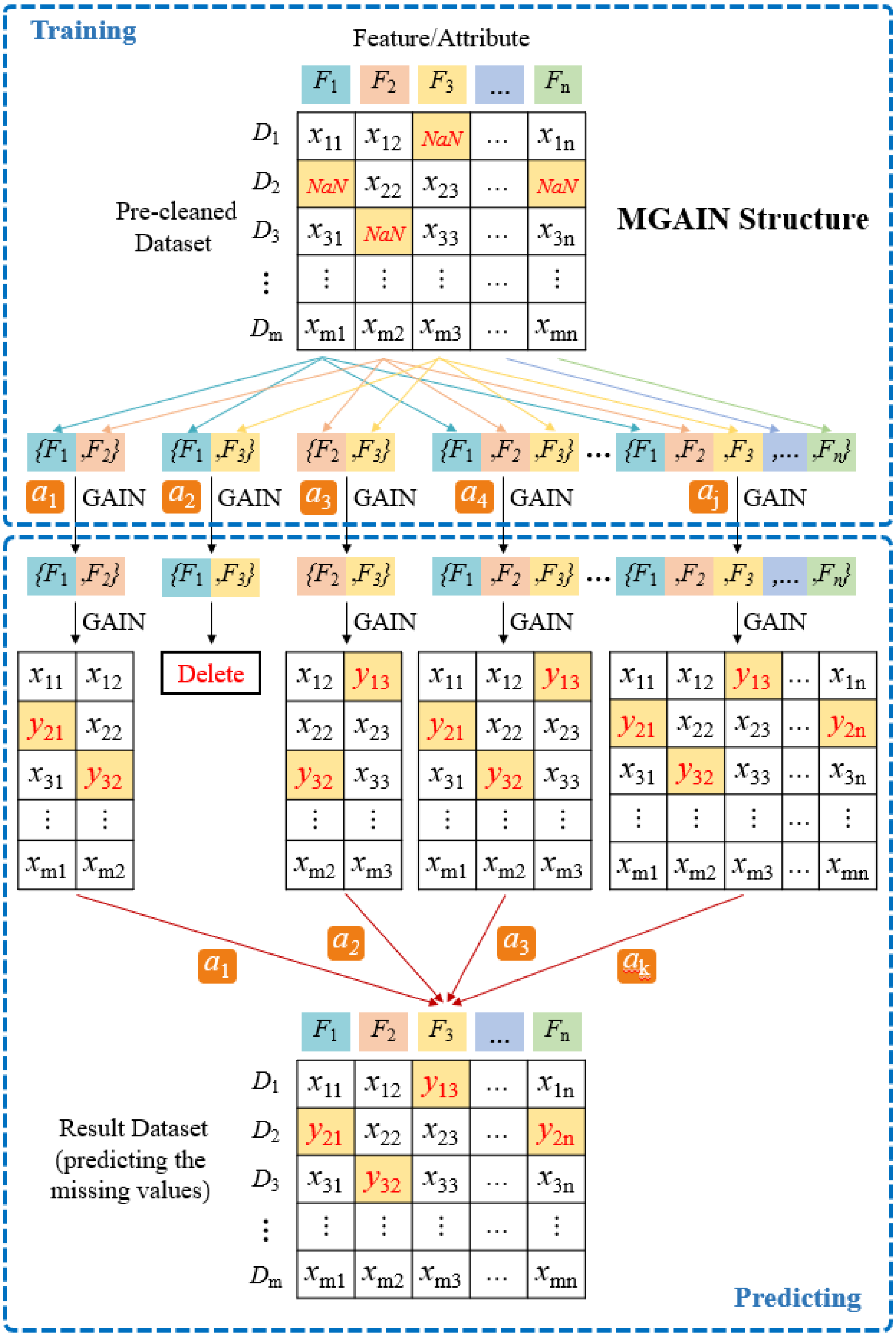


Fig. 4: The MGAIN structure for the multiple missing values prediction

The dataset processed by the MGAIN technique to predict missing data were enriched according to the data imputation approach, making the dataset contain all attributes (no more missing value remain). However, the data filled in may slightly differ from the actual values. Nevertheless, it will not affect the model performance because it will be labeled for the following process to obtain the most accurate dataset.

3.4. Class Labeling

The dataset filled with missing values is then assigned a class value suitable for the water quality used in various farms. Thus, this dataset was evaluated by experts in cultivation, fisheries, and salt farming to define the class label based on the water quality for crop farming, fish farming, and salt farming. The water quality severity level (WQSL) defined in this study was divided into three levels: normal, caution, and critical. When combined with the three farms, there are nine labeled classes (C1 to S3), as shown in Table 2.

Table 2: The class labeling C1 to S3 for various farms in the dataset

Farm type	**Water quality severity level**		
	1: Normal	**2: Caution**	**3: Critical**
Crop (C)	C1	C2	C3
Fish (F)	F1	F2	F3
Salt (S)	S1	S2	S3

The WQSL at level 1 (normal) means water quality is suitable for farming. The WQSL at level 2 (caution) means the water quality is beginning to affect farming, but it is not severely affected. Last, the WQSL at Level 3 (critical) means the water quality is highly unsuitable for the particular farming operation, which can lead to severe damage. In addition, the labeled dataset was divided into three groups: training set, validation set, and test set, in proportions of 70%, 15%, and 15%, respectively, using random split and stratified shuffle techniques for balanced class proportion.

3.5. MLPNN-Based Modeling

Generally, the MLPNN-based model has three layers: input, hidden, and output. There are six input shapes for the input layer in this work. There are three hidden layers in this neural network. The number of neurons for each hidden layer was calculated, as in (3), (4), and (5).

$$Neurons\ of\ Hidden\ Layer1 = f + c \tag{3}$$

$$Neurons\ of\ Hidden\ Layer2 = \left\lceil \frac{2}{3} f \right\rceil + c \tag{4}$$

$$Neurons\ of\ Hidden\ Layer3 = c \tag{5}$$

where f refers to the total number of features ($f = 6$, exclude an attribute 'date-time stamp'), and c refers to the total number of classes ($c = 9$).

