Article

New Graph-Based and Transformer Deep Learning Models for River Dissolved Oxygen Forecasting

Paulo Alexandre Costa Rocha [1,2,*] , Victor Oliveira Santos [1], Jesse Van Griensven Thé [1,3] and Bahram Gharabaghi [1,*]

1 School of Engineering, University of Guelph, 50 Stone Rd E, Guelph, ON N1G 2W1, Canada; volive04@uoguelph.ca (V.O.S.); jessethe@gmail.com (J.V.G.T.)
2 Mechanical Engineering Department, Technology Center, Federal University of Ceará, Fortaleza 60020-181, CE, Brazil
3 Lakes Environmental Research Inc., 170 Columbia St W, Waterloo, ON N2L 3L3, Canada
* Correspondence: pcostaro@uoguelph.ca (P.A.C.R.); bgharaba@uoguelph.ca (B.G.)

Abstract: Dissolved oxygen (DO) is a key indicator of water quality and the health of an aquatic ecosystem. Aspiring to reach a more accurate forecasting approach for DO levels of natural streams, the present work proposes new graph-based and transformer-based deep learning models. The models were trained and validated using a network of real-time hydrometric and water quality monitoring stations for the Credit River Watershed, Ontario, Canada, and the results were compared with both benchmarking and state-of-the-art approaches. The proposed new Graph Neural Network Sample and Aggregate (GNN-SAGE) model was the best-performing approach, reaching coefficient of determination (R^2) and root mean squared error (RMSE) values of 97% and 0.34 mg/L, respectively, when compared with benchmarking models. The findings from the Shapley additive explanations (SHAP) indicated that the GNN-SAGE benefited from spatiotemporal information from the surrounding stations, improving the model's results. Furthermore, temperature has been found to be a major input attribute for determining future DO levels. The results established that the proposed GNN-SAGE model outperforms the accuracy of existing models for DO forecasting, with great potential for real-time water quality management in urban watersheds.

1. Introduction

In the Anthropocene Era [1], the increasing urbanization of natural environments close to rivers and lakes has negatively affected the health of the aquatic ecosystems [2,3]. Effluents, specifically urban stormwater runoff and combined sewer outflows, serve as significant sources of pollutants that consume the oxygen in the water through chemical and biological oxygen demands, disrupting the physical, chemical, and biological integrity of the aquatic ecosystem [4,5]. Monitoring the river water quality is important to keep track of the river's health condition, helping develop management strategies aimed at the river's preservation and conservation [6]. One major indicator of the river's water quality is the dissolved oxygen (DO) concentration. The DO is highly variable in both time and space within the stream network, under the constant influence of both groundwater discharge and surface runoff to the stream [7,8]. Fluctuations on DO concentrations can be related to adverse thermal impacts of stormwater runoff and eutrophication due to excessive growth of aquatic plants, impairing aquatic habitat [9–12].

Given the major role of DO in maintaining the health of aquatic ecosystems, it is crucial to forecast and monitor its level with high accuracy [13]. The DO levels can be evaluated by different methods, such as chemical analysis, and by the employment of

sensors. While accurate, chemical analysis, like iodometric titration, is difficult to perform and best suited for lab use. Polarography, an electrochemical-based sensor approach, is widely used but requires regular maintenance and cannot continuously measure oxygen. The fluorescence method, a sensor based on the optical properties of DO, offers quick and sensitive measurements without oxygen depletion. Current studies aim to enhance materials in electrochemical and optical sensors. Intelligent sensors now offer real-time, accurate oxygen measurements through smart signal processing [14,15]. Accurate identification of DO levels helps stakeholders' decision-making and managerial actions, mitigating the impacts of anthropogenic changes on the aquatic environment and aiding the development of strategies to prevent further degradation of the environment.

In this context, machine learning, a subfield of artificial intelligence, has emerged as a powerful tool for enhancing DO monitoring and prediction, providing insight into the factors underlying future DO levels. This paradigm makes use of mathematical tools like probability, optimization, linear algebra, vector calculus, analytic geometry, and matrix decomposition, enabling the analysis of large quantities of information, as well as learning intricate relationships in the data [16]. The principle of machine learning is to create an algorithm that can identify the statistical structures that determine the rules of a dataset through the results provided to it by training rather than being explicitly programmed. In recent years, thanks to the progress in hardware development [17,18], this approach has drawn attention from researchers due to its improved computational performance and generalization capabilities when compared with numerical models. It can also identify complex non-linear associations in the dataset without predefined algorithms to specifically perform this task [19,20]. Despite their attested good performance, ML models may suffer from low data quality, availability, and overfitting. They are also a black-box approach, which hinders the interpretability of their results.

When used in time-series forecasting applications, the ML paradigm reached excellent results in different knowledge fields [21–24], also showing improvements when using past information, i.e., time lags [25–27]. Extensive literature can be found for hydrological studies using ML tools, including studies of water quality and DO. A plethora of studies have shown that different ML and hybrid ML models can forecast DO levels in distinct locations [28–30]. In the context of urban rivers, the ML paradigm also proved to be a valuable asset in DO estimation [6,31–33].

Deep learning (DL) models represent an advanced approach that addresses the limitations of traditional methods in predicting future DO concentrations. Dissolved oxygen estimation is a spatiotemporal problem where recurrent neural networks and their variations are popular among researchers in this field. In [34], the authors used remote sensing data to estimate future DO concentrations using prior DO information in the Rawal watershed in Pakistan. The best result for future DO estimation was returned by the bi-directional long short-term memory (Bi-LSTM) paradigm, with an RMSE equal to 0.199 mg/L. Hybrid configuration merging the LSTM and the convolutional neural network (CNN) paradigm was evaluated in [35] to forecast short-term DO in Small Prespa Lake, Greece. Compared with the standalone models and conventional ML paradigms such as decision trees (DTs) and support vector machine (SVM), the hybrid configuration surpassed all of them, reaching an RMSE of 0.518 mg/L. Other application using CEEMDAN together with DL models was proposed in [36]. In their paper, the authors compared several DL models with real-time forecasting of DO concentrations in Xin'anjiang River, China. They compared the standalone DL approaches CNN, LSTM, the hybrid CNN–LSTM, and combinations of these models with the CEEMDAN pre-processing. The authors discovered that the CEEMDAN–CNN–LSTM hybrid configuration was able to provide the best outcomes for different multi-step ahead forecasting for DO, with the best RMSE value of 0.26 mg/L for 4 h ahead forecasting. Application of DL models in an urban river was investigated in [37]. In their work, the authors implemented a recurrent neural network to forecast DO in Fanno Creek in Oregon, USA, achieving excellent results for DO estimation for 1, 3, and 7 days ahead. In recent years, graph neural network (GNN) models have been implemented to time-series

forecasting within the DL field, reaching cutting-edge results [25,26,38]. However, little is known about implementing such a paradigm in the water quality research area, with few works addressing this subject [39–42].

As the present bibliographic review elucidates, ML applications in hydrological and environmental situations constitute a vast field of study. Aiming to deepen the understanding of ML, specifically addressing the graph-based approaches, the present work aims to contribute to filling the knowledge gap regarding these models when applied to DO forecasting. To this end, historical data comprising the years 2016 to 2020 for Credit River, Ontario, were collected and implemented to the Deep Neural Network Transformer (DNN-Transformer) and the Graph Neural Network Sample and Aggregate (GNN-SAGE) to determine future levels of DO on the river, a highly non-conservative substance. The present study aims to contribute to the field by

1. Developing a cutting-edge model to predict DO concentration with elevated precision and accuracy.
2. Verifying the temporal effect over the estimated DO concentration, a highly non-conservative substance, by implementing different time lags for different predictive models, namely XGBoost, DNN-Transformer, and the proposed GNN-SAGE.
3. Conducting a Shapley additive explanation (SHAP) analysis to assess the significance of different input variables, allowing meaningful inputs over the models' forecasting and its functioning.
4. Enabling the development of a water quality forecasting system for urban rivers, aiding the elaboration of risk management strategies and environmental policies.

2. Case Study

The Credit River is a major river in Southern Ontario, Canada, part of the Toronto Greater Area [43] (Figure 1). Starting at Orangeville, the Credit River stretches for 90 km before finally reaching Lake Ontario [44,45]. The river has a mainly rural configuration in its upper and middle areas, having significant forest coverage [46]. As it approaches Lake Ontario, the urbanization of the river's area increases [6]. Urban development in the river area, however, endangers aquatic life and the people who rely on the river, increasing the risk of floods [44,47–49]. Table 1 compiles the Credit River's watershed characteristics. Figure 1 shows the location of the studied site in Ontario province.
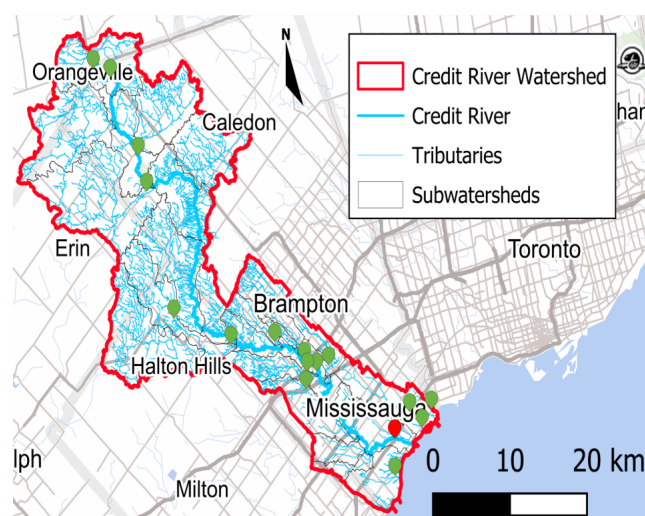

Figure 1. Location of the Credit River in Ontario Province. The points showcase the position of the measurement stations along the river's course. The red point in the map represents the study station where the dissolved oxygen is predicted, namely "Credit River @ MGCC", i.e., Mississauga Golf and Country Club. The green points are the used neighboring monitoring stations, which provide spatiotemporal information to the forecasting model.

**Table 1.** Characteristics of the Credit River watershed [43–46,50].

Credit River Watershed Characteristics	
Drainage area	93,000 ha
Credit River length	90 km
Altitude	190–521 m
Area used for agriculture	35%
Area used for urban settlement	27%
Area of natural habitats	38%
Estimated population within the watershed area	1 million people

This study developed the forecasting models using historical data from 2016 to 2020. The data were made available by the Credit Valley Conservation Organization (data available at https://cvc.ca/real-timemonitoring/, accessed on 1 November 2023). The monitored data include weather attributes and physical–chemical information collected from stations throughout the river's watershed (Figure 1). The dataset, originally with a temporal resolution of 15 min, was resampled to 1 h intervals using the moving average technique. This adjustment involved calculating the average value for each variable, except for precipitation, which was summed. These processed data were then used as input for the models under evaluation [6,50]. Table 2 shows the statistical description for DO data for all the monitoring stations and the reference station alone, in mg/L. Figure 2 shows the measured DO levels at the reference station "Credit River @ MGCC", for different time windows. After that, the data correlogram is presented in Figure 3, where the target attribute is the dissolved oxygen (first row and first column).

