How to cite: KAYA Y. Z., DEMİRCİ M., TAŞAR B., GÜZEL H. (2024) Prediction of the Dissolved Oxygen by Using Multi-Layer Perceptron and KNN Approaches: A Case Study in Coosa River, Alabama, USA. *2024 "Air and Water – Components of the Environment" Conference Proceedings*, Cluj-Napoca, Romania, p. 243-250, DOI: 10.24193/AWC2024_21.

PREDICTION OF THE DISSOLVED OXYGEN BY USING MULTI-LAYER PERCEPTRON AND KNN APPROACHES: A CASE STUDY IN COOSA RIVER, ALABAMA, USA

YUNUS ZİYA KAYA[1], MUSTAFA DEMİRCİ[2], BESTAMİ TAŞAR[2], HASAN GÜZEL[2]

DOI: 10.24193/AWC2024_21

ABSTRACT. Prediction of the Dissolved Oxygen by Using Multi-Layer Perceptron and KNN Approaches: A Case Study in Coosa River, Alabama, USA. The dissolved oxygen amount of a water body, such as a reservoir, stream, or river, is an important water quality parameter that may affect society's health directly. The daily mean dissolved oxygen of the Coosa River was investigated in this presented study. The multi-layer Perceptron (MLP) approach and k-nearest neighbor (KNN) algorithm, recently widely used for hydrological and environmental problems, was chosen for the prediction. Daily water temperature (Max, Min, and Mean), daily mean specific conductivity, daily median water pH, and discharge parameters were inputs in the MLP and KNN models. A total of 3535 daily records were implemented into the model. 2951 daily data were used as the training set, while 584 was the test set. Results were compared with each other by using statistical evaluation methods. The KNN approach was also generated by applying the same training and test sets. Based on the results, it is evident that the MLP (Multilayer Perceptron) model provided satisfactory dissolved oxygen prediction results. However, the KNN (K-Nearest Neighbors) model outperformed the MLP approach, despite having a lower correlation coefficient than the MLP.

Keywords: Water quality; MLP, KNN, Dissolved oxygen

1. INTRODUCTION

Dissolved oxygen is an important parameter for the determination of water quality. Monitoring the dissolved oxygen has a significant effect on the management of the water systems. Mohan & Kumar, 2016 defined dissolved oxygen as an indicator that gives an idea about the pollution of the water system. As it has great importance on society's health directly, it is hard to monitor the dissolved oxygen because of the economic conditions of the sensors used for the measurements (Zhu et al., 2021). Due to the expensiveness and economic life of the measurement sensors, researchers recently used different machine-learning approaches to determine the dissolved oxygen in rivers, reservoirs, wetlands, etc.

[1] Osmaniye Korkut Ata University, Engineering Faculty, Civil Eng. Department, yunuszkaya@osmaniye.edu.tr

[2] Iskenderun Technical University, Eng. Faculty, Civil Engineering Department, mustafa.demirci@iste.edu.tr, bestami.tasar@iste.edu.tr, hasan.guzel@iste.edu.tr

Xiao, 2017 used the neural network approach to predict the dissolved oxygen. The area chosen for the mentioned study was Beihai, Guangxi, a traditional aquaculture base in southern China. They compared the model results with the support vector machines, auto-regressive model, curve fitting, and grey model approaches. As a result, they have obtained that the backpropagation neural network approach has better performance in predicting the dissolved oxygen. Huang, 2021 underlined the importance of dissolved oxygen for growing healthy fish in intensive aquaculture ponds. To improve the prediction precision of the dissolved oxygen, they have proposed a gated recurrent unit (GRU) with a generalized opposition-based learning particle swarm optimization algorithm (GOBLPSO) and a combination of complete ensemble empirical mode decomposition with an adaptive noise Lempel-Ziv complex (CEEMDAN-LZC) approaches. Kisi et al., 2021 predicted hourly dissolved oxygen by using Bayesian model averaging approach. They compared the models with some of the soft computing approaches. They considered hourly temperature, pH, and specific conductivity parameters as inputs in the created models. According to the results, they underlined that the proposed Bayesian model averaging approach has better performance on the prediction of dissolved oxygen. Basant et al., 2010 used linear and nonlinear approaches to predict the dissolved oxygen and biochemical oxygen demand of the surface water. More studies about the implications of the soft computing approaches on the prediction of dissolved oxygen can be found in Yang, 2023; Li et al., 2021; Xu et al., 2021; Šiljić et al.,2018.