Therefore, the number of neurons were 15, 13, and 9 for the first, second, and third hidden layer, respectively. Each neuron has computed the output (y) as the sum of the weight (w) and input (x) with bias (b), as in (6) (Kasasbeh et al., 2022).

$$y = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \tag{6}$$

where n refers to the number of input neurons, w_i refers to the weight of input i-th, x_i refers to input i-th, b refers to a bias of neuron, and f refers to the activation function.

The Sigmoid function is used as a generic function for neural network models. However, the values for the Sigmoid function will range from 0 to 1 with less saturation. Training will become slower as a result. The Rectified Linear Unit (ReLU) activation function was used in this experiment to tune the MLPNN-based model efficiently. Three hidden layers were set with the ReLU activation function for the hidden layer architecture. Each neuron's output was processed using the ReLU activation function, as in (7) (Oostwal et al., 2021), and represented in Figure 5.

$$f(x) = \max\{0, x\} = \begin{cases} o, & x \le 0 \\ x, & x > 0 \end{cases} \tag{7}$$

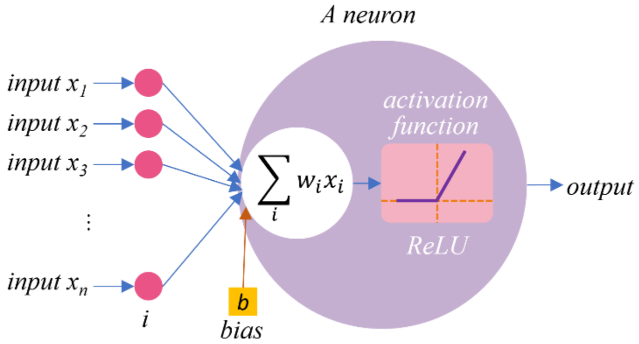


Fig. 5: A neuron input and output with ReLU activation function diagram

In addition, the batch normalization technique is used to improve the training process and generalization performance. Batch normalization addresses this issue by normalizing the inputs to a layer for each mini-batch during training. Normalization involves subtracting the batch mean and dividing by the batch standard deviation. By reducing internal covariate shifts, batch normalization allows the model to converge faster during training. This enables the use of higher learning rates and accelerates the optimization process. Moreover, batch normalization acts as a form of regularization, reducing the need for other regularization or dropout techniques in models with few parameters. It can help prevent overfitting to some extent. Thus, batch normalization was processed following the dense layer on each hidden layer. The batch size is an essential hyperparameter in training neural network models, and its value can significantly impact the training process and model performance. The batch size represents the number of samples processed together in a single forward and backward pass during each training iteration. For the small dataset, the batch size was set to 16, but in this research, a batch size of 64 was consistent with the number of data in the dataset and gave a better model performance. Besides, the Adaptive Moment Estimation (Adam) optimizer, which adapts the learning rates of individual parameters by calculating an adaptive learning rate for each parameter based on their historical gradients, was applied with a learning rate was 0.001 during the MLPNN models training. Finally, in the output layer, the number of outputs was the same as the number of classes where the activation function was Softmax as default. However, the activation function will apply the MCSVM classifier for the output layer. In addition, the number of MLPNN learning cycles was set at 600 epochs. The total number of parameters used to train the model was 529 parameters. The layer architecture and configuration of the MLPNN-based Softmax or MCSVM model are described in Table 3 and illustrated in Figure 6.

Table 3: The layer configuration of the MLPNN-based model

Layer	No. of input	No. of output	Activation function/Classifier	No. of parameter
Input	6	6	n/a	0
Hidden 1	6	15	ReLU	105
Hidden 2	15	13	ReLU	208
Hidden 3	13	9	ReLU	126
Output	9	9	Softmax (or MCSVM)	90