Table 2. Statistical description for DO levels recorded by all the monitoring stations and the reference station. The values are in mg/L.

|	All Monitoring Stations	Study Station Credit River @ MGCC
Minimum	0.9	5.6
Maximum	16.7	16.7
Mean	10.3	10.8
Standard deviation	2.2	2.2
25% Quantile	8.9	8.9
50% Quantile (median)	10.4	11.0
75% Quantile	12.0	12.7

In Figure 2, the variance in the DO levels is not visibly meaningful. Panel (a) shows yearly variation in DO for the whole dataset, elucidating that the measured levels' peaks and valleys are well comprised within the maximum and minimum values described in Table 2. Figure 2b elucidates that, for an annual measurement of DO, the temperature effect dominates the behavior of the oxygen levels in the watershed, decreasing during the hotter months of the year and returning to increase as the climate cools down. In Figure 2c, we observe that the temperature still dominates the behavior of DO oscillations, albeit in a time-scale variation ranging from 5 to 7 days. For the daily variance of DO levels presented in Figure 2d, the solar irradiance now dominates the variations on the dissolved oxygen. The data presented in Figure 2 allow us to conclude that the reference station complies with the Canadian Council of Ministers of the Environment (CCME) for DO levels in the freshwater for conservation of the aquatic life [51], indicating that the efforts on maintaining the aquatic ecosystem are yielding positive results.
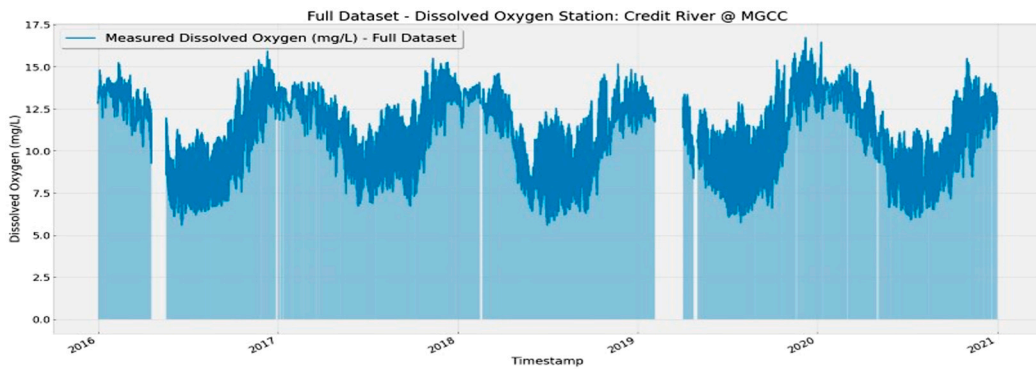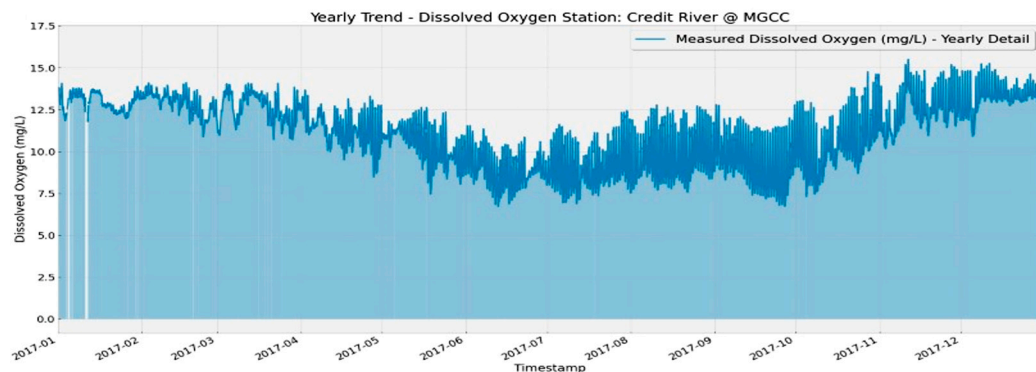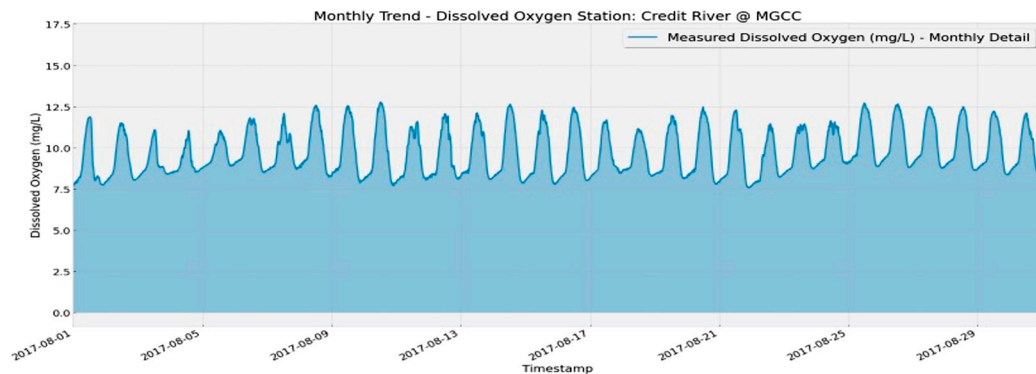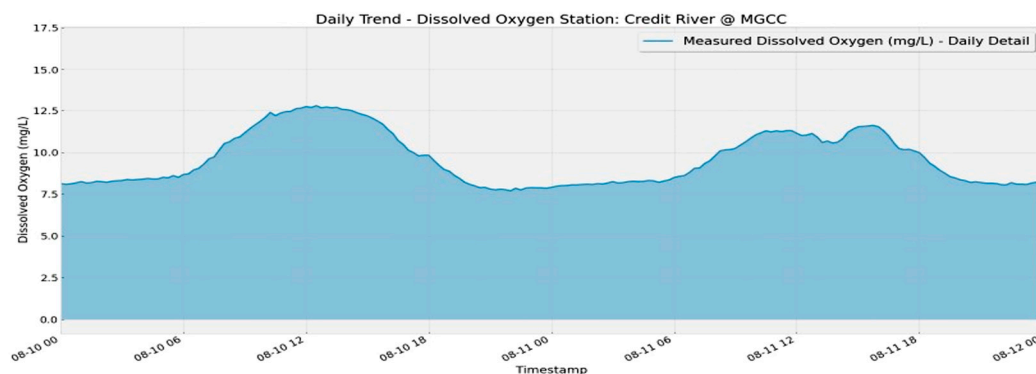
Figure 2. Dissolved oxygen measurement for (**a**) the whole dataset, (**b**) one year, (**c**) one month, and (**d**) two sequential days at the reference station "Credit River @ MGCC".
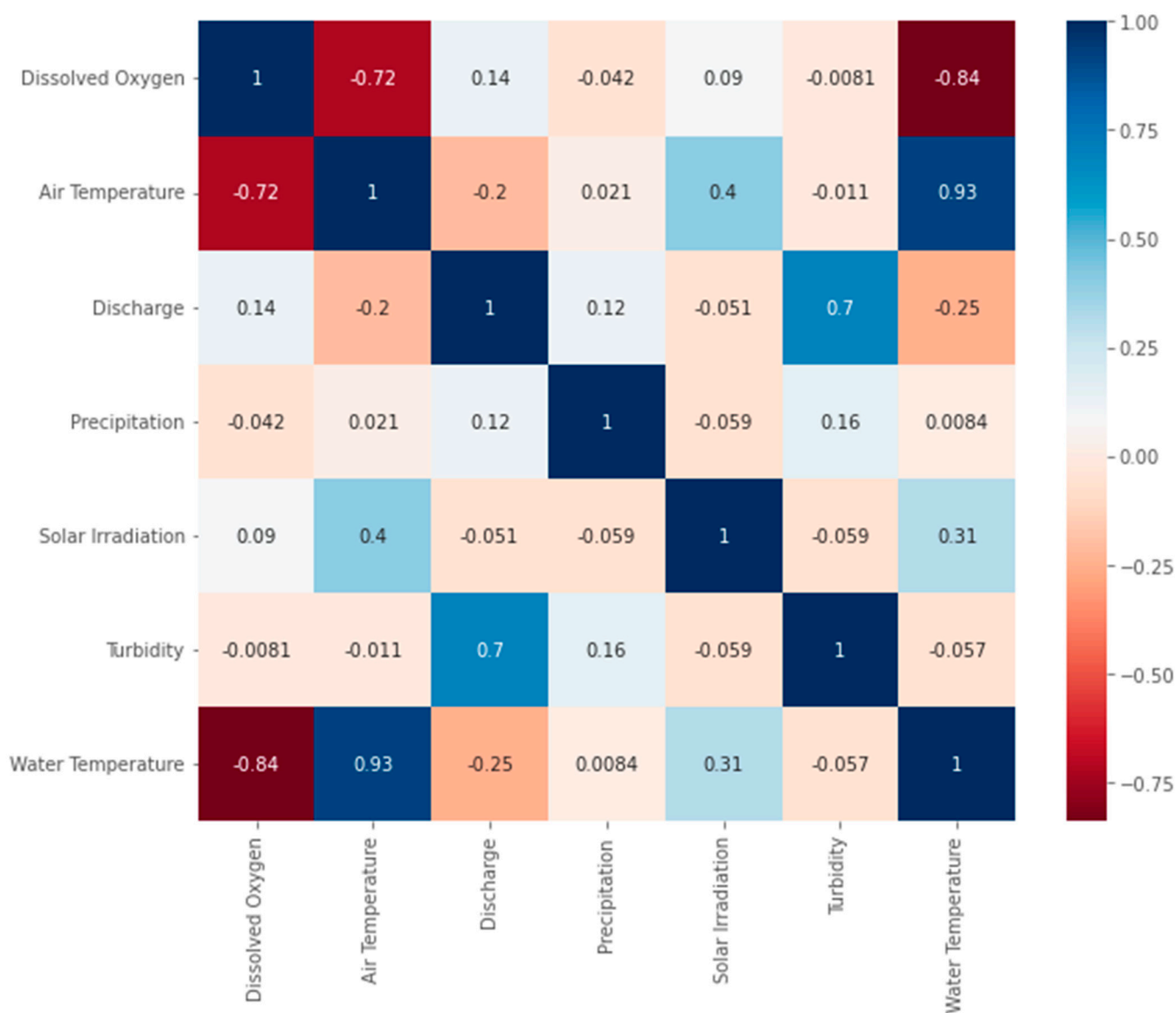
Figure 3. Correlation matrix for the Credit River variables used for dissolved oxygen concentration forecasting using Pearson correlation coefficient [52]. The blue colors indicate positive correlation, and the red colors show negative correlation.

In Figure 3, the used input data present a strong negative correlation between water temperature and air temperature, an expected behavior for DO since higher temperatures impair gas solubility in water [53–55]. We observe that positive correlation occurs between the discharge variable and DO due to the discharge's influence on the disturbance to the water body, leading to oxygen transport throughout the river's course. Similar behavior can be noticed between DO and solar irradiation, given the occurrence of the photosynthesis, which also increases oxygen levels on the river. The solar irradiance, albeit responsible to define daily cycles within the river, has its effect as not immediate in this sense. In fact, the exposure to sun light will affect the temperature in the river, which is related to the DO levels in the watershed. This can be visualized in Figure 3, where the correlation between solar irradiance and water temperature (0.31) is similar to that between solar irradiance and air temperature (0.4), indicating a positive correlation between them.

The remaining precipitation and turbidity showed a very small correlation with the dissolved oxygen attribute, indicating that they may not directly affect the DO levels in the

water body. Also, the negative value of these variables indicates that they may reduce the photosynthetic process by the aquatic flora [56,57].

It is important to note that the occurrence of high correlation values in different input attributes does not necessarily positively impact the forecasting result. This may introduce high variance into the predictive model due to collinearity, sometimes damaging its performance [58,59]. However, incorporating input variables with low correlation values into the forecasting model may be advantageous. These inputs may contain valuable spatiotemporal information that can help reveal the connection between the input and output parameters, ultimately improving the model's predictions [50].

3. Methodology

3.1. Benchmarking Models

The achievements of the assessed models were evaluated against the results of two benchmarking models: the persistence and eXtreme Gradient Boosting (XGBoost). The persistence model is a simple benchmarking method widely used to evaluate any forecasting model. It simply states that the current value of the target variable is the same as the previously measured one. This approach has good performance on forecasting values in short time windows, but as the forecasting horizon increases, it can no longer capture the complex non-linear behavior underneath the dataset nor external factors influencing future values, quickly deteriorating its results [60,61].

XGBoost is a tree-based ML model, improving the random forests approach. Based on the bagging sampling, the XGBoost algorithm generates smaller tree models. It later combines them into a unique bigger one, reducing the final variance and, therefore, the risk of overfitting [62–64]. The XGBoost model performs remarkably well in handling missing values in the dataset, and by being scalable to different applications [65], it is successfully applied in various research fields [66–68].

3.2. DNN-Transformer Model

The evaluated DNN-Transformer model utilizes a deep learning technique that incorporates a self-attention mechanism. This enables the model to concentrate on important features of the input information, improving its decision-making capabilities [69]. On its original configuration [70], the transformer architecture determines how the data attend to each one of its elements by implementing positional encoding together with multi-head attention, later being processed by a feed-forward structure that will output the processed data by the encoder. The generated information from the encoder is later used in the decoder structure, where a multi-head structure will again process it before being fed into another feed-forward internal model, which will output the transformer result.