As it is known, the applications of artificial intelligence in environmental and civil engineering are not limited to the prediction of dissolved oxygen. There have been many studies applied to different types of hydrological or environmental problems recently. Some of the prediction approaches recently performed on civil and environmental engineering can be found in (Kasiviswanathan, K. S., & Sudheer, K. P., 2017; Unes, F. & Demirci, M., 2015; Altay et al.,2023; Noori, N., & Kalin, L., 2016; Karatas, I., & Budak, A. 2021; Khazaee Poul, 2019; Unes at al. 2020)

In this study, two of the most tested methods in civil and environmental engineering, namely Multi-Layer Perceptron (MLP) and K-nearest neighbor (KNN), were used to predict the daily dissolved oxygen of the Coosa River, Alabama, USA. Since both approaches were found successful in the dissolved oxygen prediction, it is seen that the KNN approach gave more satisfactory results on prediction.

2. METHODOLOGY

In this study, the dissolved oxygen of the Coosa River, Alabama, USA, was investigated by using the MLP and KNN approaches. The data was downloaded from the United States Geological Survey's website (USGS). The data consists of a total of 3535 daily data. The data sets were divided into training and test sets. The last 583 daily data was taken as the test set. The data starts from 1.1. 2013 and it ends in 1.1.2023. Detailed information about the selected parameters is shared in Table 1.

Table1. ***Statistical information about the data sets***

Parameter/ Statistic	Training Set				Unit	Test Set			
	Min	Max	Mean	Std		Min	Max	Mean	Std
Water Temperature (Min)	3.8	32.9	17.99	7.3	C^0	5.9	30.2	19.12	6.58
Water Temperature (Max)	4.3	36.5	19.23	7.85	C^0	6.3	32.6	20.41	6.98
Water Temperature (Mean)	4.1	34.1	18.5	7.48	C^0	6.1	31.1	19.67	6.71
Specific Conductivity	51	355	150.4	44.74	(µS/cm at 25 °C)	77	218	143.2	30.09
PH	6.6.	9.3	7.59	0.38	-	7	8.6	7.57	0.29
Discharge	1	998	15.74	80.99	$(ft^3/s)x10^3$	1.44	27	6.87	5.09
Dissolved Oxygen	5	16.4	9.05	1.5	Mg/L	4.7	12	8.69	1.43

The data was prepared by eliminating the estimated daily values and the daily missing records. No data estimations were made to complete the missing daily data.

2.1 Study area

The monitoring location of the Coosa River is associated with a Stream in Cherokee County, Alabama. The monitoring station is a U.S. Geological Survey agency. The location map, downloaded from the USGS website, is shared in Figure 1.

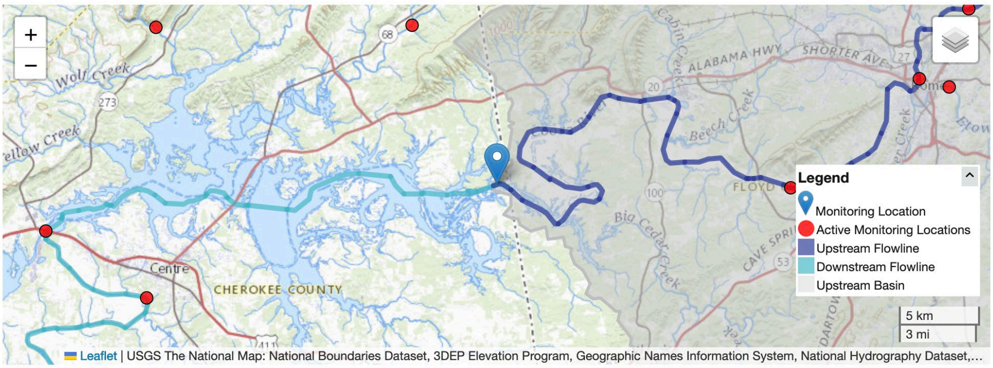


Figure 1. ***The location of the monitoring station (USGS, 2023)***

The name of the site was Coosa River at State Line, AL/GA. The decimal latitude of the location is 34.20194444, and the decimal longitude is -85.45325.

2.2 Multi-Layer Perceptron (MLP)

A multi-layer perceptron is a type of neural network with a wide range of usage areas. In this type of neural network, the signals have one direction which is from input(s) to outputs. The general structure of the MLP contains an input layer, a hidden layer, and an output layer. In this study, the MLP structure was automatically built up using WEKA software. The WEKA software lets users build the model structure by hand or set it up using a heuristic. For further information about the MLP, readers can check Almeida, 1997. The structure of the model created for this study is given in Figure 2.