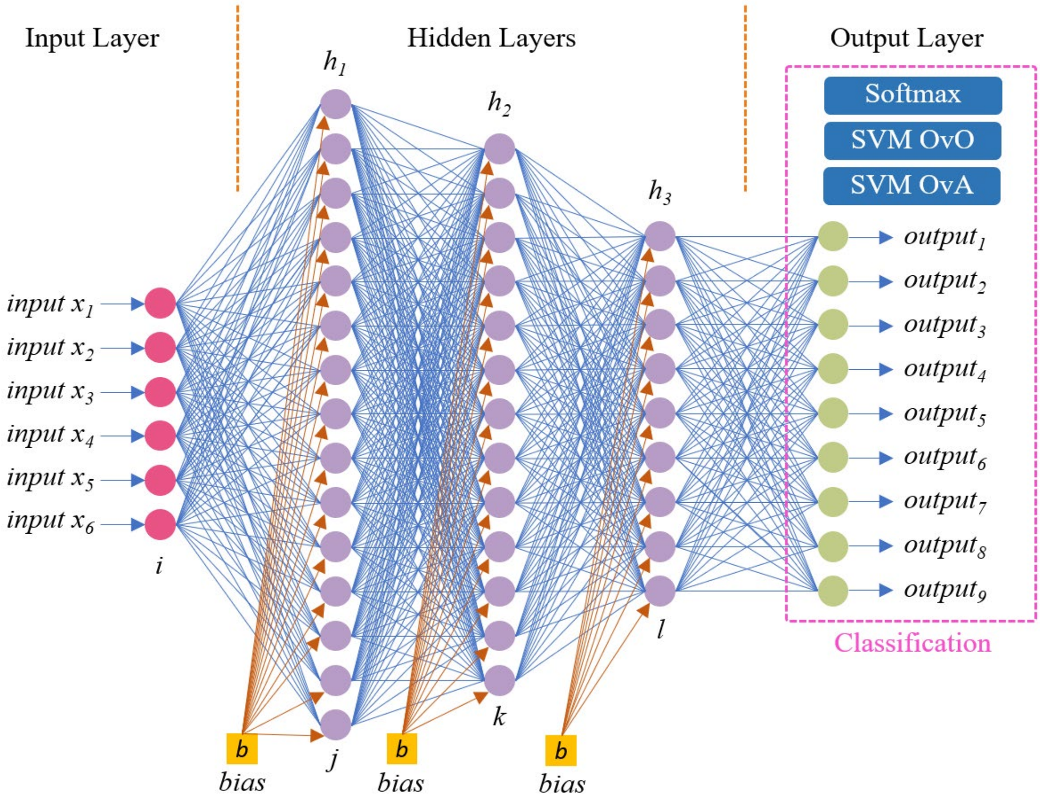


Fig. 6: The proposed MLPNN-based model architecture

3.6. Blending ANN and MCSVM

This process is a combination of the MLPNN-based model and other classifiers. In general, the Softmax activation function is implemented in the output layer as a result class classifier. It classifies the feature extraction from the last hidden layer of the model. Thus, the final step can be modified to use another classifier that may result in a different classifier. Some studies have found that some classifiers, such as SVM, produce better results than MLPNN (Zhu et al., 2022). SVM is inherently designed to find the optimal hyperplane that separates linearly separable classes. If the data is linearly separable, SVM can achieve excellent results without the need for complex non-linear transformations. In contrast, MLPNN might require deeper architectures and complex non-linear activation functions to handle non-linearly separable data. Thus, this work applied the MCSVM to classify the feature extraction from the MLPNN-based models (see Figure 6). The MCSVM is based on several binary classification problems between support vectors of two classes and the maximum margin through the hyperplanes. There are two popular approaches used in MCSVM: OvO and OvA. For the SVM OvO approach, the data was split into binary classes where each class against the other class was considered the decision hyperplane. Assuming c refers to the total number of classes, the number of classifiers or against times (t) to split the data for decision hyperplane separates the two classes' probability can be calculated as in (8).

$$t = \frac{1}{2}\left(c \times (c-1)\right) \tag{8}$$

While the SVM OvA approach split the data of one class for separating hyperplane against others where the against times was c, this SVM OvA approach takes less computing time than the SVM OvO approach. The decision hyperplane based on linear regression is defined in (9) and (10) (Meng & Wu, 2022).

$$w^T x + b = 0 \tag{9}$$

$$\begin{cases} w^T x + b > 0, & if\ 1\ for\ positive\ side\ hyperplane \\ w^T x + b < 0, & if\ -1\ for\ negative\ side\ hyperplan \end{cases} \tag{10}$$

where w refers to the weight vector, x refers to the input feature vector or data point to be classified, and b refers to the bias.

Thus, MCSVM will be classified between an individual or all classes in Lagrangian multipliers, as in (11) (Guo et al., 2021).

$$\min \frac{1}{2}\sum_{k=1}^{K}\|w\|_2^2 + C\sum_{i=1}^{n}\sum_{k \neq y_i}\xi_{ik}, \tag{11}$$

$$\textit{subject to } (w_{yi} - w_k).\phi(x_i) + b_{y_i} - b_k \geq 1 - \xi_{ik},\ \xi_{ik} \geq 0$$

where $||w||$ refers to the Euclidean norm, K refers to the variables that must be classified, C refers to the hyperparameter for n samples, x_i refers to the input vector of i-th sample, and y_i refers to the output of i-th sample.

However, this research has developed the MLPNN-based model to compare the model efficiency under different classifiers between Softmax, SVM OvO, and SVM OvA. Thus, three models have been developed to classify water quality for various farms.

3.7. The Model Evaluation

There are three developed models in this work with differences in classifiers for the output layer of MLPNN: Softmax, SVM OvO, and SVM OvA. The evaluation metrics of the models include accuracy, precision, recall, and F1-score, which are most commonly used to evaluate model performance in many works. In addition, the Mean Absolute Error (MAE) was used to evaluate the models because of it is robustness to outliers and equal weighting of errors that might be desirable when each data point's prediction error carries the same importance. It provides a clear and interpretable measure of the average prediction error in the original unit of the target variable and sometimes results in faster convergence during training. Moreover, the categorical cross-entropy loss was used to evaluate the model efficiency in this work. It is a widely used loss function for evaluating multiclass classification models and provides a strong signal during training. The loss decreases as the model learns to classify examples correctly, providing a clear optimization objective for the learning process. Further, the confusion matrix was also used for evaluating the classification of different classes. This provides a better understanding of the classification accuracy of each result class. These evaluation metrics can be formulae in (12) to (17) (Huo et al., 2021; Kaddoura, 2022; Nagarajaiah et al., 2022; Nuanmeesri, 2022, Nuanmeesri et al., 2021).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F1\text{-}score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{15}$$

where:

TP refers to the true positive predicted of the water quality class that is corrected to the actual positive value;

TN refers to the true negative predicted of the water quality class that is corrected to the actual negative value;

FP refers to the false positive predicted of the water quality class that is incorrect to the actual positive value;

FN refers to the false negative predicted of the water quality class that is incorrect to the actual negative value.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{16}$$

where N refers to the number of instances, y_i refers to the label of an instance i-th, and $\hat{y}_i$ refers to the average value of all instances.

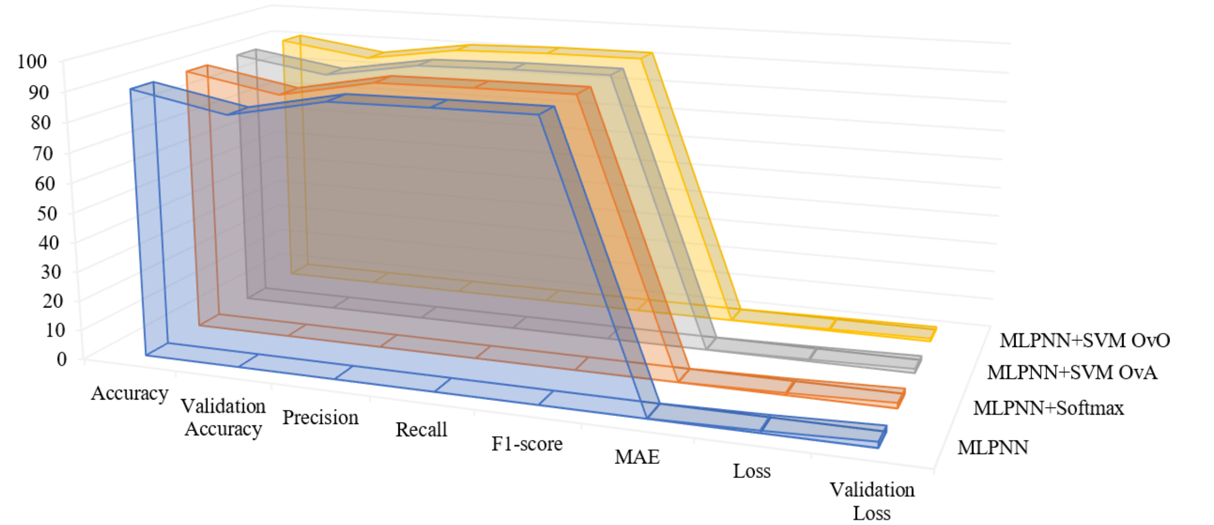
$$Cross\text{-}entropy\ Loss = -\frac{1}{M}\sum_{k=1}^{K}\sum_{m=1}^{M} y_m^k \log(h_\theta(x_m, k)) \tag{17}$$

where f' K refers to the class total, M refers to the input data total, y_m^k refer to the target label for input m-th that corresponds to class k-th, h_θ refers to the model with weights θ, and x_m refers to the input m-th.