For the present work, the transformer architecture was adapted so it can be applied to a time-series forecasting model. The DNN-Transformer architecture used in the present study is depicted in Figure 4.
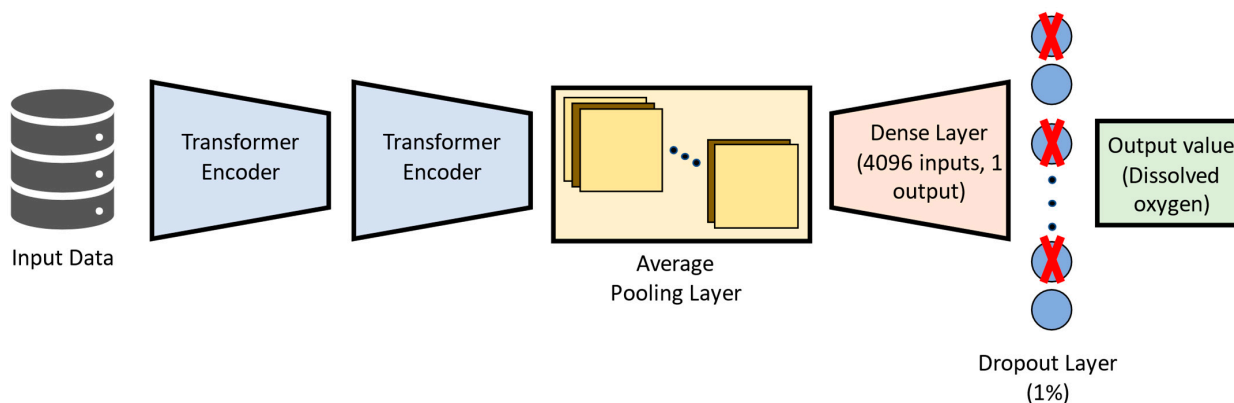

Figure 4. Applied DNN-Transformer model structure.

In Figure 4, the input data pass through two transformer structures. Then, their output is processed by an average pooling layer, which connects to a dense layer. The dropout layer [71] is then used to avoid overfitting by relying on an ensemble of different models after the training phase, thus reducing the model's variance [63,72]. Finally, the forecasted value for dissolved oxygen is output by the DNN-Transformer model.

3.3. GNN-SAGE Model

The Graph Neural Network Sample and Aggregate (GNN-SAGE) is a cutting-edge model that applies both graph theory and deep learning techniques. Unlike traditional ML models, the graph neural network approach naturally considers graph-structured data, retaining relevant spatiotemporal information ruling the behavior of input variables and target variables, boosting time-series forecasting models [73,74]. For the SAGE paradigm, the model can generalize unknown data by sampling a constant set of nodes, later aggregating them using an aggregator function [75–77], achieving excellence in extracting spatiotemporal information underneath the data [25,38]. The structure used in the present work is depicted in Figure 5.
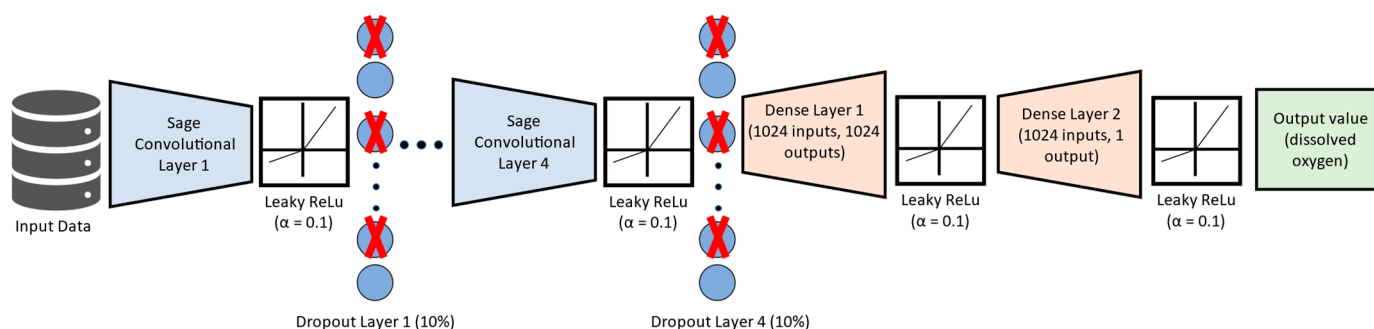

Figure 5. Applied GNN-SAGE model structure.

Figure 5 feeds the input spatiotemporal data into the GNN-SAGE model. First, they pass through a succession of graph convolutional layers, followed by the Leaky ReLU function [78,79] and a subsequent dropout layer. After four convolution processes, the data are fed to a dense layer containing 1024 inputs and 1024 outputs, followed again by the Leaky ReLU activation function and a final dense layer containing 1024 inputs and 1 output. Finally, future dissolved oxygen values are then generated by GNN-SAGE.

One key advantage of graph-based architectures, like GNN-SAGE, over traditional DL approaches is their natural ability to process multi-spatiotemporal data from surrounding monitoring stations. This enables them to effectively uncover the inherent connection among input and target variables when implemented in predictive applications [50,73,74]. The capacity of GNN-SAGE to extract spatiotemporal information from the data is relevant to this study, given that DO levels depend on the dataset's temporal and spatial features.

Each model was developed using spatiotemporal data from the monitoring stations and past information called time lags. In this work, the time lag consists of a subseries of the original time series dataset, representing past information in hours, that is fed to the model. That is, being x_i the attribute value taken in the i-th hour, its time lag is defined as a subgroup of the attribute x, where $\{x_{i-1}, x_{i-2}, \ldots, x_{i-n}\}$ contains up to n time lags. The time lag in this study has temporal resolution of 1 h to match the resampling mentioned in the previous section.

For training and testing phases, the dataset comprising the years 2016 to 2020 was split into the ratio 80/20. In this case, the data from 2016 to 2019 were used for training the XGBoost, DNN-Transformer, and GNN-SAGE models, while the remaining year 2020 was reserved for testing these models for a forecasting horizon of 6 h.

3.4. SHAP Analysis

Despite the excellent results achieved by ML and DL models in recent studies, these models still lack interpretability, posing as a major hindrance to a full understanding and interpretation of how they work [80–82]. Aiming to solve this problem, the Shapley additive explanation (SHAP) statistic was developed to be a valuable tool for better understanding the black-box behavior of ML applications. This game-theory-based technique provides a deeper understanding of ML models by examining the relationship among input attributes and the resulting outcome. It does so by combining additive feature attribution in terms of each input variable's significance, correlation, and impact on the final prediction, resulting in a better understanding of how complex models work [83].

3.5. Evaluation Metrics

The studied models were evaluated using different error metrics, commonly found in the ML field for time-series applications. Error metrics are an important indicator about the ML models' performance and accuracy, allowing posterior comparison with their peers [84]. The used metrics were RMSE, nRMSE, MAE, nMAE, MAPE, MBE, forecast skill, and R^2 [85,86].

4. Results

4.1. Evaluation of Different Time Lags over the Model's Performance

The performances of the DNN-Transformer and GNN-SAGE approaches were compared with the benchmarking ones for a different number of time lags. In this part, each model was evaluated regarding the RMSE metric. The results are presented in Figure 6 for the training set, analyzing time lags ranging from 1 h to 96 h of previous information.
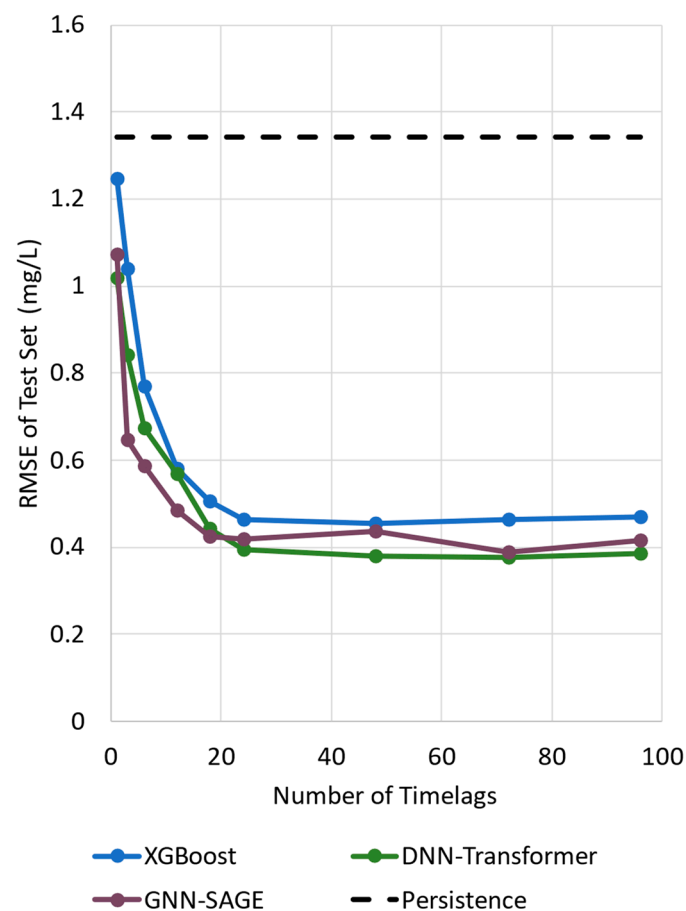

Figure 6. Impact of different number of time lags for dissolved oxygen forecasting.

Figure 6 illustrates the evaluated models' performance using different time lags by using only dissolved oxygen as the input parameter. All the models can surpass the persistence and XGBoost benchmarking for the assessed time lag values, presenting similar behavior: a steep decrease in the RMSE value, followed by convergence after 20 time lags. The deep learning paradigms, DNN-Transformers, and GNN-SAGE benefited more as time lags increased. Both XGBoost and DNN-Transformers presented more stable outcomes when compared with GNN-SAGE, which demonstrated a small variation in its results.

For the evaluated models, the best results were achieved by DNN-Transformers for 48 time lags (RMSE = 0.38 mg/L). The proposed GNN-SAGE approach was the second best-performing model, reaching the optimal RMSE value of 0.39 for 72 time lags, a comparable value for DNN-Transformer. The DNN-Transformer model improved dissolved oxygen forecasting by 72% compared with the persistence model and by 17% in comparison with XGBoost. For GNN-SAGE, the improvements were 71% and 15%, when compared with the persistence and XGBoost models, respectively. Finally, using the results depicted in Figure 6, we implemented 72 time lags for the input attributes, i.e., data from up to the previous 72 h for the variables displayed in Figure 3, including dissolved oxygen concentrations, to determine future DO levels using GNN-SAGE, while DNN-Transformer was implemented using 48 time lags.

### 4.2. Results of Dissolved Oxygen for 6 h Ahead

To determine the optimal configuration for forecasting dissolved oxygen levels, various input variables were tested. Using a step-by-step approach [50] for the considered parameters, the models' results were initially evaluated using just past dissolved oxygen information. Additional input parameters were then added to the models, and the resulting forecasts were analyzed. If the addition of an attribute did not improve the model's forecasting performance, leading to a greater error than previous configurations, that attribute was removed. This process was repeated for all models until all attributes depicted in Figure 3 were assessed, considering a 6 h forecasting horizon. The results are presented in Figure 7.

