SEMARAK ILMU PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

A Comparison of Regression Models for Dissolved Oxygen Prediction in Koh Yor, Thailand

Warisara Na lampang[1], Apidet Booranawong[1], Nattha Jindapetch[1], Kiattisak Sengchuai[1,*]

[1] Department of Electrical and Biomedical Engineering, Faculty of Engineering, Prince of Songkla University, Songkhla, Thailand

ARTICLE INFO

Article history:
Received 2 January 2024
Received in revised form 12 June 2024
Accepted 21 August 2024
Available online 19 September 2024

Keywords:
Dissolved oxygen prediction; regression models; Gaussian process regression

ABSTRACT

Dissolved oxygen is an indicator of the health of a water source and aquatic life. In Koh Yor, fishermen encountered the problem of sea bass dying from changes in dissolved oxygen and did not have information on dissolved oxygen levels in the water. In this paper, a system and methodology for dissolved oxygen prediction in Koh Yor water, Thailand, for breeding sea bass in cages are introduced. The major objective of this study is to find the optimal regression models that achieve the best prediction accuracy, where soft sensor information including temperature, pH, and salinity data are used. The regression model is suitable for predicting DO in open systems with non-linear water quality data. The regression models for evaluation and comparison include Gaussian Process Regression (GPR), Medium Gaussian SVM (SVM), Least Squares Regression (LSR), and Medium Neural Network (MNN), respectively. The performances of the regression models are validated with the water data set of each village collected from Koh Yor, Thailand. Experimental results demonstrate that the GPR model provides the best prediction accuracy of 91.8%, where the prediction accuracy of all villages around Koh Yor is over 90%.

1. Introduction

Water quality is an important factor in aquaculture, where water quality depends on changes in water quality parameters. Physicochemical and biological characteristics can explain the uncertainty of water quality parameters and their dynamic changes [1-3]. In addition, environmental factors and human activities around water sources also directly affect water quality. Koh Yor, Songkhla, Thailand, is the second-largest area for breeding sea bass in cages in Thailand. In recent years, information from the Coastal Aquatic Animal Technology and Innovation Research and Development Centre, Songkhla, Thailand shows that Koh Yor fishermen have encountered problems with the death of sea bass. The cause of this problem is a sudden change in the environment, and fishermen are unable to prevent and correct sudden changes in water quality in time. As a result, the amount of dissolved oxygen in the water decreased rapidly.

[*] *Corresponding author.*
E-mail address: kiattisak.se@psu.ac.th

https://doi.org/10.37934/araset.51.2.227240

Water quality assessment should be based on water quality parameters including salinity, pH, transparency, temperature, dissolved oxygen (DO), BOD, turbidity, and others, but one important parameter that indicates the health and state of an aquatic ecosystem is dissolved oxygen [4-6]. DO levels serve as an important indicator of water quality. Sufficient DO is necessary for the decomposition of organic matter in water. Bacteria that break down organic material require oxygen for their metabolic processes. Insufficient levels of DO, often referred to as hypoxia, can be harmful to aquatic organisms [7,8]. Fish and other species may experience stress, reduced growth, or mortality when DO levels are too low. At present, water quality measurement with soft sensors is popular because measuring water quality using soft sensors involves using computational models or algorithms to estimate the water quality parameters indirectly, often based on data from other available sensors. Thus, soft sensors are particularly useful when direct measurements are difficult, expensive, or not feasible in real-time.

From a review of the literature on estimating or predicting the level of dissolved oxygen in water, it can be summarized in Table 1. There are several prediction models that have been applied to predict the level of DO. Popular ones include Artificial Neural Networks (ANN) [9-15], Adaptive Neuro-Fuzzy Inference System (ANFIS) [9], Least Squares Support Vector Regression (LSSVR) [16-19], Gaussian Process Regression (GPR) [20], K-Nearest Neighbor Regression (KNN-R) [22], Support Vector Machine Regression (SVR) [22], Extreme Learning Machine (ELM) [22], Long Short-Term Memory Neural Network (LSTM) [22] and Ensemble Techniques [22-24]. Several types of neural networks (NN) used to predict DO include Multilayer Perceptron Neural Network (MLPNN) [11,13], General Regression Neural Network (GRNN) [12], a deep learning-based model called Marine Deep Jointly Informed Neural Network (M-DJINN) [10] and Radial Basis Function Neural Network (RBFNN) [13,14]. Many types of regression models are also used for predicting DO including SVR, GPR, KNN-R, etc. The main objective of the entire article is to design and develop models to predict DO, evaluate the performance of the prediction models, and increase the prediction efficiency. Predicting dissolved oxygen requires water quality parameters such as pH, temperature, turbidity, salinity-specific conductance, suspended solids, total hardness, total alkalinity, and ammonium nitrogen as the input data for the model. The input data are trained until the optimal model is formed, and the output of the system is DO.

This paper presents a system and methodology for dissolved oxygen prediction in Koh Yor water that is caused by a combination of three types of water, including salt water, brackish water, and fresh water. The research gap is a study an optimal model for predicting dissolved oxygen that can assess water quality in all areas and can be implemented. The major objective of the study is to find the optimal regression models that achieve the best prediction accuracy using the data from soft sensors. The water quality parameters used for the study consist of dissolved oxygen, temperature, pH, and salinity. The regression models including Gaussian Process Regression (GPR), Medium Gaussian SVM (SVM), Least Squares Regression (LSR), and Medium Neural Network (MNN) are evaluated and compared. The performances of the best regression model are validated with the water data set of each village collected from Koh Yor, Thailand. The organization of the paper is as follows: Section 2 describes the system overview, the data collection, the prediction models, the experiments, and the performance evaluation. Section 3 presents the experimental results and discussion. In Section 4, the conclusions and the future work are given.