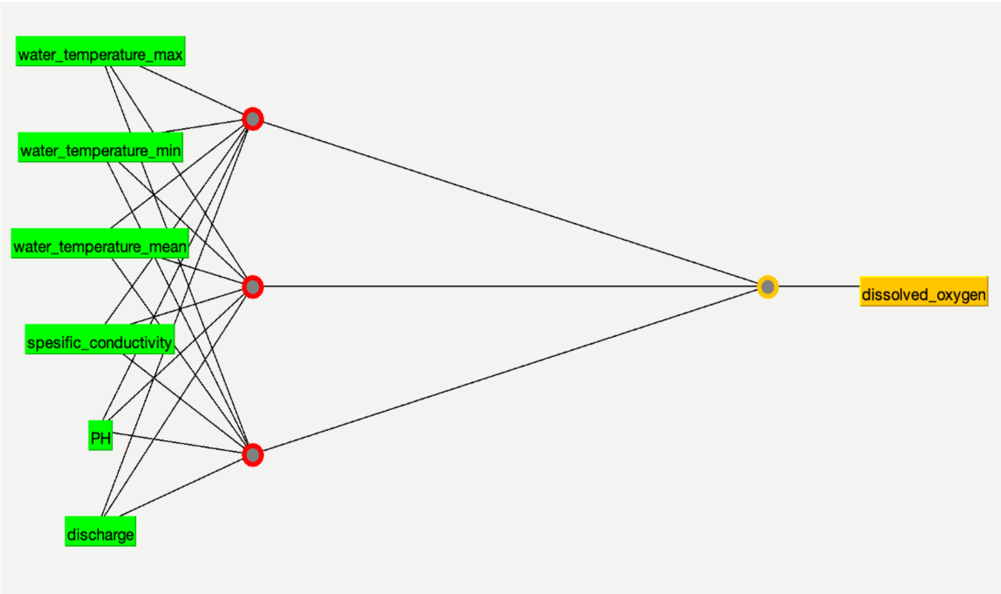


Figure 2. ***Structure of the created MLP model for the prediction of MLP***

2.3 K nearest neighbor (KNN)

The k-nearest neighbor approach is a machine-learning approach that is considered a lazy method. The KNN approach is used for prediction based on the determination of the number of nearest k, computation of the distances between training values and the output, grouping the training records, and assigning a prediction value for the output. More information about the KNN algorithm can be found in D. Aha & D. Kibler, 1991.

3. RESULTS AND DISCUSSIONS

In this study, daily water temperature (max, min, mean), specific conductance, PH, and discharge values of the Coosa River were added to the MLP and KNN approaches to predict the daily dissolved oxygen. The correlation coefficients of each input parameter against the output (Dissolved Oxygen) parameter are given in Table 2.

Table 2. ***The correlation matrix of the used parameters***

	Water Temperature max	*Water Temperature min*	*Water Temperaure Mean*	*Spesific Conductance*	*PH*	*Discharge (10^3)*	*Dissolved oxygen*
Water Temperature max	1,000						
Water Temperature min	0,996	1,000					
Water Temperaure Mean	0,998	0,999	1,000				
Spesific Conductance	0,454	0,441	0,446	1,000			
PH	0,452	0,420	0,434	0,605	1,000		
Discharge (10^3)	-0,031	-0,031	-0,031	0,157	0,040	1,000	
Dissolved oxygen	-0,471	-0,508	-0,494	0,027	0,433	0,022	1,000

When Table 2 is examined, it will be seen that there is a negative correlation between water temperature and dissolved oxygen. The most highly positively correlated parameter with dissolved oxygen is calculated as PH. The correlation between the discharge and dissolved oxygen can be neglected. Meanwhile, there is no negative or positive correlation between the discharge and dissolved oxygen. The correlation between the specific conductance and the dissolved oxygen is almost the same as the relation between dissolved oxygen and the discharge.

The MLP distribution graph and scatter chart are shared in Figure 3. As mentioned in the methodology section, the MLP model was created by using water temperature (min, max, mean), specific conductance, PH, and discharge parameters. All input parameters are in daily format.