4. Results and Discussion

4.1. The Efficiency Result of the Developed MLPNN-Based Models

The results evaluation of the technique for optimizing predictive modeling for water quality in various farms with a blended artificial neural network and multi-class SVM of the three models is shown in Figure 7.



	Accuracy	Validation Accuracy	Precision	Recall	F1-score	MAE	Loss	Validation Loss
MLPNN	90.24	84.11	90.62	91.04	91.23	0.11013	0.80103	1.95021
MLPNN+Softmax	90.31	84.49	90.69	91.11	91.30	0.10928	0.75364	1.80755
MLPNN+SVM OvA	90.58	85.54	90.78	91.53	91.63	0.10626	0.64998	1.45402
MLPNN+SVM OvO	90.69	86.11	90.85	91.66	91.75	0.10497	0.60126	1.01671

Fig. 7: The efficiency result of the developed MLPNN-based models

According to Figure 7, the result has shown that the model developed using blended MLPNN-based with SVM OvO (MLPNN+SVM OvO) has the highest efficiency during training processing compared to other models. The efficiency of this model is as follows: the accuracy was 90.69%, the precision was 90.85%, the recall was 96.66%, the F1-score was 91.75%, the MAE was 0.10497, and the loss was 0.60126. The models with the next efficiency are MLPNN+SVM OvA, MLPNN+Softmax, and MLPNN with accuracy of 90.58%, 90.31%, and 90.24%; the MAE of 0.10624, 0.10928, and 0.11013, respectively. The model blended the SVM OvO classifier at the output layer of MLPNN; it has higher efficiency than SVM OvA, Softmax classifier, and MLPNN-based, respectively. The experimental results found that the MLPNN-based model, which had an accuracy of 90.24%, when the classifiers were modified to Softmax, SVM OvA, and SVM OvO, the model efficiency slightly increased, respectively. Increased precision and recall values can determine it. The increase in precision is due to decreased false positives while reducing false negatives leads to higher recall values. In contrast, MAE and categorical cross-entropy loss gradually decreased with Softmax, SVM OvA, and SVM OvO, respectively. It confirms that blending MLPNN and MCSVM can improve model performance even further, consistent with research by Garg et al. (2021) that integrating MLP and SVM can improve model performance. Consider the results from existing studies that classify water quality, as shown in Table 4.

Table 4: The comparison of the proposed method with existing studies

Research	Method	No. of			Accuracy
		Features	Classes	Samples	
Al-Adhaileh & Alsaade (2021)	Feed-Forward Neural Network	8	4	1,679	100.00%
Ladjal et al. (2023)	PCA + (ANN fusion SVM) + DT	4	3	1,800	99.24%
Ladjal et al. (2023)	PCA+MPLNN	4	3	1,800	99.13%
Ladjal et al. (2023)	PCA+SVM	4	3	1,800	98.48%
Aldhyani et al. (2020)	SVM	7	5	1,679	97.01%
Radhakrishnan & Pillai (2020)	SVM	4	5	550	95.63%
Sillberg et al. (2021)	SVM	6	4	9,780	95.00%
Our proposed method	MLPNN + SVM OvO	6	9	4,997	90.69%
Our proposed method	MLPNN + SVM OvA	6	9	4,997	90.58%
Our proposed method	MLPNN + Softmax	6	9	4,997	90.31%
Our proposed method	MLPNN	6	9	4,997	90.24%
Radhakrishnan & Pillai (2020)	SVM	4	5	134	87.10%
Ladjal et al. (2023)	SVM	8	3	1,800	86.88%
Ladjal et al. (2023)	MLPNN	8	3	1,800	86.61%
Ahmed et al. (2019)	MLPNN	4	5	663	85.07%
Ahmed et al. (2019)	SVM	4	5	663	79.79%

According to Table 4, the validity values obtained from other studies were higher than in this work. For example, Al-Adhaileh & Alsaade (2021) used the Feed-Forward Neural Network method with 100.00% accuracy. Next was Ladjal et al. (2023), which used the PCA method as feature selection to reduce the number of input features and then modeled with ANN or SVM with an accuracy of at least 98.48%. The results of Aldhyani et al. (2020), Radhakrishnan & Pillai (2020), and Sillberg et al. (2021) all used SVM to classify water quality with an accuracy of 95.00% or higher. When looking at the number of classes, these studies found only between three and five classes, which is less than the number of classes in this study (nine classes). This increased number of classes may also result in a lower accuracy rate due to increased classification complexity resulting in a higher chance of misclassification. In some cases, when the number of classes is imbalanced, the model might focus on the majority classes and perform poorly on the minority classes. This can lead to an overall decrease in accuracy. It is also seen that these studies use water quality parameters with only four features, which may result in less complex feature extraction compared to models using eight features without PCA feature selection. When the number of features increases, the model performance decreases as well. Therefore, appropriate feature selection techniques should be considered to assist in selecting features affecting the result class, increasing the model's efficiency.

4.2. The Training and Validation Result

Considering the efficiency during the validation stage, the MLPNN-based model blended with SVM OvO (MLPNN+SVM OvO) is the suitable model; it had a validation accuracy of 86.11%, which was a decrease of 4.58%, while the validation loss was 1.01671 which was about 0.42 higher than the loss of

0.60126, as shown in Figure 8.