Table 1
A summary comparison of related works and this work

Refs	Objectives	Water quality parameters	Prediction models
[9]	To compare the models and estimate the DO in the Fei Tsui Reservoir of northern Taiwan	Input: temperature, pH, conductivity, turbidity, suspended solids, hardness, alkalinity, ammonium nitrogen Output: DO	ANN, BPNN, ANFIS, and MLR
[10]	To propose a model for predicting marine DO concentration	Input: temperature, salinity, phosphate Output: DO	M-DJINN
[11]	To model the DO concentration using the MRSM and MLPNN and compare performance	Input: river discharge, pH, conductance, turbidity Output: DO	MRSM and MLPNN
[12]	To evaluate the performance of the prediction models by using the mean absolute percentage error (MAPE) and the correlation coefficient (R-value) and assess prediction accuracy	Input: Temperature, Salinity, Conductivity, pH Output: DO	GRNN
[13]	To compare the model for modelling and forecasting DO in the Klamath River, USA	Input: pH, temperature, conductance, depth Output: DO	RBFNN and MLPNN
[14]	To predict the DO saturation level in the lower reaches using the principal component RBF network model	Input: unit discharge, depth, discharge amount, turbulence intensity Output: DO	RBFNN
[16]	To develop a prediction model for DO content in aquaculture	Input: pH, salinity, temperature, turbidity Output: DO	FOA and LSSVR
[17]	To propose and investigate the predictive performance of the regression models for predicting DO	Input: freshwater quality, stream and river health, severity of aquatic pollution Output: DO	LSSVR
[18]	To improve the model for modelling the DO concentration	Input: temperature, discharge, pH, conductance Output: DO	LSSVM-BA
[20]	To develop models for indirectly measure the DO level using machine learning techniques	Input: temperature, pH, conductance Output: DO	MLR, ANN-R, SVR, GPR, and KNN-R
[21]	To propose a regression model for predicting water quality in aquaculture	Input: pH, conductivity, water temperature, solar radiation, air temperature, wind speed Output: DO	RGA-SVR
[22]	To develop and compare the models for modelling the DO concentration in the Kinta River, Malaysia	Input: BOD, chemical oxygen demand (COD), temperature, ammonia (NH_3), Total solids (TS), chlorides (Cl), Calcium (Ca), pH, sodium (Na) Output: DO	LSTM, ELM, HW, and GRNN
This work	To study and compare the regression models for DO prediction in Koh Yor, Thailand	Input: temperature, pH, salinity Output: DO	GPR, SVM, LSR, and MNN

2. Materials and Methods
2.1 System Overview

The block diagram of the proposed framework for DO prediction and water quality evaluation is shown in Figure 1. Our proposed system is divided into three main parts, including data collection, prediction model evaluation, and prediction model validation, respectively.

The model evaluation is divided into four processes. First, the collected data were processed to remove missing values and outliers. Then the data set was divided into 80% training set and 20% testing set. After the data were segmented, the training data set was applied to train the regression models, including GPR, Medium Gaussian SVM, LSR, and MNN. Finally, the performance of the model was evaluated using the testing data set.

For the model validation part, six sets of data according to the village locations were applied. The regression model that provides the best performance (i.e., the highest accuracy) was tested with the zone datasets. At the same time, a selected model was evaluated.

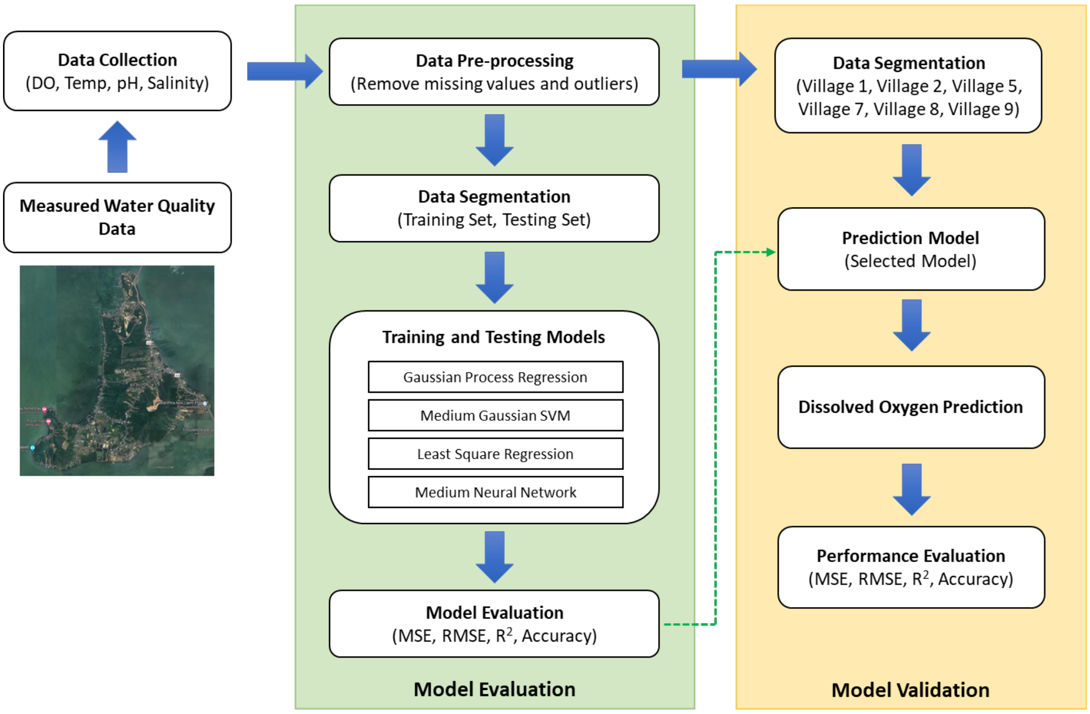


Fig. 1. The proposed framework for DO prediction and water quality evaluation

2.2 Data Collection

In this study, the water quality data were measured from the sea water around Koh Yor Island. The island is located at the lower Songkhla Lake in Muang District, Songkhla, Thailand, as shown in Figure 2. The water around Koh Yor has three characteristics, including fresh water (Songkhla Lake), brackish water, and salt water (sea), respectively. All data were obtained from the Coastal Aquaculture Research and Development Centre, Songkhla, Thailand.