Figure 7 presents the different forecasting performances for each evaluated model. In Figure 7a, it is possible to observe that the XGBoost approach reached improved predictive results as more input variables were added. Its best configuration was set as using dissolved oxygen, air temperature, precipitation, and water temperature, achieving an RMSE value of 0.37 mg/L. Interestingly, the DL transformer-based model's best configuration reached the same error value as XGBoost. Analyzing Figure 7b, using only dissolved oxygen as an input attribute, DNN-Transformer achieved an RMSE outcome of 0.37 mg/L. The model did not need to use any exogenous variables to achieve its best result. Contrarywise, the last panel indicates that GNN-SAGE also benefited from increased input parameters. In fact, it was the best-performing model, reaching an RMSE of 0.34 mg/L when fed with dissolved oxygen, air temperature, precipitation, and water temperature. Interestingly, despite the high correlation depicted in Figure 3 between air and water temperature, the further addition of the latter proved beneficial for the XGBoost and GNN-SAGE models alike, increasing their performance by a tiny margin. The graph-based model enhanced its DO forecasting performance by 8% when compared with both XGBoost and DNN-Transformer.
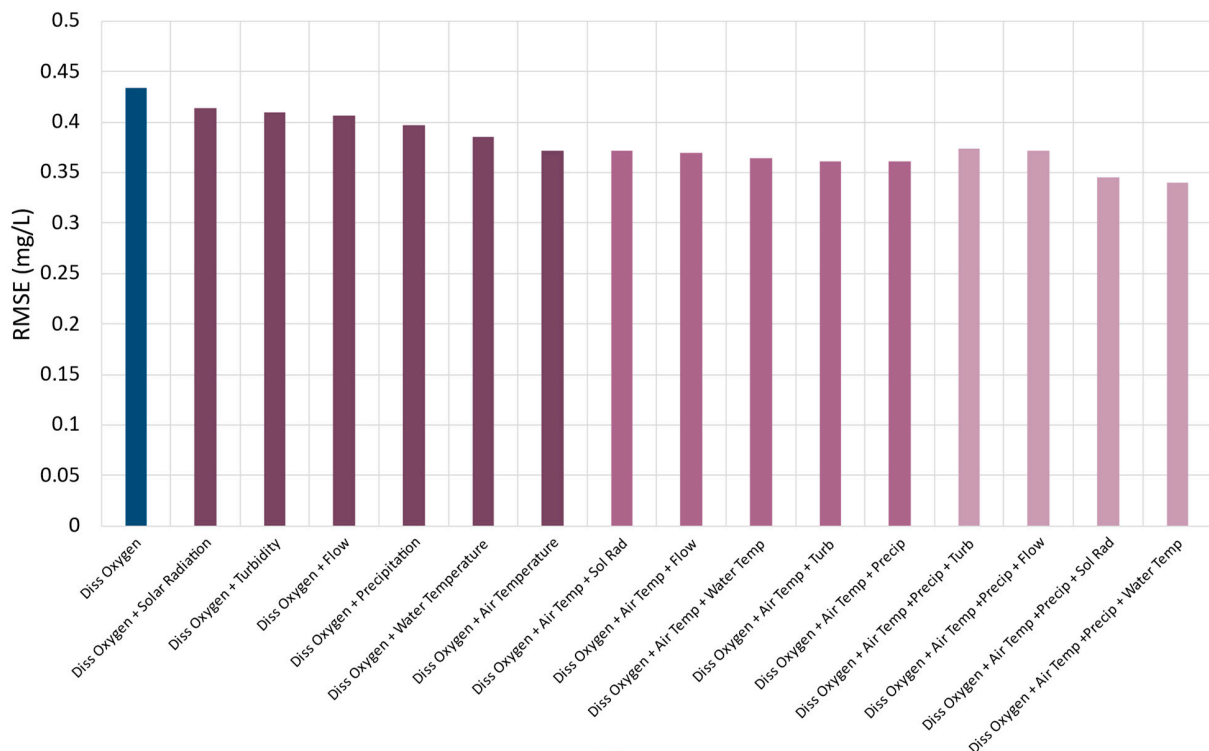
The models XGBoost, DNN-Transformer, and GNN-SAGE can be evaluated by analyzing Figures 8–10, which contain a scatter plot and marginal distributions for the real measured dissolved oxygen and the predicted values by the models, in mg/L.
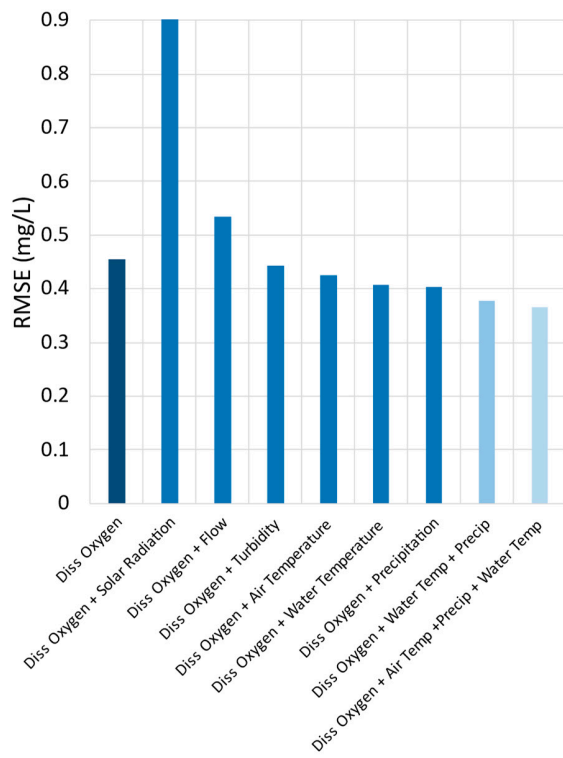
(a)
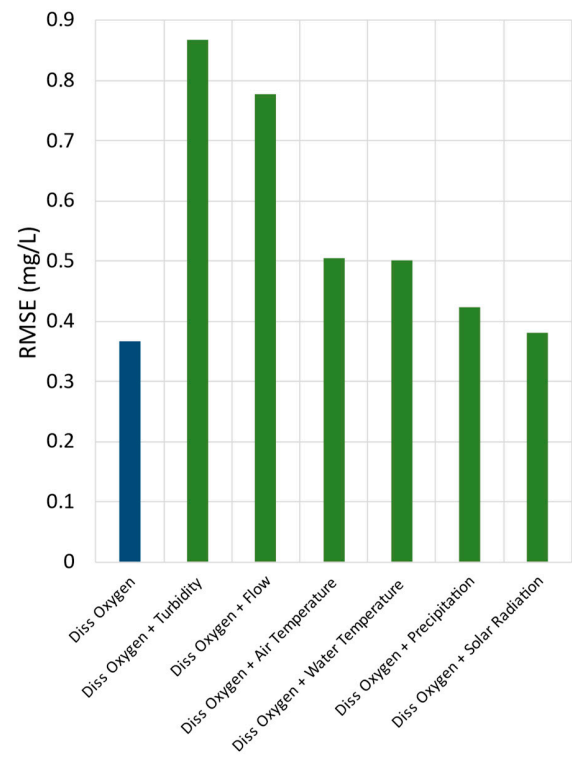
(b)

(c)

Figure 7. This figure depicts the results for this approach, where the *x*-axis shows the used input variables combination and, in the *y*-axis, their respective RMSE values in mg/L. In Figure 7, the lighter the column color, the lower the RMSE value. The effect of different input parameters for dissolved oxygen 6 h ahead estimation using (**a**) XGBoost, (**b**) DNN-Transformer, and (**c**) GNN-SAGE models.
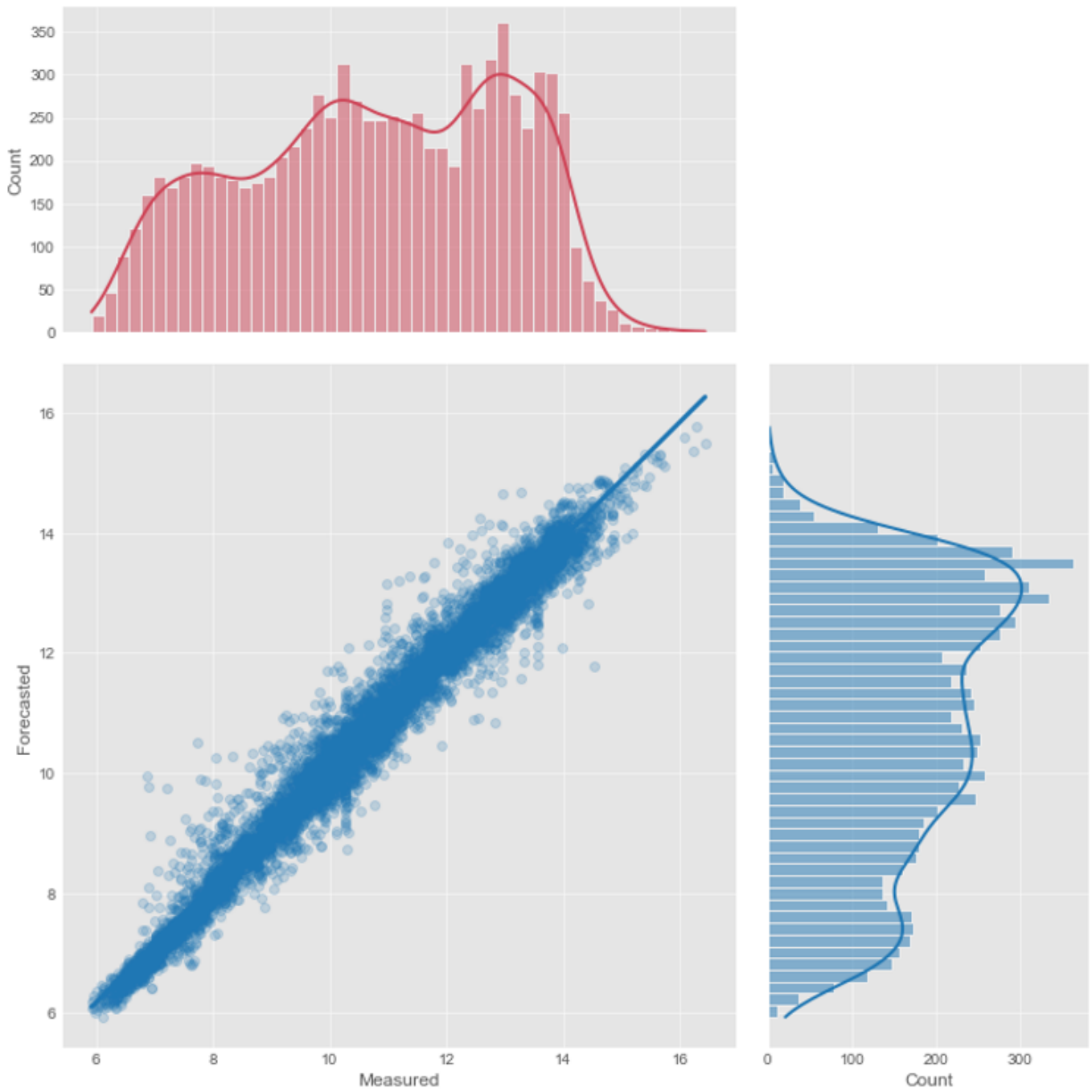
Figure 8. Scatter plot illustrating the estimated and observed dissolved oxygen levels for a forecast window of 6 h using XGBoost.
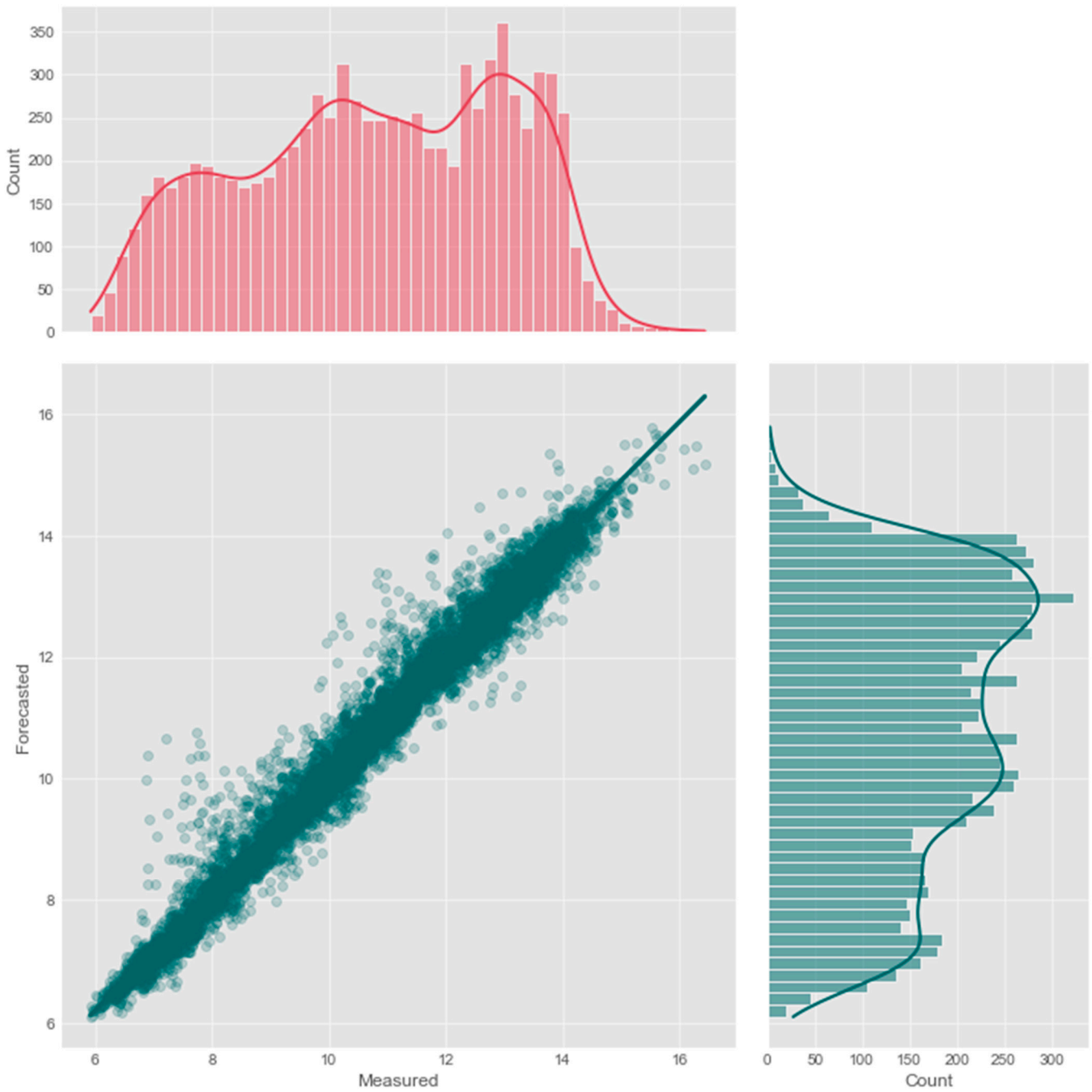
Figure 9. Scatter plot illustrating the estimated and observed dissolved oxygen levels for a forecast window of 6 h using DNN-Transformer.
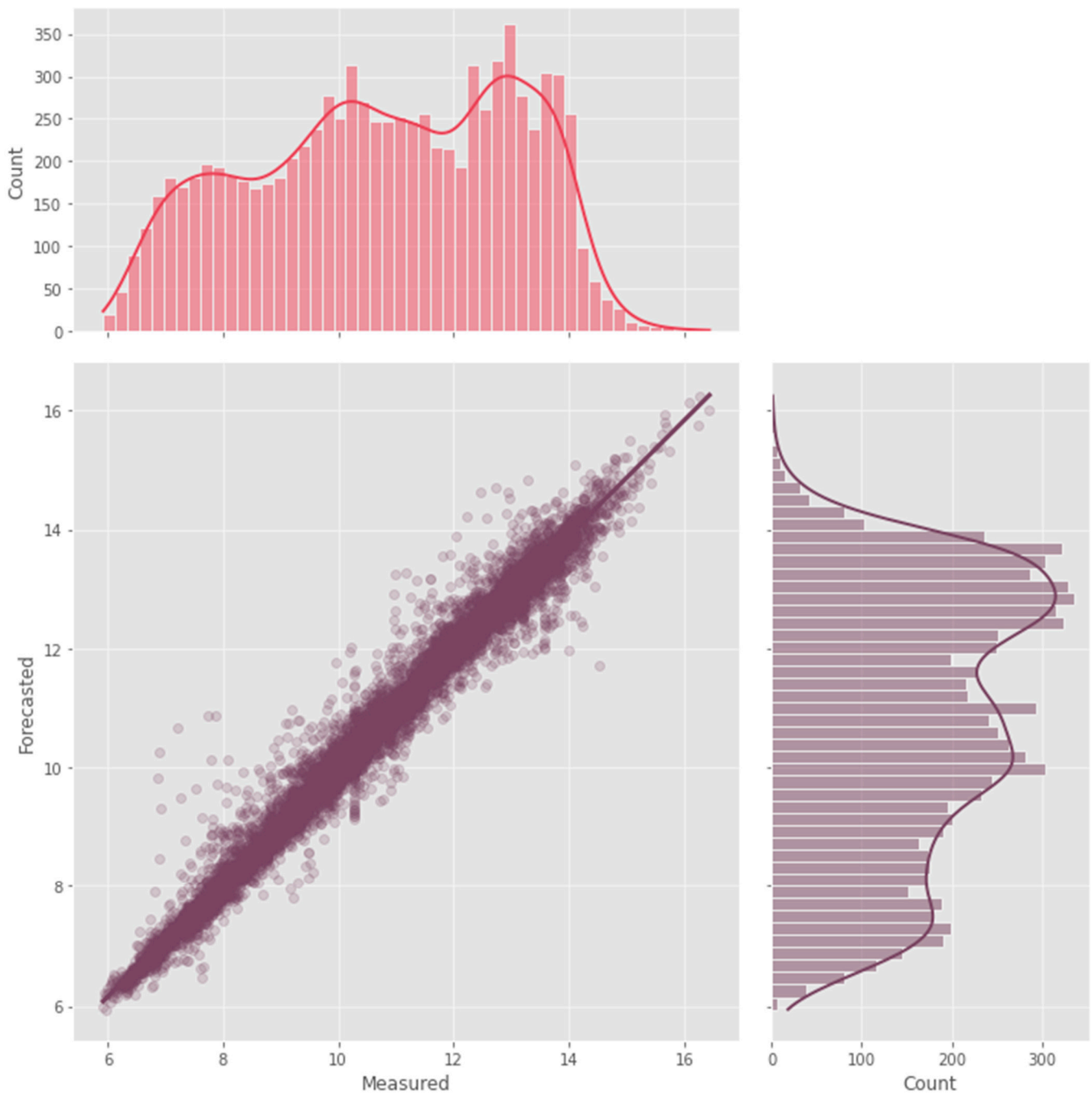
Figure 10. Scatter plot illustrating the estimated and observed dissolved oxygen levels for a forecast window of 6 h using GNN-SAGE.