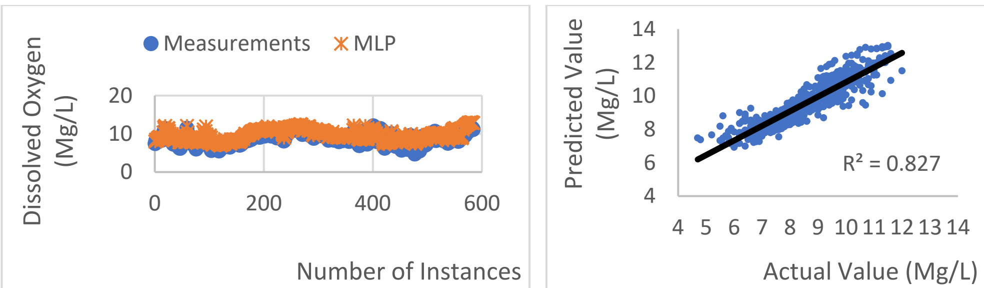
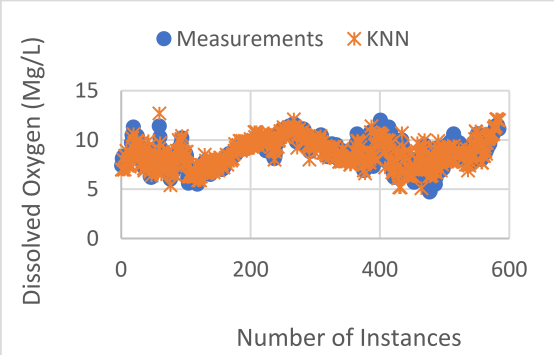


Figure 3. ***The MLP model results***

When Figure 3 is examined, it can be seen that the MLP results and measurement values largely overlap. The determination coefficient was calculated as 0.83 for the MLP. Mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE), and root relative squared error (RRSE) statistics were calculated as 1.02, 1.16, 84.12%, and 78.89%, respectively.

The KNN results are shared in Figure 4. The KNN model was created by using the same inputs as the MLP model.



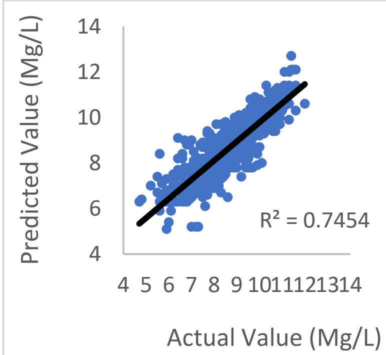


Figure 4. ***The KNN model results***

As can be seen in Figure 4 determination coefficient was calculated for KNN models as 0.75. This value is lower than the determination coefficient of the MLP. However, the error statistics of the KNN model approach have much lower values than the MLP. Comparative statistics are shared in Table 3.

Table 3. ***Statistical evaluation of the MLP and KNN models***

Statistics	MLP	KNN
Correlation coefficient	**0,91**	0,86
MAE	1,02	**0,57**
RMSE	1,16	**0,74**
RAE	84,12	**46,87**
RRSE	78,89	**50,11**
Total Number of Instances	584,00	584,00

In Figure 4, the relationship between the KNN results and the measured dissolved oxygen is displayed. The distribution graph of the KNN model indicates that it accurately captures the measured values. Although the determination coefficient was found to be lower than that of the MLP model, all calculated error statistics demonstrate that the KNN model has lower errors in predicting daily dissolved oxygen.

4. CONCLUSION

In this paper, the authors focused on predicting the daily dissolved oxygen values of Coosa River, AL, US. Two of the most popular soft computing approaches, namely Multi-Layer Perceptron and k-nearest neighbor, were selected for prediction of the investigated parameter. Results were compared with each other by using various statistical error calculations and correlation coefficients. The relation of the parameters was also investigated by calculating the correlation matrix of all selected parameters. The correlation matrix showed that there is a negative correlation between water temperature and dissolved oxygen. No correlation was detected between the specific conductance and the discharge. Only, A positive correlation between PH and dissolved oxygen was determined based on the correlation calculations.

According to the model results, MLP has acceptable results in the case of dissolved oxygen by given parameters. However, the results of the MLP have serious error statistics. To minimize these errors, the number of epochs was increased during the modeling process, but the errors did not decrease more than presented in the results and discussion section. KNN results were found to be more reliable in this case study due to the lower error calculations. In conclusion, both methods were found useful in the case of dissolved oxygen prediction of the Coosa River. Model performances may be increased by using different input parameters. The reliability of the models also needs to be tested under different conditions for various types of water bodies. The performance of the models can also be tested by using various combinations of inputs. Similar results may be obtained even by using fewer input parameters.

ACKNOWLEDGEMENT

We greatly appreciate the U.S Geological Survey for their invaluable contribution in collecting and disseminating critical water quality data over an extended period. Their unwavering commitment to this cause has undoubtedly helped in enhancing our understanding of water quality and improving our ability to address environmental challenges.