Fig. 2. The study area: Koh Yor Island, Thailand

The water quality parameters used for the study consist of dissolved oxygen, or DO (i.e., target), temperature (temp), pH, and salinity. The data set of the water around the study area was measured from six villages, including Village 1, Village 2, Village 5, Village 7, Village 8, and Village 9. A total of 2256 (i.e., 564*4) samples were collected over a 4-year period (i.e., from 2019–2022). All data sets were pre-processed to remove missing values and outliers before being used for training and testing models. The measured values in the time series of the DO, the temperature, the pH, and the salinity are shown in Figure 3. Here, the values of temperature and salinity in the Koh Yor water depend on the season. In the summer, the temperature and the salinity are high, but the DO is relatively low. On the other hand, the temperature and the salinity in the rainy season are low, but the DO is high. Therefore, the value of DO has an inverse relationship with the temperature and the salinity. Table 2 also illustrates the water quality parameter data, where their minimum, maximum, mean, and standard values are provided.

Table 2
Water quality parameter data

Parameters	Minimum	Maximum	Mean	Standard deviation
DO (mg/L)	2.60	7.60	5.10	0.92
Temp (°C)	25.50	31.40	28.70	1.18
Salinity (ppt)	0.0	33.80	15.10	10.87
pH	7.00	8.50	7.80	0.30

These parameters are obtained from statistical analysis of raw water quality data, as shown in Figure 3.

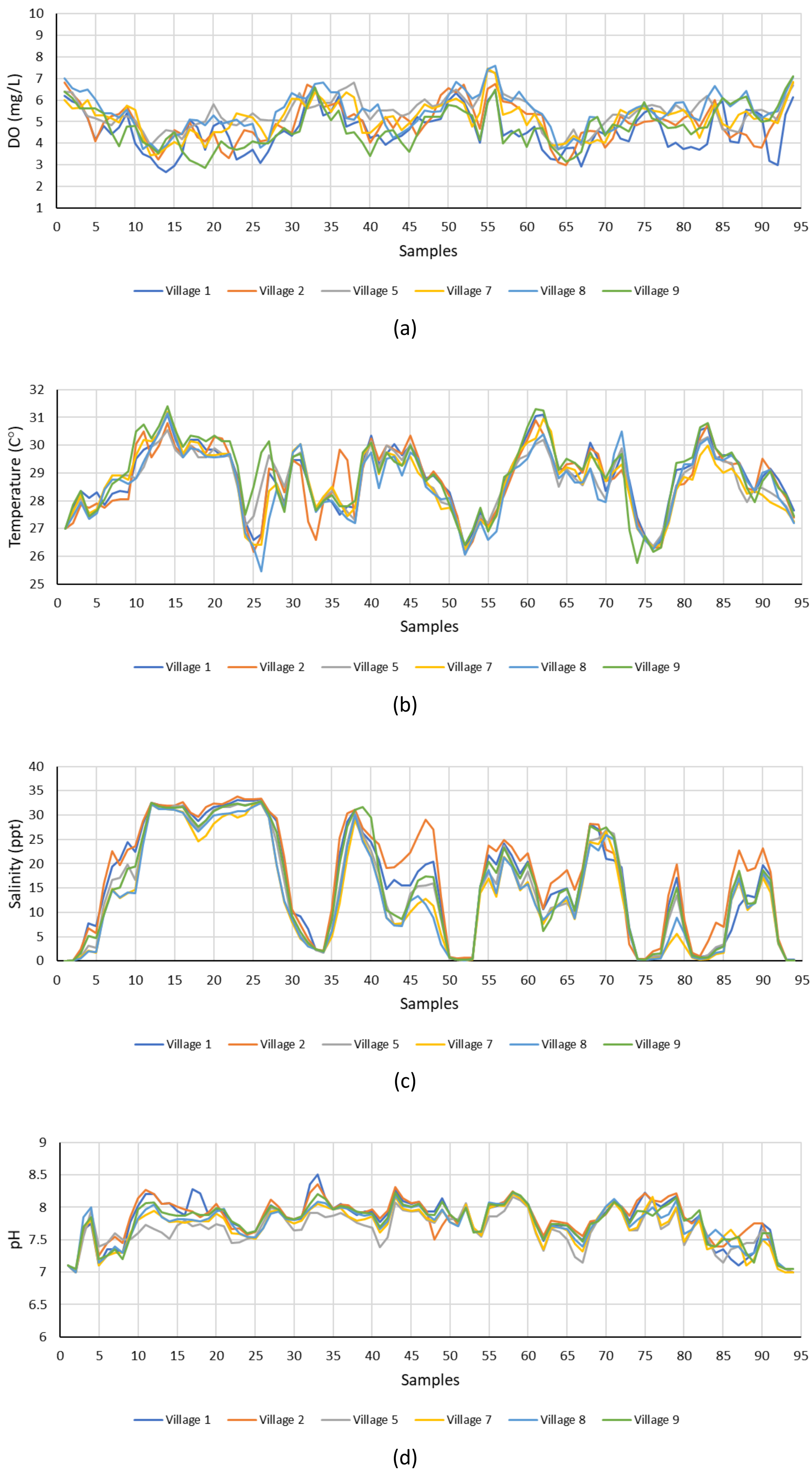


Fig. 3. The water quality data; (a) the dissolved oxygen, (b) the temperature, (c) the salinity, and (d) the pH