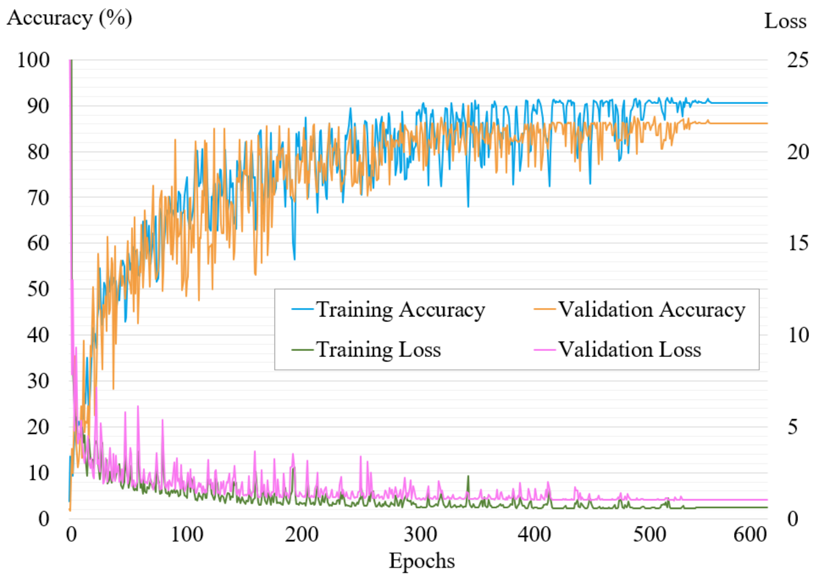


Fig. 8: The comparison of training and validation of the MLPNN+SVM OvO model

According to Figure 8, it was found that the training and validation accuracy had a few exchange rates from the 519th epoch until it was almost constant from the 445th epoch onwards. The efficiencies of the remaining three models were similar: validation accuracy decreased, and validation loss increased. Comparing the validation accuracy and validation loss of the remaining three models, it was found that the validation accuracy was 85.54%, 84.49%, and 84.11% for the MLPNN+SVM OvA, MLPNN+Softmax, and MLPNN models. Further, the validation loss was 1.45402, 1.80755, and 1.95021, respectively (see Figure 7). It has been shown that blending MLPNN with MCSVM can improve model performance.

Moreover, the classification accuracy for each class can be presented in the normalized confusion matrix, as shown in Figure 9. The result of the normalized confusion matrix has shown that the predicted class ‘C3’, a WQSL at Level 3 (critical) for corp farms, has the highest accuracy of 0.9274. This means that water quality classification, a ‘critical’ class for corp farms, was accurate up to 92.74%. The following predicted classes are ‘F3’ (a critical level for fish farms) and ‘S3’(a critical level for salt farms), with an accuracy of 0.9201 and 0.9106, respectively. These three classes deal with water quality, which is critically important for corp, fish, and salt farming. In addition, water quality classification within the same farm type found that classes ‘C2’, ‘F2’, and ‘S2’, which are WQSL at level 2 (caution), had higher accuracy than classes ‘C1’, ‘F1’, and ‘S1’ (WQSL at level 1: normal) for the MLPNN-based SVM OvO model. According to experimental results, it was found that the MLPNN-based SVM OvO model can best classify the water quality values in cases where the water quality values are at critical levels. These will help farmers accurately recognize the criticality of water quality through this developed model. The feature extraction obtained from the MLPNN-based model consists of three hidden layers, and the number of neurons is proportional to the number of classes and features. This may be a limitation of the experiment. Therefore, designing a neural network architecture may conduct further experiments to compare the number of hidden layers and different neurons and customize additional hyperparameters, which may contribute to the optimal model for water quality classification.