Figures 8–10 present very similar results for the three evaluated models. They have well-aligned scatter points with the regression line, and very similar data distribution, as indicated by the histograms. However, the GNN-SAGE model has superior performance over the tree-based and transformer-based approaches, reaching an excellent coefficient of determination of 97.6%. Figures 11–13 show the model's results over 30 days, displayed on the *x*-axis, comparing them with real measured data for dissolved oxygen in mg/L, displayed on the *y*-axis.

Figure 11. Comparison of test data and forecasted data by XGBoost model.

Figure 12. Comparison of test data and forecasted data by DNN-Transformer model.

Figure 13. Comparison of test data and forecasted data by GNN-SAGE model.

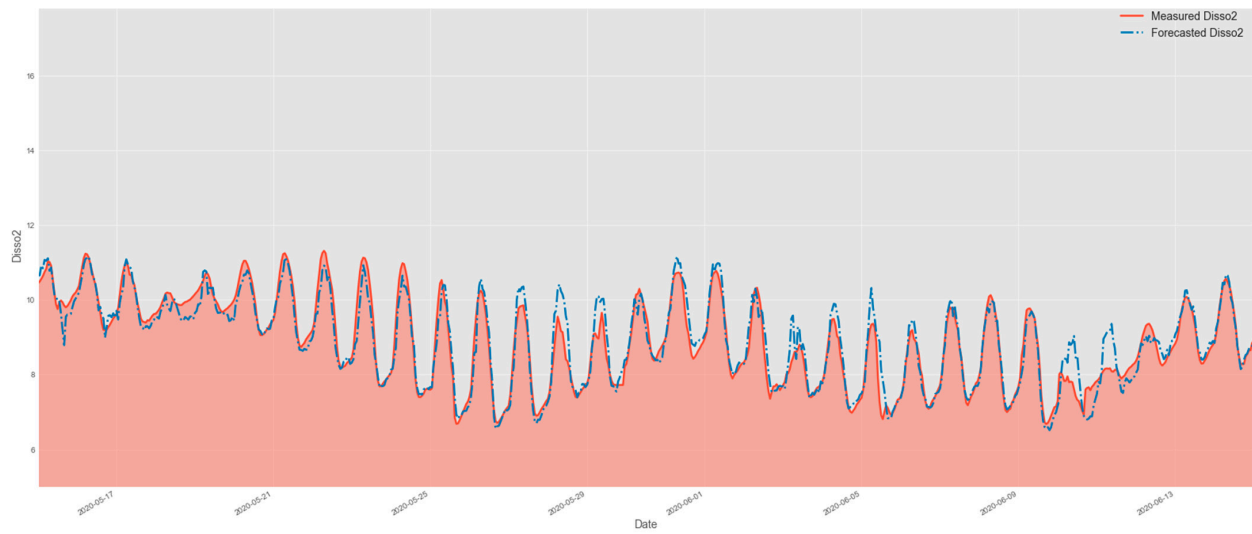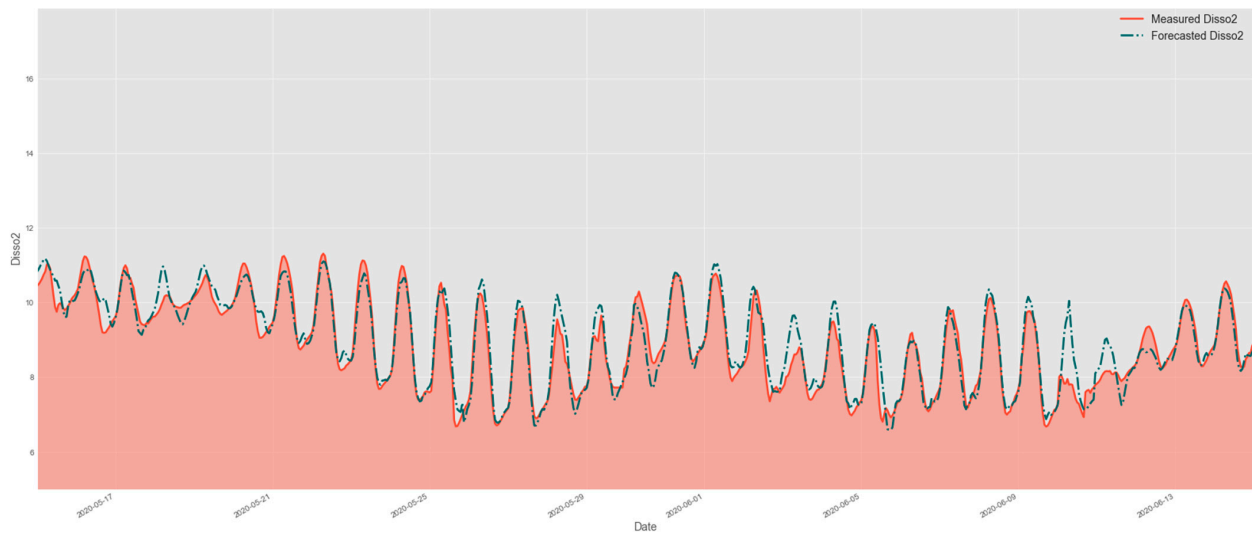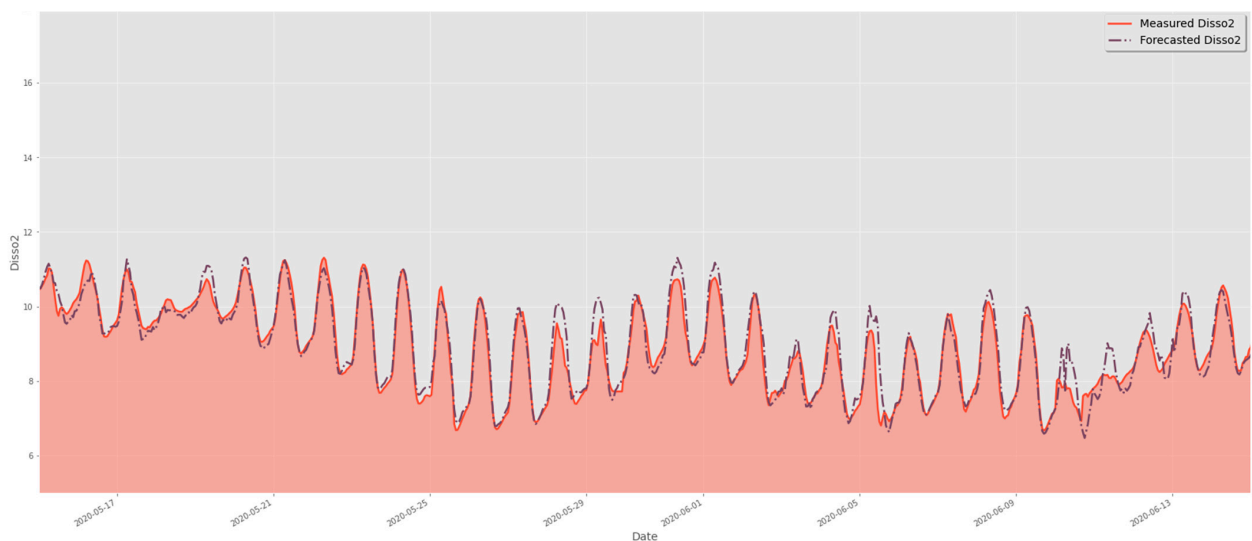In Figures 11–13, comparing the evaluated models' results for the same assessed period, we can observe that the graph-based structure has a superior performance in identifying the peak concentrations of dissolved oxygen, without any lagging in its results. Comparing XGBoost with GNN-SAGE, we notice that the graph-based approach has superior performance in identifying DO peaks, closely following their shape with a marginal overestimation. Comparing the transformer-based and GNN-SAGE models, we can notice that the DNN-Transformer model tends to maintain the historical behavior of the DO, as seen in the fifth peak from right to left in Figure 12. This is an expected behavior from the transformer approach since it determines the most likely outcome based on the previous sequential data [70]. Contrariwise, GNN-SAGE can aggregate the influence of the spatiotemporal relationships of the input parameters, deeming it sensitive to changes in the DO concentration and thus identifying the peaks within the assessed period, as seen in the same fifth peak from right to left in Figure 13.