2.3 Prediction Models

As mentioned before, according to the literature survey, there are several models used to predict DO, for example, Neural Networks [9-14], Least Squares Support Vector Regression [15-17], Gaussian Process Regression [18], Support Vector Machine Regression [19], and Long Short-Term Memory Neural Network [20], etc. One method suitable for predicting DO in open systems with non-linear water quality data is a regression model. A regression model is a statistical technique used to analyse the relationship between one or more independent variables (predictors) and a dependent variable (the outcome or response variable). Regression models aim to understand and quantify the relationship between variables, allowing us to make predictions, draw insights, and perform hypothesis testing. A regression model produces an equation that represents the relationship between the independent and dependent variables, as shown in Eq. (1).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon \tag{1}$$

where Y is the dependent variable, $X_1, X_2, \ldots, X_n$ are independent variables, $\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ are coefficients of model, and ε represents the error term.

There are many types of regression models for predicting water quality. In this paper, we use the Regression Learner toolbox from MATLAB to find a suitable regression model for predicting DO at Koh Yor. According to the literature survey related to DO prediction, as described above, Gaussian Process Regression, Support Vector Machine, Least Squares Regression, and Neural Network are selected for testing in this work. Thus, a description of such models is summarized below.

2.3.1 Gaussian process regression

Gaussian Process Regression is a probabilistic approach to regression modelling that leverages Gaussian processes to capture the relationships between input and output data. It provides probabilistic predictions, allowing for a more nuanced understanding of uncertainty in predictions, and is particularly useful when dealing with limited data or when accurate uncertainty estimates are essential.

2.3.2 Medium Gaussian SVM

A Medium Gaussian Support Vector Machine (SVM) is a machine learning model for classification and regression tasks. It employs a Gaussian (RBF) kernel to map input data into a higher-dimensional space, enhancing its ability to capture complex relationships. This medium-sized SVM strikes a balance between computational efficiency and model expressiveness, making it suitable for moderately sized datasets. Training involves finding the optimal hyperparameters that define the decision boundary, with a focus on maximizing the margin between different classes. Its versatility and performance make it a common choice in various applications.

2.3.3 Least squares regression

Least Squares Regression, often referred to simply as linear regression, is a fundamental statistical and machine learning technique used for modelling the relationship between a dependent variable (also known as the response variable) and one or more independent variables (predictors or features). The primary goal of linear regression is to find the best-fitting linear equation that describes the relationship between these variables.

2.3.4 Medium neural network

A medium neural network is a computational model comprising layers of interconnected nodes (neurons) that process information. It falls between small and large networks in terms of size and complexity. Typically used for tasks like image recognition or natural language processing, it strikes a balance between computational efficiency and expressive power. Training involves adjusting the network's parameters through iterative optimization algorithms, such as gradient descent. The medium size allows for capturing intricate patterns in data without excessive computational demands.

2.4 Experiments

In this study, the experiments for regression model evaluation and DO prediction were divided into two cases. In the first experiment, the regression models, including GPR, Medium Gaussian SVM, LSR, and MNN, were assessed to find the best model that obtained the highest accuracy. The evaluation metrics were root mean square error (RMSE), mean square error (MSE), coefficient of determination, or R-squared (R^2), and accuracy. The water quality data *(*i.e., temperature, pH, and salinity*)* were processed before being used for training and testing the models. All data sets were divided into the training set (80%) and the testing set (20%) at a ratio of 4*:*1. The trained models were tested by the testing data set. Model training was using the Regression Learner App in MATLAB. The training parameters for each model were set to automatic and the model validation was set to 5-fold cross-validation. The performance results of each model were compared, and the highest performance model was selected for validation in the next experiment.

For the second experiment, the selected regression model that provided the highest efficiency was validated with the data set collected from each village. The aim of this experiment was to validate the regression model to predict the DO of each area. The validating data set was divided into six sets according to the villages around Koh Yor where sea basses are raised. The data set for each village was used to test the model. At the same time, the efficiency of the prediction was reported.

2.5 Performance Evaluation

As mentioned above, to evaluate the performance of the regression models, RMSE, MSE, R^2, and accuracy were selected as the performance evaluation metrics, as shown in Table 3. The variables DO_{mes}, DO_{pre}, and N were the measured DO, the predicted DO, and the prediction samples, respectively. The RMSE and the MSE are used to calculate the error of the data prediction compared to the observed values. The error values closer to 0 mean that the data has a low prediction error. In addition, the R-squared and the accuracy are used to calculate the predictive performance of the model. The R-squared and the accuracy values closer to 1 mean that the model has high predictive performance. Therefore, the regression model with high predictive performance should have a low prediction error and high R-squared and accuracy values.