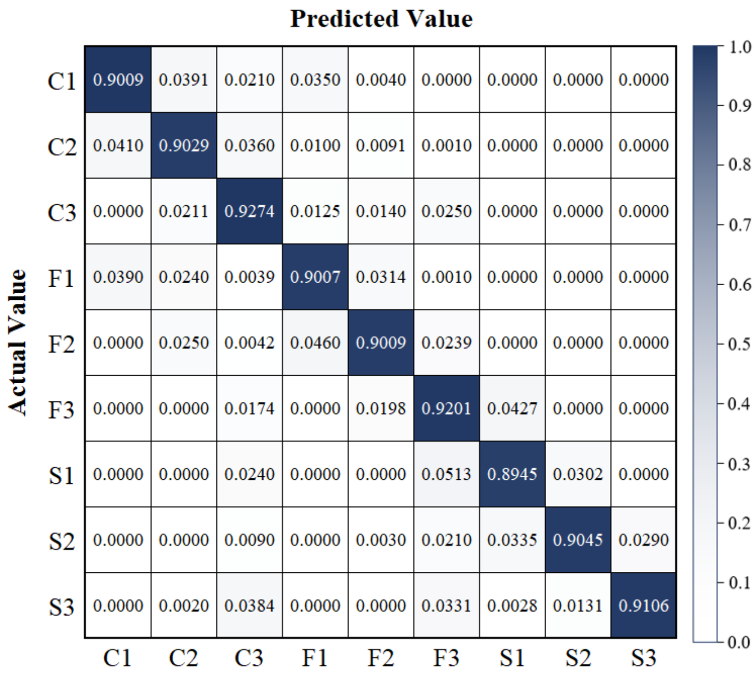


Fig. 9: The normalized confusion matrix of the MLPNN+SVM OvO model

5. Conclusions

Most water quality research focuses on single water sources in specific contexts. In contrast, this study analyzed multiple connected coastal water sources affecting diverse occupations like aquaculture, salt production, and agriculture. Additionally, few studies combine Multi-Layer Perceptron Neural Networks (MLPNN) and Multi-Class Support Vector Machines (MCSVM) for water quality classification. This research blended MLPNN and MCSVM algorithms to analyze and predict coastal water quality impacts on farming. Since real-world data can have missing values, Multiple Generative Adversarial Imputation Networks filled gaps, improving model performance. MLPNN models with different output classifiers like MCSVM One-vs-One, MCSVM One-vs-All, and Softmax were compared, finding One-vs-One performed best. The optimized MLPNN-MCSVM model effectively classified critical water quality levels for all farm types, demonstrating its ability to identify high-risk conditions. The experimental results show that the feature extraction obtained from the last hidden layer of the MLPNN-based model can be further categorized by MCSVM or other conventional machine learning techniques. Further, choosing a classifier, the amount of data, the number of attributes, the number of classes, the amount of data, and the data type in the dataset should be considered appropriately to obtain a model that can be used optimally. In summary, blended neural networks and machine learning show promise for modeling complex coastal water systems to inform policy and management. However, this experiment and the amount of data were collected for a limited period (about two months). If the volume of data is gradually increasing, which is large or equivalent to big data, the model’s learning will be more accurate. This may also result in variations in model performance, particularly MCSVM One-vs-All, which has fewer classifiers or against times than MCSVM One-vs-One, resulting in less processing time then and may answer the question in some datasets. In addition, several parameters of water quality affect the water quality index. Thus, increasing the number of water quality parameters will help to make the model more accurate according to the water environment. There are other environmental factors, for example, the amount of rainfall in the rainy season, heat radiated from the sun, and rising and falling sea conditions causing water quality values to change throughout the year. Therefore, water quality data should be collected throughout the year, increasing the frequency of measuring and reading the water quality from every hour to every ten minutes. Also, consider the frequency and response time of each sensor. Future work could further optimize hyperparameters, incorporate multimodal data like images, add recurrent neural networks for time series forecasting, and develop user applications. This could continue improving predictive capabilities and support for coastal industries facing water quality challenges.

Acknowledgements

The authors express their gratitude to the Institute for Research and Development of Suan Sunandha Rajabhat University for their support and the opportunity to conduct this research. The research conducted in this study received support from the national budget for the year 2023, provided by Suan Sunandha Rajabhat University. The project was carried out under grant/contract number 11803/2566, with a research ethics certificate number COE.1-039/2023.

References

Abuzir, S. Y., & Abuzir, Y. S. (2022). Machine learning for water quality classification. *Water Quality Research Journal*, 57(3), 152-164. https://doi.org/10.2166/wqrj.2022.004

Ahmad, Z., Rahim, N. A., Bahadori, A., & Zhang, J. (2016). Improving water quality index prediction in Perak River basin Malaysia through a combination of multiple neural networks. *International Journal of River Basin Management*, 15(1), 79-87. https://doi.org/10.1080/15715124.2016.1256297

Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R., & García-Nieto, J. (2019). Efficient water quality prediction using supervised machine learning. *Water*, 11(11). https://doi.org/10.3390/w11112210

Al-Adhaileh, M. H., & Alsaade, F. W. (2021). Modelling and prediction of water quality by using artificial intelligence. *Sustainability*, 13(8). https://doi.org/10.3390/su13084259

Aldhyani, T. H. H., Al-Yaari, M., Alkahtani, H., Maashi, M., & Algalil, F. A. (2020). Water quality prediction using artificial intelligence algorithms. *Applied Bionics and Biomechanics*, 2020, 1-12. https://doi.org/10.1155/2020/6659314

Belghit, A., Lazri, M., Ouallouche, F., Labadi, K., & Ameur, S. (2023). Optimization of One versus All-SVM using AdaBoost algorithm for rainfall classification and estimation from multispectral MSG data. *Advances in Space Research*, 71(1), 946-963. https://doi.org/10.1016/j.asr.2022.08.075

Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H., & Kazakis, N. (2020). Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science of The Total Environment*, 721. https://doi.org/10.1016/j.scitotenv.2020.137612

Chawla, P., Cao, X., Fu, Y., Hu, C.-m., Wang, M., Wang, S., & Gao, J. Z. (2021). Water quality prediction of salton sea using machine learning and big data techniques. *International Journal of Environmental Analytical Chemistry*, 1-24. https://doi.org/10.1080/03067319.2021.1963713

de Luna Souto, A. G., Cavalcante, L. F., de Melo, E. N., Cavalcante, Í. H. L., da Silva, R. Í. L., de Lima, G. S., Gheyi, H. R., Pereira, W. E., de Paiva Neto, V. B., de Oliveira, C. J. A., & de Oliveira Mesquita, F. (2023). Salinity and Mulching Effects on Nutrition and Production of Grafted Sour Passion Fruit. *Plants*, 12(5). https://doi.org/10.3390/plants12051035

Emamgholizadeh, S., Kashi, H., Marofpoor, I., & Zalaghi, E. (2013). Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *International Journal of Environmental Science and Technology*, 11(3), 645-656. https://doi.org/10.1007/s13762-013-0378-x

Gaafar, M., Mahmoud, S. H., Gan, T. Y., & Davies, E. G. R. (2020). A practical GIS-based hazard assessment framework for water quality in stormwater systems. *Journal of Cleaner Production*, 245. https://doi.org/10.1016/j.jclepro.2019.118855

Gakii, C., & Jepkoech, J. (2019). A classification model for water quality analysis using Decision Tree. *European Journal of Computer Science and Information Technology*, 17(3), 1-8.

Garg, G., Kumar, D., Pal, A., Sonker, Y., & Garg, R. (2021). A hybrid MLP-SVM model for classification using spatial-spectral features on hyper-spectral images. *In arXiv e-prints* (pp. arXiv:2101.00214 [cs.CV]).

Giri, A., & Singh, N. B. (2014). Comparison of Artificial Neural Network algorithm for water quality prediction of River Ganga. *Environmental Research Journal*, 8(2), 55-63.

Günther, F., & Fritsch, S. (2010). neuralnet: Training of Neural Networks. *The R Journal*, 2(1). https://doi.org/10.32614/rj-2010-006

Guo, Y., Zhang, Z., & Tang, F. (2021). Feature selection with kernelized multi-class support vector machine. *Pattern Recognition*, 117. https://doi.org/10.1016/j.patcog.2021.107988

Haghiabi, A. H., Nasrolahi, A. H., & Parsaie, A. (2018). Water quality prediction using machine learning methods. *Water Quality Research Journal*, 53(1), 3-13. https://doi.org/10.2166/wqrj.2018.025

Harpaz, S., Hakim, Y., Slosman, T., & Eroldogan, O. T. (2005). Effects of adding salt to the diet of Asian sea bass Lates calcarifer reared in fresh or salt water recirculating tanks, on growth and brush border enzyme activity. *Aquaculture*, 248(1-4), 315-324. https://doi.org/10.1016/j.aquaculture.2005.03.007