5. Discussion of the Results

5.1. Analysis of the Results of Dissolved Oxygen for 6 h Ahead

When applied to the proposed scenario, the graph-based model GNN-SAGE presented the best performance, providing cutting-edge results for DO forecasting, a highly non-conservative substance, for 6 h ahead forecasting. By using dissolved oxygen, air temperature, precipitation, and water temperature as input attributes, the proposed GNN-SAGE model showed that, when compared with XGBoost and DNN-Transformer, the graph structure can better model and analyze the DO levels. GNN-SAGE can satisfactorily extract spatiotemporal data from the study station and its surroundings, substantially improving the DO forecasting results, reaching R^2 of 97.6% and RMSE of 0.34 mg/L, representing improved DO forecasting by 8% compared with both XGBoost and DNN-Transformer RMSE. The superior performance of GNN-SAGE for hydrological applications has been attested in previous studies [38,50].

The proposed model also showed better tracking of DO peaks for 4 weeks, as presented in Figures 8–10. In contrast to the transformer-based approach, the GNN-SAGE model proved to be more sensitive to the spatiotemporal attributes, being able to adapt its forecasting to the current DO concentrations, while DNN-Transformer posed a much more rigid paradigm, tending to maintain the data trend for different DO measurements during the same assessed period, leading to a significant difference between the forecasted and real measured DO. The capacity of GNN-SAGE to adapt to the input spatiotemporal information from the monitoring stations reveals its potential to be used as a real-time monitoring tool for water quality forecasting.

Figures 11–13 show that the assessed models tend to slightly overestimate the DO values for a 6 h ahead time window. This behavior is well known in the ML community. It occurs when the forecasting parameter is too far into the future, leading to a disconnection between the input dataset and the output result. This results in average predicted outcomes [25]. Such difference can also be explained by the lack of sufficient spatiotemporal data provided to the forecasting models under scrutiny, resulting in lagged results as observed in Figures 11–13 [38,87–89].

In addition to these factors, discrepancies between predicted and measured DO values can also be caused by abiotic and biotic factors composing the geochemical environment of the river. Salinity, atmospheric pressure, turbulence, and other geochemical parameters can impact the accuracy of the predictions. Similarly, biotic factors, like photosynthesis and respiration rates of the aquatic flora, can also contribute to the observed discrepancies. The inclusion of these parameters as inputs to the GNN-SAGE approach can improve the model's performance and reduce the difference between the forecasted and real DO values.

Forecasted values greater than the actual DO values can prevent proper decision-making by managers responsible for keeping the watershed's water quality in accordance with regulatory requirements, leading to a deficit in DO. In this sense, the proposed GNN-

SAGE model was the best-performing assessed model, returning the least overestimation on DO levels 6 h ahead, achieving the lowest MAPE value of 2.22%.

5.2. Analysis of Results of the SHAP Analysis

Figure 14 shows the results of the SHAP analysis for the 6 h forecasting horizon, where the attributes are arranged in a descending manner from the variable with the most influence on the one with the least influence when used to determine future DO concentrations. The rightmost bar is a scale for the feature value based on the correlation between the variable and the forecasted attribute, indicating that an elevated correlation also has an elevated value. SHAP values greater than zero indicate that positive attribute values positively contribute to DO forecasting, while negative values negatively influence the model's prediction.
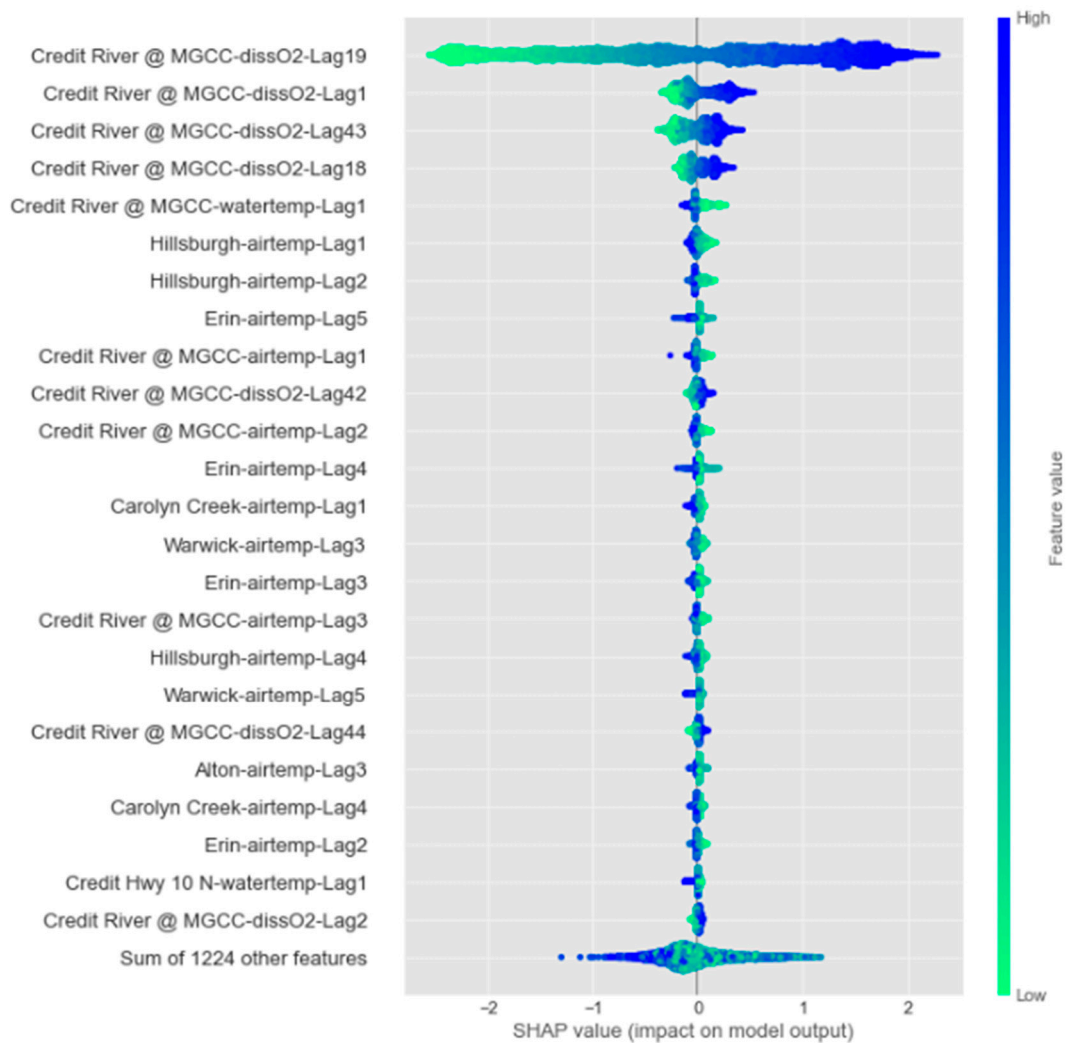
Figure 14. Results from the SHAP analysis for predicting DO concentrations 6 h ahead using GNN-SAGE model.

Figure 14 elucidates that using past DO data, i.e., "dissO2-Lag19", "dissO2-Lag1", "dissO2-Lag43", and "dissO2-Lag18", from the reference station "Credit River @ MGCC" provides the most significant amount of spatiotemporal data for the model's forecasting. After that, water and air temperatures are other important factors contributing to the model's DO predictions. For these variables, verifying the relevant influence of spatiotemporal data from the neighboring stations, such as "Hillsburgh" and "Erin", is important, indicating that the model benefits from their information.

The relevant influence of the temperature attribute over the model's result was expected. Figure 3 depicts the elevated correlation between dissolved oxygen and temperature, as elevated temperatures directly affect the oxygen solubility on the water, also contributing to the photosynthesis process and consequent oxygen concentration on the watershed. The importance of the temperature in future estimations of DO has also been studied and verified in previous works [28,42,90], attesting its significant contribution to forecasting models.

The findings from the SHAP analysis provide better insight into the workings of the GNN-SAGE model, showing that it can satisfactorily capture the relationship between the input variables from the study station and its neighboring locations. The spatiotemporal data retrieved from the dataset by the model improved its forecasting capability, with the past DO and temperature data having a relevant influence.

5.3. Analysis of the Comparison between the GNN-SAGE Results and Literature-Found Values

In order to understand where the proposed GNN-SAGE approach stands among the current forecasting models, the results yielded by the graph-based model in the current study were compared with the ones found in the literature. Nonetheless, such comparison may not provide a truthful picture between the compared models, meaning that their results are inherent to their specific methodologies, input data, geographic location, and application. Yet, a comparison is still a viable approach to determine how well models can perform when compared with their peers. To facilitate the comparison between the proposed GNN-SAGE with the results from the literature, the following Table 3 compiles the results for different metrics for this model when applied to a 6 h forecasting horizon, and Table 4 presents the results from the literature. For further information on the applied metrics, we recommend reference [85].

Table 3. Values of the error metrics for a 6 h forecasting horizon using GNN-SAGE paradigm.

Metric	Value
RMSE	0.34 mg/L
nRMSE	3.17%
MAE	0.23 mg/L
nMAE	2.14%
MAPE	2.22%
MBE	0.01 mg/L
Forecast skill	74.30%
R^2	97.63%

Table 4. Values for dissolved oxygen concentrations forecasting found in the literature.

Model	Metric Value	Author
Delft3D	RMSE 1.18 mg/L	Oliveira et al. [7]
Delft3D	MAE 1.03 mg/L MAPE 15.9%	Curbani et al. [91]
Prophet	RMSE 0.71 mg/L MAE 0.55 mg/L	Kogekar et al. [92]
LSSVM-BA	RMSE Mean value 0.79 mg/L MAE Mean value 0.94 mg/L	Yaseen et al. [93]

Table 4. *Cont.*

Model	Metric Value	Author
Bi-LSTM	RMSE 0.2 mg/L MAE 0.15 mg/L	Ahmed et al. [34]
CEEMDAN–CNN–LSTM	RMSE 0.26 mg/L for 4 h forecasting horizon 0.28 mg/L for 8 h forecasting horizon 0.31 mg/L for 12 h forecasting horizon 0.34 mg/L for 16 h forecasting horizon 0.39 mg/L for 20 h forecasting horizon 0.48 mg/L for 24 h forecasting horizon (Average RMSE of 0.34 mg/L) MAPE 2.55% for 4 h forecasting horizon 2.79% for 8 h forecasting horizon 3.00% for 12 h forecasting horizon 3.30% for 16 h forecasting horizon 3.65% for 20 h forecasting horizon 4.56% for 24 h forecasting horizon (Average MAPE of 3.31%)	Sha et al. [36]

In references [7,92], DO was modeled by the physical approach of Delft3D for rivers located in Portugal and Brazil, respectively. From the comparison between the GNN-SAGE results and those using a physics-based approach, it can be concluded that the graph model provides superior results considering RMSE and MAPE, substantially improving DO forecasting. Also, the MAPE comparison allows us to conclude that GNN-SAGE provides more accurate estimates for DO than the model used in [91]. Furthermore, it is worth noting that GNN-SAGE, a graph-based model, can inherently process multi-spatiotemporal information contained in the input dataset, without requiring explicit programming and/or feature preprocessing. Another advantage of the GNN-SAGE paradigm is its simpler implementation: differently from the Delft3D model, which models the DO levels by the finite difference method and thus requires the domain to be defined by a mesh grid, GNN-SAGE uses only the inputs presented in Figure 2.