Table 3
Performance evaluation metrics

Metrics	Formula	Description
Mean Squared Error (MSE)	$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(DO_{mes(i)} - DO_{pre(i)}\right)^2$	The MSE measures the average difference between predicted values and measured values. It is calculated as the mean of the squared differences between those values.
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(DO_{mes(i)} - DO_{pre(i)}\right)^2}$	The RMSE is the square root of the MSE. It is consistent with the scale of the measured values.
R-Squared (R^2)	$R^2 = 1 - \left(\frac{\sum_{i=1}^{N}\left(DO_{mes(i)} - DO_{pre(i)}\right)^2}{\sum_{i=1}^{N}\left(DO_{mes(i)} - \overline{DO}_{mes}\right)^2}\right)$	The R^2 evaluates the model's ability to explain the variance in the observed data. It is the ratio of the sum of squared differences between predicted and measured values to the total variance of the measured values.
Accuracy	$Accuracy = \frac{1}{N}\sum_{i=1}^{N}\left(1 - \frac{\left\|DO_{mes(i)} - DO_{pre(i)}\right\|}{DO_{mes(i)}}\right)$	Accuracy is a measure of how close the predicted value is to the measured value. It is a ratio of the difference between measured and predicted values divided by the measured value.

3. Results and Discussion

This section presents the experimental results for predicting DO using regression models. The performance comparison of the models to find the best-performing regression model and the performance evaluation results of the selected model for DO prediction in different areas (villages) are shown.

3.1 Performance Comparison of the Regression Models

The results of the first experiment to find the best performance regression model are demonstrated in this section. The training and testing results of all regression models are compared, as shown in Table 4 and Figure 4, where Figure 4 illustrates the test results. In addition, the correlation between the measured and predicted values of DO for each regression model is shown in Figure 5.

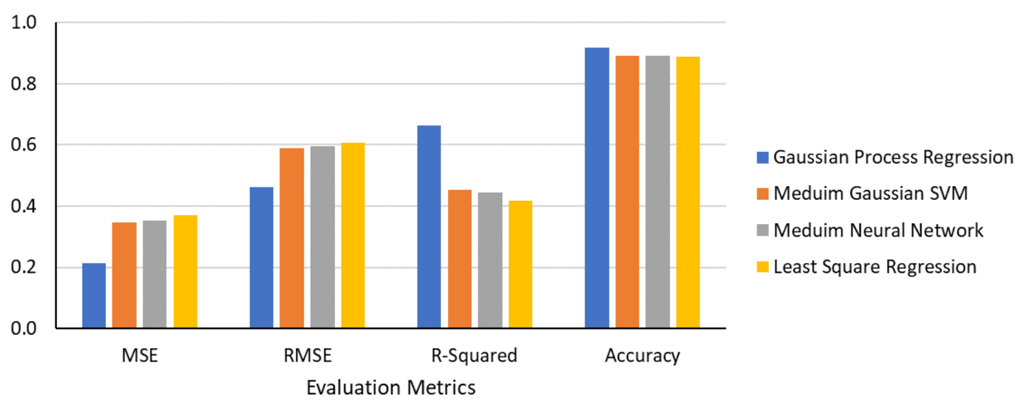


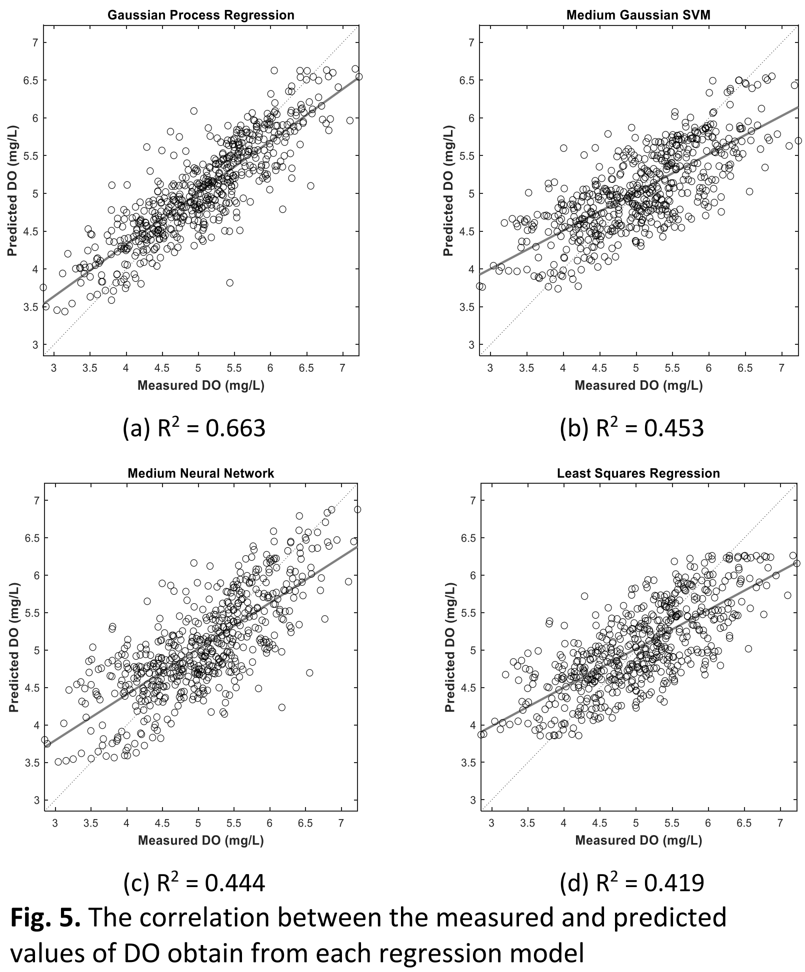
Fig. 4. Illustration of the performance of the regression models for testing

In Table 4, the experimental results demonstrate that the GPR model gives the lowest prediction error as measured by the MSE of (0.286, 0.214) and the RMSE of (0.534, 0.463) for training and testing results. The LSR model provides a high prediction error with an MSE of (0.342, 0.370) and an RMSE of (0.584, 0.608). The SVM model provides the lowest training time of 2.4218 seconds and the LSR provides the highest training time of 19.911 seconds. For the accuracy of the prediction models, the highest accuracy for DO prediction in the testing is from the GPR model, with an R-squared of 0.663 and an accuracy of 91.8*%*. The regression model has few errors, meaning the predicted DO is close to the measured DO.