Hassan, H. U., Ali, Q. M., Ahmed, A. E., Gabol, K., Swelum, A. A., Masood, Z., Mushtaq, S., Saeed, Gul, Y., Rizwan, S., Zulfiqar, T., & Siddique, M. A. M. (2024). Growth performance and survivability of the Asian seabass Lates calcarifer (Bloch, 1790) reared under hyper-saline, hypo-saline and freshwater environments in a closed aquaculture system. *Brazilian Journal of Biology*, 84. https://doi.org/10.1590/1519-6984.254161

Hassan, M. M., Akter, L., Rahman, M. M., Zaman, S., Hasib, K. M., Jahan, N., Smrity, R. N., Farhana, J., Raihan, M., & Mollick, S. (2021). Efficient prediction of water quality index (WQI) using machine learning algorithms. *Human-Centric Intelligent Systems*, 1(3-4). https://doi.org/10.2991/hcis.k.211203.001

Ho, J. Y., Afan, H. A., El-Shafie, A. H., Koting, S. B., Mohd, N. S., Jaafar, W. Z. B., Lai Sai, H., Malek, M. A., Ahmed, A. N., Mohtar, W. H. M. W., Elshorbagy, A., & El-Shafie, A. (2019). Towards a time and cost effective approach to water quality index class prediction. *Journal of Hydrology*, 575, 148-165. https://doi.org/10.1016/j.jhydrol.2019.05.016

Hong, L. C., Tee, C., & Goh, M. K. O. (2022). Activities of daily living recognition using deep learning approaches. *Journal of Logistics, Informations and Service Science*, 9(4), 12-148. https://doi.org/10.33138/LISS.2022.0410

Huo, W., Li, W., Zhang, Z., Sun, C., Zhou, F., & Gong, G. (2021). Performance prediction of proton-exchange membrane fuel cell based on convolutional neural network and random forest feature selection. *Energy Conversion and Management*, 243. https://doi.org/10.1016/j.enconman.2021.114367

Islam, A. R. M. T., Kabir, M. M., Faruk, S., Al Jahin, J., Bodrud-Doza, M., Didar-ul-Alam, M., Bahadur, N. M., Mohinuzzaman, M., Fatema, K. J., Safiur Rahman, M., & Choudhury, T. R. (2021). Sustainable groundwater quality in southeast coastal Bangladesh: co-dispersions, sources, and probabilistic health risk assessment. *Environment, Development and Sustainability*, 23(12), 18394-18423. https://doi.org/10.1007/s10668-021-01447-4

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. https://doi.org/10.1016/j.eswa.2022.117160

Kaddoura, S. (2022). Evaluation of machine learning algorithm on drinking water quality for better sustainability. *Sustainability*, 14(18). https://doi.org/10.3390/su141811478

Kasasbeh, B., Aldabaybah, B., & Ahmad, H. (2022). Multilayer perceptron artificial neural networks-based model for credit card fraud detection. *Indonesian Journal of Electrical Engineering and Computer Science*, 26(1). https://doi.org/10.11591/ijeecs.v26.i1.pp362-373

Kaur, G., Braveen, M., Krishnapriya, S., Wawale, S. G., Castillo-Picon, J., Malhotra, D., Osei-Owusu, J., & Khan, R. (2023). machine learning integrated multivariate water quality control framework for prawn harvesting from fresh water ponds. *Journal of Food Quality*, 2023, 1-9. https://doi.org/10.1155/2023/3841882

Kim, B., Lee, S.-m., Kang, S.-h., Jeong, M.-s., Gim, G. H., Park, J., & Lim, C. (2020). Aquavoltaic system for harvesting salt and electricity at the salt farm floor: Concept and field test. *Solar Energy Materials and Solar Cells*, 204. https://doi.org/10.1016/j.solmat.2019.110234

Ladjal, M., Bouamar, M., Brik, Y., & Djerioui, M. (2023). A decision fusion method based on classification models for water quality monitoring. *Environmental Science and Pollution Research*, 30, 22532-22549. https://doi.org/10.1007/s11356-022-23418-6

Li, H., Cui, Z., Cui, H., Bai, Y., Yin, Z., & Qu, K. (2023). Hazardous substances and their removal in recirculating aquaculture systems: A review. *Aquaculture*, 569. https://doi.org/10.1016/j.aquaculture.2023.739399

López, J., & Maldonado, S. (2016). Multi-class second-order cone programming Support Vector Machines. *Information Sciences*, 330, 328-341. https://doi.org/10.1016/j.ins.2015.10.016

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

Malek, N. H. A., Wan Yaacob, W. F., Md Nasir, S. A., & Shaadan, N. (2022). Prediction of water quality classification of the Kelantan River Basin, Malaysia, using machine learning techniques. *Water*, 14(7). https://doi.org/10.3390/w14071067

Manoj, K. D. P, & Ananda, B. J. (2022). Neural network-based game theory approach for personalized privacy preservation in data publishing. *Journal of System and Management Sciences*, 12(1), 498-520. https://doi.org/10.33168/JSMS.2022.0132

Meng, L., & Wu, C.-H. (2022). The promotion effect of the improved ISCA model on the application of accounting informatization in small- and medium-sized enterprises in the cloud computing environment. *Mobile Information Systems*, 2022, 1-13. https://doi.org/10.1155/2022/4228178

Mishra, R. K. (2023). Fresh water availability and its global challenge. *British Journal of Multidisciplinary and Advanced Studies*, 4(3), 1-78. https://doi.org/10.37745/bjmas.2022.0208

Muhammad, S. Y., Makhtar, M., Rozaimee, A., Aziz, A. A., & Jamal, A. A. (2015). Classification model for water quality using machine learning techniques. *International Journal of Software Engineering and Its Applications*, 9(6), 45-52.