REFERENCES

1. Almeida, L.B. Multilayer perceptrons, in Handbook of Neural Computation, IOP Publishing Ltd and Oxford University Press, 1997.
2. Altay, G., Kayadelen, C. & Kara, M. Model selection for prediction of strong ground motion peaks in Türkiye. Nat Hazards (2023). https://doi.org/10.1007/s11069-023-06252-y
3. Basant, N., Gupta, S., Malik, A., & Singh, K. P. (2010). Linear and nonlinear modeling for simultaneous prediction of dissolved oxygen and biochemical

oxygen demand of the surface water—a case study. *Chemometrics and Intelligent Laboratory Systems*, *104*(2), 172-180.

4. D. Aha, D. Kibler (1991). Instance-based learning algorithms. Machine Learning. 6:37-66.
5. Huang, J., Liu, S., Hassan, S. G., Xu, L., & Huang, C. (2021). A hybrid model for short-term dissolved oxygen content prediction. *Computers and Electronics in Agriculture*, *186*, 106216. https://doi.org/10.1016/j.compag.2021.106216
6. Karatas, I., & BUDAK, A. (2021). Prediction of Labor Activity Recognition in Construction with Machine Learning Algorithms . Icontech International Journal, 5(3), 38–47. https://doi.org/10.46291/ICONTECHvol5iss3pp38-47
7. Kasiviswanathan, K. S., & Sudheer, K. P. (2017). Methods used for quantifying the prediction uncertainty of artificial neural network based hydrologic models. *Stochastic environmental research and risk assessment*, *31*, 1659-1670.
8. Khazaee Poul, A., Shourian, M., & Ebrahimi, H. (2019). A comparative study of MLR, KNN, ANN and ANFIS models with wavelet transform in monthly stream flow prediction. *Water Resources Management*, *33*, 2907-2923.
9. Kisi, O., Alizamir, M., & Docheshmeh Gorgij, A. (2020). Dissolved oxygen prediction using a new ensemble method. *Environmental Science and Pollution Research*, *27*, 9589-9603.
10. Li, W., Wu, H., Zhu, N., Jiang, Y., Tan, J., & Guo, Y. (2021). Prediction of dissolved oxygen in a fishery pond based on gated recurrent unit (GRU). *Information Processing in Agriculture*, *8*(1), 185-193.
11. Mohan, S., & Kumar, K. P. (2016). Waste load allocation using machine scheduling: model application. *Environmental Processes*, *3*, 139-151.
12. Noori, N., & Kalin, L. (2016). Coupling SWAT and ANN models for enhanced daily streamflow prediction. *Journal of Hydrology*, *533*, 141-151.
13. Šiljić Tomić, A., Antanasijević, D., Ristić, M., Perić-Grujić, A., & Pocajt, V. (2018). Application of experimental design for the optimization of artificial neural network-based water quality model: a case study of dissolved oxygen prediction. *Environmental Science and Pollution Research*, *25*, 9360-9370.
14. Üneş, F., Demirci, M., Zelenakova, M., Çalışıcı, M., Taşar, B., Vranay, F., & Kaya, Y. Z. (2020). River flow estimation using artificial intelligence and fuzzy techniques. *Water*, *12*(9), 2427.
15. Üneş, F., & Demirci, M. (2015). Generalized regression neural networks for reservoir level modeling. International Journal of Advanced Computational Engineering and Networking, 3, 81-84.
16. USGS, United States Geological Survey, 2023. URL: https://waterdata.usgs.gov/monitoring-location/02397530/#parameterCode=00065&period=P7D&showMedian=false. Access date: 20.12.2023
17. Xiao, Z., Peng, L., Chen, Y., Liu, H., Wang, J., & Nie, Y. (2017). The dissolved oxygen prediction method based on neural network. *Complexity*, *2017*.
18. Xu, C., Chen, X., & Zhang, L. (2021). Predicting river dissolved oxygen time series based on stand-alone models and hybrid wavelet-based models. *Journal of Environmental Management*, *295*, 113085.
19. Yang, J. (2023). Predicting water quality through daily concentration of dissolved oxygen using improved artificial intelligence. *Scientific Reports*, *13*(1), 20370.
20. Zhu, N., Ji, X., Tan, J., Jiang, Y., & Guo, Y. (2021). Prediction of dissolved oxygen concentration in aquatic systems based on transfer learning. *Computers and Electronics in Agriculture*, *180*, 105888.