Table 4
Performance comparison of the regression models

Models	Train				Test			
	MSE	RMSE	R-Squared	Training time (sec)	MSE	RMSE	R-Squared	Accuracy (%)
GPR	0.286	0.534	0.567	10.177	0.214	0.463	0.663	91.8
SVM	0.326	0.571	0.506	2.4218	0.348	0.590	0.453	89.1
MNN	0.352	0.593	0.467	14.262	0.354	0.595	0.444	89.2
LSR	0.342	0.584	0.483	19.911	0.370	0.608	0.419	88.9

The results in Figure 5 also show the correlation between the measured and predicted values of DO, where the GPR regression model provides good correlation results, as corresponding to an R^2 of 0.663 (also see Table 4).



(a) R^2 = 0.663
(b) R^2 = 0.453
(c) R^2 = 0.444
(d) R^2 = 0.419

Fig. 5. The correlation between the measured and predicted values of DO obtain from each regression model

3.2 Performance Evaluation of the Selected Model and DO Prediction for Different Areas

According to the results of the first experiment, it was found that the GPR was the most effective in predicting DO. Therefore, the second experiment uses this model to predict DO when six sets of water quality data (six villages) are applied. The aim of this experiment is to examine the accuracy of the GPR in predicting DO in different areas.

The performance evaluation results of the six data sets are shown in Table 5 and Figure 6. From Table 5, the results show that the village 7 area has an MSE of 0.118 and an RMSE of 0.343, the smallest error compared to other datasets. When considering the prediction results of all data sets, it was found that the prediction results of the village 7 area had an accuracy 95.1%. However, the results in Table 5 also indicate that the GPR regression model performs well in DO prediction for all villages, since the prediction accuracy is higher than 90% (i.e., min = 90.5% for village 1, max = 95.1% for village 7).

Table 5
Performance evaluation results of the GPR regression models in different areas

Areas	Performance evaluation			
	MSE	RMSE	R-Squared	Accuracy
Village 1	0.249	0.499	0.592	0.905
Village 2	0.154	0.392	0.748	0.939
Village 5	0.200	0.447	0.543	0.938
Village 7	0.118	0.343	0.750	0.951
Village 8	0.161	0.401	0.742	0.941
Village 9	0.152	0.390	0.775	0.931

An example graph of R-square results is shown in Figure 6, which consists of results from villages 1, 2, and 7, respectively. Village 1 has an R-square of 0.592 and an accuracy of 90.5%, which has the lowest prediction performance among the six villages. Considering the graph between the measured DO and the predicted DO of this village, there is a high error, resulting in the lowest R-square and accuracy. When considering the R-square graphs of village 7, which have an R-square of 0.750, Here, the results show that village 7 has the highest R-square and approaches 1, resulting in a maximum accuracy of 95.1% because the water quality parameters of Village 7 have low variation and this area is an area with low ecological change. The results presented in this section show that the GPR model can efficiently predict dissolved oxygen in different areas and has an average prediction accuracy of 93.4%.