Nagarajaiah, K., Chandramouli, S. M., & Ramakrishna, L. M. (2022). Mobile app usage behaviour based user authentication using Fuzzy Random Forest. *Journal of System and Management Sciences*, 12(2), 497-511. https://doi.org/10.33168/JSMS.2022.0227

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. https://doi.org/10.1016/j.jwpe.2022.102920

Nuanmeesri, S. (2022). Development of community tourism enhancement in emerging cities using gamification and adaptive tourism recommendation. *Journal of King Saud University - Computer and Information Sciences*, 34(10), 8549-8563. https://doi.org/10.1016/j.jksuci.2021.04.007

Nuanmeesri, S., Chopvitayakun, S., Kadmateekarun, P., & Poomhiran, L. (2021). Marigold flower disease prediction through deep neural network with multimodal image. International Journal of Engineering Trends and Technology, 69(7), 174-180. https://doi.org/10.14445/22315381/IJETT-V69I7P224

Nuanmeesri, S., Poomhiran, L., Chopvitayakun, S., & Kadmateekarun, P. (2022). Improving dropout forecasting during the COVID-19 pandemic through feature selection and Multilayer Perceptron Neural Network. *International Journal of Information and Education Technology*, 12(9), 851-857. https://doi.org/10.18178/ijiet.2022.12.9.1693

Ömer Faruk, D. (2010). A hybrid neural network and ARIMA model for water quality time series prediction. *Engineering Applications of Artificial Intelligence*, 23(4), 586-594. https://doi.org/10.1016/j.engappai.2009.09.015

Oostwal, E., Straat, M., & Biehl, M. (2021). Hidden unit specialization in layered neural networks: ReLU vs. sigmoidal activation. Physica A: *Statistical Mechanics and its Applications*, 564. https://doi.org/10.1016/j.physa.2020.125517

Paltán, H. A., Pant, R., Plummer Braeckman, J., & Dadson, S. J. (2021). Increased water risks to global hydropower in 1.5 °C and 2.0 °C warmer worlds. *Journal of Hydrology*, 599. https://doi.org/10.1016/j.jhydrol.2021.126503

Radhakrishnan, N., & Pillai, A. S. (2020). Comparison of water quality classification models using machine learning. In *the 2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 1183-1188. https://doi.org10.1109/icces48766.2020.9137903

Rahmati, O., Choubin, B., Fathabadi, A., Coulon, F., Soltani, E., Shahabi, H., Mollaefar, E., Tiefenbacher, J., Cipullo, S., Ahmad, B. B., & Tien Bui, D. (2019). Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. *Science of The Total Environment*, 688, 855-866. https://doi.org/10.1016/j.scitotenv.2019.06.320

Rajaee, T., Khani, S., & Ravansalar, M. (2020). Artificial intelligence-based single and hybrid models for prediction of water quality in rivers: A review. *Chemometrics and Intelligent Laboratory Systems*, 200. https://doi.org/10.1016/j.chemolab.2020.103978

Sakaa, B., Elbeltagi, A., Boudibi, S., Chaffaï, H., Islam, A. R. M. T., Kulimushi, L. C., Choudhari, P., Hani, A., Brouziyne, Y., & Wong, Y. J. (2022). Water quality index modeling using random forest and improved SMO algorithm for support vector machine in Saf-Saf river basin. *Environmental Science and Pollution Research*, 29(32), 48491-48508. https://doi.org/10.1007/s11356-022-18644-x

Sengorur, B., Koklu, R., & Ates, A. (2015). Water quality assessment using artificial intelligence techniques: SOM and ANN—A case study of Melen River Turkey. *Water Quality, Exposure and Health*, 7(4), 469-490. https://doi.org/10.1007/s12403-015-0163-9

Shan, W., Cai, S., & Liu, C. (2018). A new comprehensive evaluation method for water quality: Improved Fuzzy Support Vector Machine. *Water*, 10(10). https://doi.org/10.3390/w10101303

Sillberg, C., Kullavanijaya, P., & Chavalparit, O. (2021). Water quality classification by integration of attribute-realization and Support Vector Machine for the Chao Phraya River. *Journal of Ecological Engineering*, 22(9), 70-86. https://doi.org/10.12911/22998993/141364

Silva, W. A., & Villela, S. M. (2020). Improving the one-against-all binary approach for multiclass classification using balancing techniques. *Applied Intelligence*, 51(1), 396-415. https://doi.org/10.1007/s10489-020-01805-1

Swetha, P., Rasheed, A. H. K. P., & Harigovindan, V. P. (2023). Random Forest Regression based water quality prediction for smart aquaculture. In *the 2023 4th International Conference on Computing and Communication Systems (I3CS)*, 1-5. https://doi.org/10.1109/i3cs58314.2023.10127488

Tan, G., Yan, J., Gao, C., & Yang, S. (2012). Prediction of water quality time series data based on least squares support vector machine. *Procedia Engineering*, 31, 1194-1199. https://doi.org/10.1016/j.proeng.2012.01.1162

Vijay, S., & Kamaraj, K. (2019). Ground water quality prediction using machine learning algorithms in R. *International Journal of Research and Analytical Reviews*, 6(1), 743-749.

Wang, Y., Zheng, T., Zhao, Y., Jiang, J., Wang, Y., Guo, L., & Wang, P. (2013). Monthly water quality forecasting and uncertainty assessment via bootstrapped wavelet neural networks under missing data for Harbin, China. *Environmental Science and Pollution Research*, 20(12), 8909-8923. https://doi.org/10.1007/s11356-013-1874-8

Yoon, J., Jordon, J., & van der Schaar, M. (2018). GAIN: Missing Data imputation using Generative Adversarial Nets. *In arXiv e-prints* (pp. arXiv:1806.02920 [cs.LG]).

Yumang, A. N., Cruz, G. A., & Fontanilla, L. A. (2020). Combination of acoustic and vibration sensor data using Support Vector Machines and One-Versus-All technique data fusion for detecting objects. In *Proceedings of the 2020 12th International Conference on Computer and Automation Engineering*, 56-59. https://doi.org/10.1145/3384613.3384626

Zhang, Q., Wang, H., & Yoon, S. W. (2020). A 1-norm regularized linear programming nonparallel hyperplane support vector machine for binary classification problems. *Neurocomputing*, 376, 141-152. https://doi.org/10.1016/j.neucom.2019.09.068

Zhao, F., Lu, Y., Li, X., Wang, L., Song, Y., Fan, D., Zhang, C., & Chen, X. (2022). Multiple imputation method of missing credit risk assessment data based on Generative Adversarial Networks. *Applied Soft Computing*, 126, 109273. https://doi.org/10.1016/j.asoc.2022.109273

Zhu, M., Wang, J., Yang, X., Zhang, Y., Zhang, L., Ren, H., Wu, B., & Ye, L. (2022). A review of the application of machine learning in water quality evaluation. *Eco-Environment & Health*, 1(2), 107-116. https://doi.org/10.1016/j.eehl.2022.06.001