In reference [92], the authors implemented the Prophet ML model. The Prophet was developed and made available by Facebook's Core Data Science group [94] and is a cumulative approach that excels in analyzing and forecasting non-linear data trends [95]. Their work implemented the Prophet for a DO forecasting task for a river in India. Comparing the GNN-SAGE results with those in [92], the proposed model clearly provides superior results over a simpler ML model implementation such as the Prophet. The superiority of GNN-SAGE is expected in this scenario, once it can process and identify seasonality and spatial information from the dataset, while the Prophet may handle temporal data only. From this comparison, we can conclude that spatial data play a major role in determining future DO levels by incorporating relevant data for the model's prediction, as stated in the SHAP analysis results in Figure 13.

In work [93], the authors proposed a hybrid model called the Least Square Support Vector Machine-Bat Algorithm (LSSVM-BA) for monthly DO estimation in the USA. Their results for the best-performing LSSVM-BA for monthly DO predictions were, on average, 0.79 mg/L and 0.94 mg/L for RMSE and MAE metrics, respectively. When directly comparing these results with those of GNN-SAGE, the graph-based model surpassed the results found in the literature with a significant difference on both metrics, again proving to be a superior approach to this task. In studies [34,36], different DL paradigms were implemented for DO forecasting. Interestingly, among the results found in the literature in Table 2, the DL models were the best-performing ones, offering significant improvement for DO estimation

over both physical and traditional ML approaches, achieving error values within the same order of magnitude. In work [34], Bi-LSTM was implemented to model future DO levels in Pakistan using remote sensing data up to three time steps ahead. The study [36] used a hybrid model consisting of CNN–LSTM and CEEMDAN for predicting DO in China up to 24 h ahead. The CNN–LSTM hybrid configuration is well known for its ability to process spatial and temporal data through its convolutional and recurrent structures, respectively. Both papers reached similar results for DO forecasting regarding the RMSE metric. However, comparison with GNN-SAGE proved that the proposed paradigm in this work can surpass the results for the DL models by a significant margin, providing more accurate and precise DO forecasting.

6. Conclusions

In this study, a new GNN-SAGE deep learning model for forecasting DO levels in natural streams was developed using a network of real-time water quality monitoring stations in an urban watershed. The accuracy of the model forecasts for a forecasting horizon of 6 h was compared with the popular, XGBoost and DNN-Transformer models. The best-performing GNN-SAGE configuration was obtained using 72 h time-lag and input parameters dissolved oxygen, air temperature, precipitation, and water temperature from the reference station "Credit River @ MGCC" and its neighboring ones.

The results show that the proposed new GNN-SAGE can gather influence from the provided network of real-time hydrometric and water quality monitoring stations and is able to more accurately forecast changes in DO levels. This leads to an improvement of 8% in the RMSE metric when compared with both XGBoost and DNN-Transformer, and it reaches an R^2 value above 97%. The SHAP analysis outcomes elucidated the importance of spatial data coming from the neighboring stations.

By comparing the GNN-SAGE results with those found in the literature, it was possible to attest the superior performance of the graph-based model, which surpassed every assessed model in Table 4 with a significant margin for both RMSE and R^2 values. These findings made the GNN-SAGE model a valuable DO forecasting tool, offering precise and cutting-edge predictions in urban river watersheds.

For future applications of the proposed GNN-SAGE model, other hydrological parameters can be assessed using this approach and directly determine the water quality index. Furthermore, more spatial information can be added to future implementations of GNN-SAGE, aiming to reduce its geographical bias once it was modeled and validated for only one location. To this end, spatiotemporal data from different rivers can be added to the model, enhancing its generalization and broadening its geographic application. Future works can also investigate the performance of the proposed GNN-SAGE when applied to estuary regions. Given the significant spatial extent of estuaries, the graph model is anticipated to be highly adaptable for water quality monitoring in these areas. The implementation of a graph-based model to assess water quality on estuaries is a promising topic worth to be further explored.

Developing a precise and accurate model to forecast real-time hydrological parameters is relevant in the current socio-economic–environmental scenario. Technologies like GNN-SAGE have the potential to enhance the creation of new legislation regarding the protection and sustainable use of watersheds, as well as to support stakeholders in their decision-making and risk management strategies. These actions can mitigate further damage from human actions over the aquatic environment when used as a real-time forecasting tool.

Author Contributions: Conceptualization, J.V.G.T. and B.G.; methodology, P.A.C.R., J.V.G.T. and B.G.; software, P.A.C.R.; validation, P.A.C.R., J.V.G.T. and B.G.; formal analysis, P.A.C.R.; investigation, P.A.C.R., J.V.G.T. and B.G.; resources, J.V.G.T. and B.G.; data curation, J.V.G.T. and B.G.; writing—original draft preparation, V.O.S. and P.A.C.R.; writing—review and editing, V.O.S., P.A.C.R., J.V.G.T. and B.G.; visualization, V.O.S. and P.A.C.R.; supervision, J.V.G.T. and B.G.; project administration, J.V.G.T. and B.G.; funding acquisition, B.G. and J.V.G.T. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://cvc.ca/real-timemonitoring/, https://drive.google.com/drive/folders/13Ef-_EklzJze8pZx1oIDoQFKU304d7NF, and https://drive.google.com/drive/folders/136RH-G-nPVScO7Ln7OOC0WEYl3kk5kDW (accessed on 22 September 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Paul, J. *Crutzen and the Anthropocene: A New Epoch in Earth's History*; Benner, S., Lax, G., Crutzen, P.J., Pöschl, U., Lelieveld, J., Brauch, H.G., Eds.; The Anthropocene: Politik—Economics—Society—Science; Springer International Publishing: Cham, Switzerland, 2021; Volume 1, ISBN 978-3-030-82201-9.
2. Freeman, L.A.; Corbett, D.R.; Fitzgerald, A.M.; Lemley, D.A.; Quigg, A.; Steppe, C.N. Impacts of Urbanization and Development on Estuarine Ecosystems and Water Quality. *Estuaries Coasts* **2019**, *42*, 1821–1838. [CrossRef]
3. Rajkumar, H.; Naik, P.K.; Rishi, M.S. A Comprehensive Water Quality Index Based on Analytical Hierarchy Process. *Ecol. Indic.* **2022**, *145*, 109582. [CrossRef]
4. Regier, P.J.; González-Pinzón, R.; Van Horn, D.J.; Reale, J.K.; Nichols, J.; Khandewal, A. Water Quality Impacts of Urban and Non-Urban Arid-Land Runoff on the Rio Grande. *Sci. Total Environ.* **2020**, *729*, 138443. [CrossRef] [PubMed]
5. Giri, S. Water Quality Prospective in Twenty First Century: Status of Water Quality in Major River Basins, Contemporary Strategies and Impediments: A Review. *Environ. Pollut.* **2021**, *271*, 116332. [CrossRef] [PubMed]
6. Stajkowski, S.; Zeynoddin, M.; Farghaly, H.; Gharabaghi, B.; Bonakdari, H. A Methodology for Forecasting Dissolved Oxygen in Urban Streams. *Water* **2020**, *12*, 2568. [CrossRef]
7. Oliveira, V.H.; Sousa, M.C.; Morgado, F.; Dias, J.M. Modeling the Impact of Extreme River Discharge on the Nutrient Dynamics and Dissolved Oxygen in Two Adjacent Estuaries (Portugal). *J. Mar. Sci. Eng.* **2019**, *7*, 412. [CrossRef]
8. Waldron, M.C.; Wiley, J.B. *Water Quality and Processes Affecting Dissolved Oxygen Concentrations in the Blackwater River, Canaan Valley, West Virginia*; U.S. Department of the Interior, US Geological Survey: Charleston, WV, USA, 1996; p. 91.
9. Yang, F.; Liang, D. Random-Walk Simulation of Non-Conservative Pollutant Transport in Shallow Water Flows. *Environ. Model. Softw.* **2020**, *134*, 104870. [CrossRef]
10. Zhi, W.; Feng, D.; Tsai, W.-P.; Sterle, G.; Harpold, A.; Shen, C.; Li, L. From Hydrometeorology to River Water Quality: Can a Deep Learning Model Predict Dissolved Oxygen at the Continental Scale? *Environ. Sci. Technol.* **2021**, *55*, 2357–2368. [CrossRef]
11. Barletta, M.; Lima, A.R.A.; Costa, M.F. Distribution, Sources and Consequences of Nutrients, Persistent Organic Pollutants, Metals and Microplastics in South American Estuaries. *Sci. Total Environ.* **2019**, *651*, 1199–1218. [CrossRef]
12. Wang, J.; Fu, Z.; Qiao, H.; Liu, F. Assessment of Eutrophication and Water Quality in the Estuarine Area of Lake Wuli, Lake Taihu, China. *Sci. Total Environ.* **2019**, *650*, 1392–1402. [CrossRef]
13. Lowe, M.; Qin, R.; Mao, X. A Review on Machine Learning, Artificial Intelligence, and Smart Technology in Water Treatment and Monitoring. *Water* **2022**, *14*, 1384. [CrossRef]
14. Wei, Y.; Jiao, Y.; An, D.; Li, D.; Li, W.; Wei, Q. Review of Dissolved Oxygen Detection Technology: From Laboratory Analysis to Online Intelligent Detection. *Sensors* **2019**, *19*, 3995. [CrossRef] [PubMed]
15. Sun, K.; Cui, W.; Chen, C. Review of Underwater Sensing Technologies and Applications. *Sensors* **2021**, *21*, 7849. [CrossRef] [PubMed]
16. Deisenroth, M.P.; Faisal, A.A.; Ong, C.S. *Mathematics for Machine Learning*; Cambridge University Press: Cambridge, UK, 2020; ISBN 978-1-108-47004-9.
17. Varadharajan, C.; Appling, A.P.; Arora, B.; Christianson, D.S.; Hendrix, V.C.; Kumar, V.; Lima, A.R.; Müller, J.; Oliver, S.; Ombadi, M.; et al. Can Machine Learning Accelerate Process Understanding and Decision-Relevant Predictions of River Water Quality? *Hydrol. Process.* **2022**, *36*, e14565. [CrossRef]
18. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
19. Rocha, P.A.C.; Santos, V.O. Global Horizontal and Direct Normal Solar Irradiance Modeling by the Machine Learning Methods XGBoost and Deep Neural Networks with CNN-LSTM Layers: A Case Study Using the GOES-16 Satellite Imagery. *Int. J. Energy Environ. Eng.* **2022**, *13*, 1271–1286. [CrossRef]
20. Costa Rocha, P.A.; Johnston, S.J.; Oliveira Santos, V.; Aliabadi, A.A.; Thé, J.V.G.; Gharabaghi, B. Deep Neural Network Modeling for CFD Simulations: Benchmarking the Fourier Neural Operator on the Lid-Driven Cavity Case. *Appl. Sci.* **2023**, *13*, 3165. [CrossRef]
21. Marinho, F.P.; Rocha, P.A.C.; Neto, A.R.R.; Bezerra, F.D.V. Short-Term Solar Irradiance Forecasting Using CNN-1D, LSTM, and CNN-LSTM Deep Neural Networks: A Case Study With the Folsom (USA) Dataset. *J. Sol. Energy Eng.* **2023**, *145*, 041002. [CrossRef]