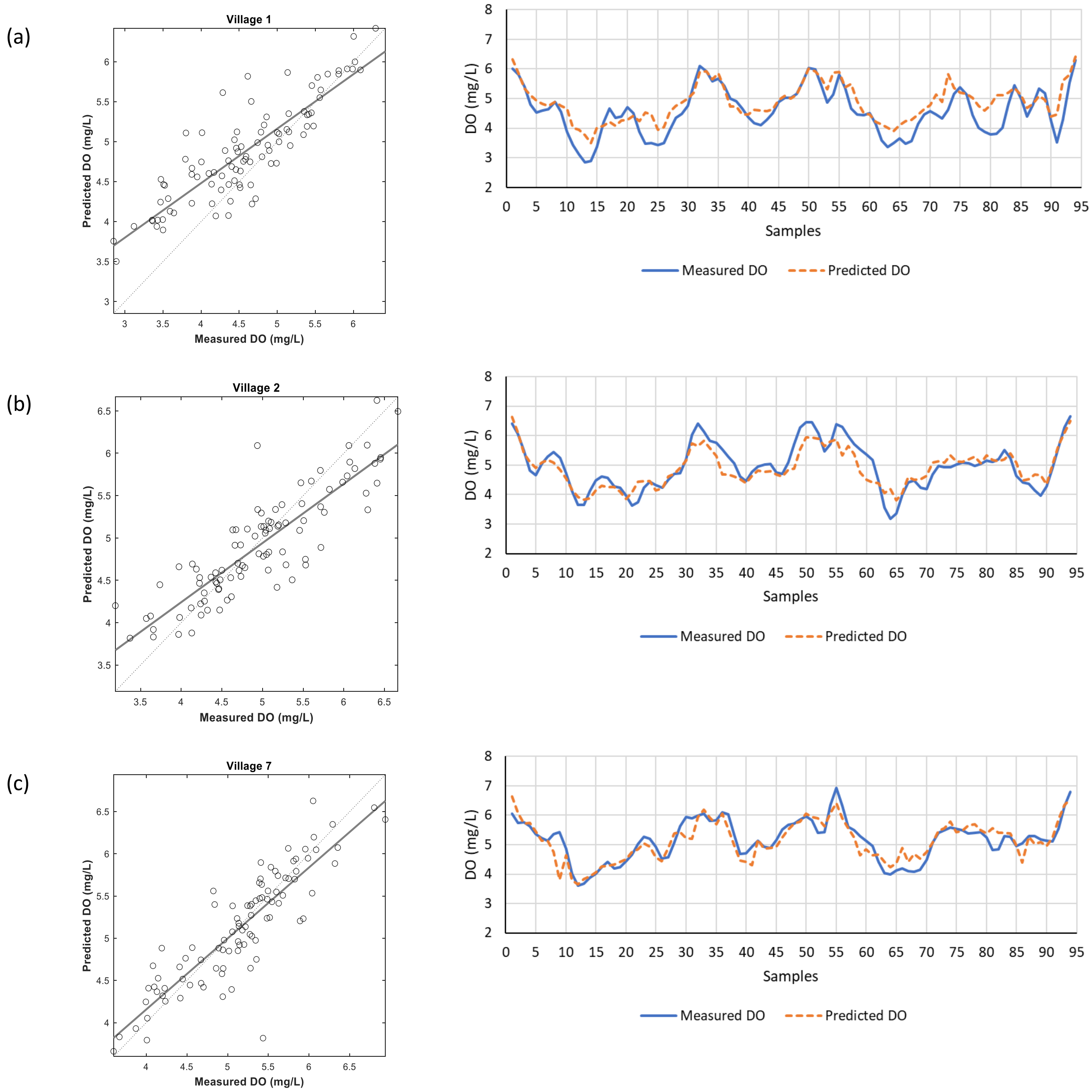


Fig. 6. Illustration of R^2 and DO prediction results using the GPR model for villages 1, 2, and 7, respectively

4. Conclusions

This work presents the system for dissolved oxygen prediction in Koh Yor water for breeding sea bass in cages. In this study, four regression models, namely Gaussian Process Regression (GPR), Medium Gaussian SVM (SVM), Least Squares Regression (LSR), and Medium Neural Network (MNN), have been applied for predicting dissolved oxygen. The models are evaluated using various performance criteria, including root mean square error (RMSE), mean square error (MSE), coefficient of determination, or R-squared (R2), and accuracy. Experimental results show that the best model with the highest prediction performance is the GPR model. During the training and testing phases, it has the highest prediction performance with an accuracy of 91.8% and R-square of 0.663, meaning that the predicted dissolved oxygen is close to the measured dissolved oxygen. The GPR model was then used to predict the DO when six sets of data from six village areas around Koh Yor were applied.

The results showed that the GPR model could effectively predict DO in all areas, with an accuracy greater than 90%. The overall results sufficiently demonstrated that the GPR model was suitable for DO prediction in Koh Yor water. In future work, the optimal regression model will be implemented to predict and monitor the DO for information to fishermen for breeding sea bass in cages in Koh Yor water, Thailand.

Acknowledgement
This work was supported by Faculty of Engineering, Academic year 2021, Prince of Songkhla University, Thailand. The authors thank the Coastal Aquatic Animal Technology and Innovation Research and Development Centre, Songkhla, Thailand, for providing the water quality data.

References

[1] Verma, A. K., and T. N. Singh. "Prediction of water quality from simple field parameters." *Environmental earth sciences* 69 (2013): 821-829. https://doi.org/10.1007/s12665-012-1967-6

[2] Lin, Weiwei, Jiajia Dai, Run Liu, Yuhong Zhai, Dingli Yue, and Qiansheng Hu. "Integrated assessment of health risk and climate effects of black carbon in the Pearl River Delta region, China." *Environmental research* 176 (2019): 108522. https://doi.org/10.1016/j.envres.2019.06.003

[3] Wu, Yingjun, Guiying Li, Yan Yang, and Taicheng An. "Pollution evaluation and health risk assessment of airborne toxic metals in both indoors and outdoors of the Pearl River Delta, China." *Environmental research* 179 (2019): 108793. https://doi.org/10.1016/j.envres.2019.108793

[4] Abba, S. I., M. S. Gaya, M. L. Yakubu, M. U. Zango, R. A. Abdulkadir, M. A. Saleh, A. N. Hamza, Ukashatu Abubakar, A. I. Tukur, and N. A. Wahab. "Modelling of uncertain system: a comparison study of linear and non-linear approaches." In *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, pp. 1-6. IEEE, 2019. https://doi.org/10.1109/I2CACIS.2019.8825085

[5] Pan, Guangbo, Youpeng Xu, Zhihui Yu, Song Song, and Yuan Zhang. "Analysis of river health variation under the background of urbanization based on entropy weight and matter-element model: A case study in Huzhou City in the Yangtze River Delta, China." *Environmental Research* 139 (2015): 31-35. https://doi.org/10.1016/j.envres.2015.02.013

[6] Wang, Dong, Dengfeng Liu, Hao Ding, Vijay P. Singh, Yuankun Wang, Xiankui Zeng, Jichun Wu, and Lachun Wang. "A cloud model-based approach for water quality assessment." *Environmental research* 148 (2016): 24-35. https://doi.org/10.1016/j.envres.2016.03.005

[7] Kisi, O., N. Akbari, M. Sanatipour, A. Hashemi, K. Teimourzadeh, and J. Shiri. "Modeling of dissolved oxygen in river water using artificial intelligence techniques." *Journal of Environmental Informatics* 22, no. 2 (2013): 92-101. https://doi.org/10.3808/jei.201300248

[8] Kisi, Ozgur, and Murat Ay. "Comparison of ANN and ANFIS techniques in modeling dissolved oxygen." In *Proceedings of the Sixteenth International Water Technology Conference (IWTC 16), Istanbul, Turkey*, pp. 7-10. 2012.

[9] Chen, Wei-Bo, and Wen-Cheng Liu. "Artificial neural network modeling of dissolved oxygen in reservoir." *Environmental Monitoring and Assessment* 186 (2014): 1203-1217. https://doi.org/10.1007/s10661-013-3450-6

[10] Wang, Lihao, Yu Jiang, and Hong Qi. "Marine dissolved oxygen prediction with tree tuned deep neural network." *IEEE Access* 8 (2020): 182431-182440. https://doi.org/10.1109/ACCESS.2020.3028863

[11] Keshtegar, Behrooz, and Salim Heddam. "Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: a comparative study." *Neural Computing and Applications* 30, no. 10 (2018): 2995-3006. https://doi.org/10.1007/s00521-017-2917-8

[12] Yang, Po-Yuan, Jyh-Homg Chou, Jinn-Tsong Tsai, Ying Ma, and Wen-Hsien Ho. "Prediction for dissolved oxygen in water of fish farm by using general regression neural network." In *2018 IEEE International Conference on Information and Automation (ICIA)*, pp. 1132-1135. IEEE, 2018. https://doi.org/10.1109/ICInfA.2018.8812400

[13] Heddam, Salim. "Simultaneous modelling and forecasting of hourly dissolved oxygen concentration (DO) using radial basis function neural network (RBFNN) based approach: a case study from the Klamath River, Oregon, USA." *Modeling Earth Systems and Environment* 2, no. 3 (2016): 135. https://doi.org/10.1007/s40808-016-0197-4

[14] Wang, Yu, and Huichao Dai. "The Application of the Oxygen Saturation Level Prediction in Dam Lower Reaches Based on the Principle Component RBF Network Model." In *2010 International Conference on E-Product E-Service and E-Entertainment*, pp. 1-6. IEEE, 2010. https://doi.org/10.1109/ICEEE.2010.5661525

[15] Amora, Dennis Jefferson A., Joesmart V. Apan, David Noel C. Layosa, Reimart C. Sarmiento, Jericho J. Tugado, and Christian Lian Paulo P. Rioflorido. "Soft Sensing Measurement of Dissolved Ammonia Nitrogen in Tank-Based Eel Aquaculture Systems Utilizing Deep Learning." In *2022 IET International Conference on Engineering Technologies and Applications (IET-ICETA)*, pp. 1-2. IEEE, 2022. https://doi.org/10.1109/IET-ICETA56553.2022.9971667

[16] Zhu, Chengyun, Xingqiao Liu, and Wangfang Ding. "Prediction model of dissolved oxygen based on FOA-LSSVR." In *2017 36th Chinese Control Conference (CCC)*, pp. 9819-9823. IEEE, 2017. https://doi.org/10.23919/ChiCC.2017.8028922

[17] Ngu, Joyce Chen Yen, and Wan Sieng Yeo. "Prediction of dissolved oxygen using least square support vector regression model." In *2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST)*, pp. 70-74. IEEE, 2022. https://doi.org/10.1109/GECOST55694.2022.10010638

[18] Yaseen, Zaher Mundher, Mohammad Ehteram, Ahmad Sharafati, Shamsuddin Shahid, Nadhir Al-Ansari, and Ahmed El-Shafie. "The integration of nature-inspired algorithms with least square support vector regression models: application to modeling river dissolved oxygen concentration." *Water* 10, no. 9 (2018): 1124. https://doi.org/10.3390/w10091124

[19] Adnan, Rana Muhammad, Xiaohui Yuan, Ozgur Kisi, and Rabia Anam. "Improving accuracy of river flow forecasting using LSSVR with gravitational search algorithm." *Advances in Meteorology* 2017, no. 1 (2017): 2391621. https://doi.org/10.1155/2017/2391621

[20] Lauguico, Sandy, Ronnie Concepcion, Rogelio Ruzcko Tobias, Jonnel Alejandrino, Dailyne Macasaet, and Elmer Dadios. "Indirect measurement of dissolved oxygen based on algae growth factors using machine learning models." In *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*, pp. 1-6. IEEE, 2020. https://doi.org/10.1109/R10-HTC49770.2020.9357014

[21] Liu, Shuangyin, Haijiang Tai, Qisheng Ding, Daoliang Li, Longqin Xu, and Yaoguang Wei. "A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction." *Mathematical and Computer Modelling* 58, no. 3-4 (2013): 458-465. https://doi.org/10.1016/j.mcm.2011.11.021

[22] Abba, Sani Isah, Nguyen Thi Thuy Linh, Jazuli Abdullahi, Shaban Ismael Albrka Ali, Quoc Bao Pham, Rabiu Aliyu Abdulkadir, Romulus Costache, and Duong Tran Anh. "Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration." *IEEE Access* 8 (2020): 157218-157237. https://doi.org/10.1109/ACCESS.2020.3017743

[23] Dheda, Dhruti, Ling Cheng, and Adnan M. Abu-Mahfouz. "Long short term memory water quality predictive model discrepancy mitigation through genetic algorithm optimisation and ensemble modeling." *IEEE Access* 10 (2022): 24638-24658. https://doi.org/10.1109/ACCESS.2022.3152818

[24] Al-Sulttani, Ali Omran, Mustafa Al-Mukhtar, Ali B. Roomi, Aitazaz Ahsan Farooque, Khaled Mohamed Khedher, and Zaher Mundher Yaseen. "Proposition of new ensemble data-intelligence models for surface water quality prediction." *IEEE Access* 9 (2021): 108527-108541. https://doi.org/10.1109/ACCESS.2021.3100490