22. Huang, C.-J.; Kuo, P.-H. A Deep CNN-LSTM Model for Particulate Matter (PM2.5) Forecasting in Smart Cities. *Sensors* **2018**, *18*, 2220. [CrossRef]

23. Demolli, H.; Dokuz, A.S.; Ecemis, A.; Gokcek, M. Wind Power Forecasting Based on Daily Wind Speed Data Using Machine Learning Algorithms. *Energy Convers. Manag.* **2019**, *198*, 111823. [CrossRef]

24. Chimmula, V.K.R.; Zhang, L. Time Series Forecasting of COVID-19 Transmission in Canada Using LSTM Networks. *Chaos Solitons Fractals* **2020**, *135*, 109864. [CrossRef] [PubMed]

25. Oliveira Santos, V.; Costa Rocha, P.A.; Scott, J.; Van Griensven Thé, J.; Gharabaghi, B. Spatiotemporal Air Pollution Forecasting in Houston-TX: A Case Study for Ozone Using Deep Graph Neural Networks. *Atmosphere* **2023**, *14*, 308. [CrossRef]

26. Oliveira Santos, V.; Costa Rocha, P.A.; Scott, J.; Van Griensven Thé, J.; Gharabaghi, B. Spatiotemporal Analysis of Bidimensional Wind Speed Forecasting: Development and Thorough Assessment of LSTM and Ensemble Graph Neural Networks on the Dutch Database. *Energy* **2023**, *278*, 127852. [CrossRef]

27. Baïle, R.; Muzy, J.-F. Leveraging Data from Nearby Stations to Improve Short-Term Wind Speed Forecasts. *Energy* **2023**, *263*, 125644. [CrossRef]

28. Csábrági, A.; Molnár, S.; Tanos, P.; Kovács, J. Application of Artificial Neural Networks to the Forecasting of Dissolved Oxygen Content in the Hungarian Section of the River Danube. *Ecol. Eng.* **2017**, *100*, 63–72. [CrossRef]

29. Lu, H.; Ma, X. Hybrid Decision Tree-Based Machine Learning Models for Short-Term Water Quality Prediction. *Chemosphere* **2020**, *249*, 126169. [CrossRef]

30. Xu, C.; Chen, X.; Zhang, L. Predicting River Dissolved Oxygen Time Series Based on Stand-Alone Models and Hybrid Wavelet-Based Models. *J. Environ. Manag.* **2021**, *295*, 113085. [CrossRef]

31. Zhu, S.; Heddam, S. Prediction of Dissolved Oxygen in Urban Rivers at the Three Gorges Reservoir, China: Extreme Learning Machines (ELM) versus Artificial Neural Network (ANN). *Water Qual. Res. J.* **2020**, *55*, 106–118. [CrossRef]

32. Moon, J.; Lee, J.; Lee, S.; Yun, H. Urban River Dissolved Oxygen Prediction Model Using Machine Learning. *Water* **2022**, *14*, 1899. [CrossRef]

33. Bolick, M.M.; Post, C.J.; Naser, M.-Z.; Mikhailova, E.A. Comparison of Machine Learning Algorithms to Predict Dissolved Oxygen in an Urban Stream. *Environ. Sci. Pollut. Res.* **2023**, *30*, 78075–78096. [CrossRef]

34. Ahmed, M.; Mumtaz, R.; Anwar, Z.; Shaukat, A.; Arif, O.; Shafait, F. A Multi–Step Approach for Optically Active and Inactive Water Quality Parameter Estimation Using Deep Learning and Remote Sensing. *Water* **2022**, *14*, 2112. [CrossRef]

35. Barzegar, R.; Aalami, M.T.; Adamowski, J. Short-Term Water Quality Variable Prediction Using a Hybrid CNN–LSTM Deep Learning Model. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 415–433. [CrossRef]

36. Sha, J.; Li, X.; Zhang, M.; Wang, Z.-L. Comparison of Forecasting Models for Real-Time Monitoring of Water Quality Parameters Based on Hybrid Deep Learning Neural Networks. *Water* **2021**, *13*, 1547. [CrossRef]

37. Moghadam, S.V.; Sharafati, A.; Feizi, H.; Marjaie, S.M.S.; Asadollah, S.B.H.S.; Motta, D. An Efficient Strategy for Predicting River Dissolved Oxygen Concentration: Application of Deep Recurrent Neural Network Model. *Environ. Monit. Assess.* **2021**, *193*, 798. [CrossRef] [PubMed]

38. Oliveira Santos, V.; Costa Rocha, P.A.; Scott, J.; Van Griensven Thé, J.; Gharabaghi, B. A New Graph-Based Deep Learning Model to Predict Flooding with Validation on a Case Study on the Humber River. *Water* **2023**, *15*, 1827. [CrossRef]

39. Li, D.; Zhao, W.; Hu, J.; Zhao, S.; Liu, S. A Long-Term Water Quality Prediction Model for Marine Ranch Based on Time-Graph Convolutional Neural Network. *Ecol. Indic.* **2023**, *154*, 110782. [CrossRef]

40. Liu, G.; Jiang, Y.; Zhong, K.; Yang, Y.; Wang, Y. A Time Series Model Adapted to Multiple Environments for Recirculating Aquaculture Systems. *Aquaculture* **2023**, *567*, 739284. [CrossRef]

41. Ni, Q.; Cao, X.; Tan, C.; Peng, W.; Kang, X. An Improved Graph Convolutional Network with Feature and Temporal Attention for Multivariate Water Quality Prediction. *Environ. Sci. Pollut. Res.* **2022**, *30*, 11516–11529. [CrossRef]

42. Fang, Y.; Liu, H. A Spatiotemporal Dissolved Oxygen Prediction Model Based on Graph Attention Networks Suitable for Missing Data. *Environ. Sci. Pollut. Res.* **2023**, *30*, 82818–82833. [CrossRef]

43. Schuster-Wallace, C.J.; Murray, S.J.; McBean, E.A. Integrating Social Dimensions into Flood Cost Forecasting. *Water Resour. Manag.* **2018**, *32*, 3175–3187. [CrossRef]

44. Allen, B.; Mandrak, N.E. Historical Changes in the Fish Communities of the Credit River Watershed. *Aquat. Ecosyst. Health Manag.* **2019**, *22*, 316–328. [CrossRef]

45. McGovarin, S.; Nishikawa, J.; Metcalfe, C.D. Vitellogenin Induction in Mucus from Brook Trout (Salvelinus Fontinalis). *Bull. Environ. Contam. Toxicol.* **2022**, *108*, 878–883. [CrossRef] [PubMed]

46. Champagne, O.; Arain, M.A.; Leduc, M.; Coulibaly, P.; McKenzie, S. Future Shift in Winter Streamflow Modulated by the Internal Variability of Climate in Southern Ontario. *Hydrol. Earth Syst. Sci.* **2020**, *24*, 3077–3096. [CrossRef]

47. Rincón, D.; Khan, U.; Armenakis, C. Flood Risk Mapping Using GIS and Multi-Criteria Analysis: A Greater Toronto Area Case Study. *Geosciences* **2018**, *8*, 275. [CrossRef]

48. Rincón, D.; Velandia, J.F.; Tsanis, I.; Khan, U.T. Stochastic Flood Risk Assessment under Climate Change Scenarios for Toronto, Canada Using CAPRA. *Water* **2022**, *14*, 227. [CrossRef]

49. Chu, C.; Minns, C.K.; Lester, N.P.; Mandrak, N.E. An Updated Assessment of Human Activities, the Environment, and Freshwater Fish Biodiversity in Canada. *Can. J. Fish. Aquat. Sci.* **2015**, *72*, 135–148. [CrossRef]

50. Oliveira Santos, V.; Costa Rocha, P.A.; Gharabaghi, B.; Thé, J.V.G. Graph-Based Deep Learning Model for Forecasting Chloride Concentration in Urban Streams to Protect Salt-Vulnerable Areas. *Environments* **2023**, *10*, 157. [CrossRef]

51. Canadian Council of Ministers of the Environment. Canadian Council of Ministers of the Environment. Canadian water quality guidelines for the protection of aquatic life: Dissolved oxygen (freshwater). In *Canadian Environmental Quality Guidelines*; Canadian Council of Ministers of the Environment: Winnipeg, MB, Canada, 1999.

52. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson Correlation Coefficient. In *Noise Reduction in Speech Processing*; Springer Topics in Signal Processing; Springer: Berlin/Heidelberg, Germany, 2009; Volume 2, pp. 1–4, ISBN 978-3-642-00295-3.

53. Jane, S.F.; Hansen, G.J.A.; Kraemer, B.M.; Leavitt, P.R.; Mincer, J.L.; North, R.L.; Pilla, R.M.; Stetler, J.T.; Williamson, C.E.; Woolway, R.I.; et al. Widespread Deoxygenation of Temperate Lakes. *Nature* **2021**, *594*, 66–70. [CrossRef]

54. Roman, M.R.; Brandt, S.B.; Houde, E.D.; Pierson, J.J. Interactive Effects of Hypoxia and Temperature on Coastal Pelagic Zooplankton and Fish. *Front. Mar. Sci.* **2019**, *6*, 139. [CrossRef]

55. Stajkowski, S.; Hotson, E.; Zorica, M.; Farghaly, H.; Bonakdari, H.; McBean, E.; Gharabaghi, B. Modeling Stormwater Management Pond Thermal Impacts during Storm Events. *J. Hydrol.* **2023**, *620*, 129413. [CrossRef]

56. Yang, R.; Sun, H.; Chen, B.; Yang, M.; Zeng, Q.; Zeng, C.; Huang, J.; Luo, H.; Lin, D. Temporal Variations in Riverine Hydrochemistry and Estimation of the Carbon Sink Produced by Coupled Carbonate Weathering with Aquatic Photosynthesis on Land: An Example from the Xijiang River, a Large Subtropical Karst-Dominated River in China. *Environ. Sci. Pollut. Res.* **2020**, *27*, 13142–13154. [CrossRef] [PubMed]

57. Xuan, Y.; Tang, C.; Liu, G.; Cao, Y. Carbon and Nitrogen Isotopic Records of Effects of Urbanization and Hydrology on Particulate and Sedimentary Organic Matter in the Highly Urbanized Pearl River Delta, China. *J. Hydrol.* **2020**, *591*, 125565. [CrossRef]

58. Dawoud, I.; Abonazel, M.R. Robust Dawoud–Kibria Estimator for Handling Multicollinearity and Outliers in the Linear Regression Model. *J. Stat. Comput. Simul.* **2021**, *91*, 3678–3692. [CrossRef]

59. Chan, J.Y.-L.; Leow, S.M.H.; Bea, K.T.; Cheng, W.K.; Phoong, S.W.; Hong, Z.-W.; Chen, Y.-L. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics* **2022**, *10*, 1283. [CrossRef]

60. Hanifi, S.; Liu, X.; Lin, Z.; Lotfian, S. A Critical Review of Wind Power Forecasting Methods—Past, Present and Future. *Energies* **2020**, *13*, 3764. [CrossRef]

61. Trebing, K.; Mehrkanoon, S. Wind Speed Prediction Using Multidimensional Convolutional Neural Networks. In Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI), Canberra, Australia, 1–4 December 2020; pp. 713–720.

62. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM Press: New York, NY, USA, 2016; pp. 785–794.

63. Ghojogh, B.; Crowley, M. The Theory Behind Overfitting, Cross Validation, Regularization, Bagging, and Boosting: Tutorial. *arXiv* **2023**, arXiv:1905.12787.

64. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A Comparative Analysis of Gradient Boosting Algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [CrossRef]

65. Bondi, A.B. Characteristics of Scalability and Their Impact on Performance. In Proceedings of the Second International Workshop on SOFTWARE and Performance—WOSP '00, Ottawa, ON, Canada, 17–20 September 2000; ACM Press: New York, NY, USA, 2000; pp. 195–203.

66. Wang, Y.; Pan, Z.; Zheng, J.; Qian, L.; Li, M. A Hybrid Ensemble Method for Pulsar Candidate Classification. *Astrophys. Space Sci.* **2019**, *364*, 139. [CrossRef]

67. Zhang, W.; Wu, C.; Zhong, H.; Li, Y.; Wang, L. Prediction of Undrained Shear Strength Using Extreme Gradient Boosting and Random Forest Based on Bayesian Optimization. *Geosci. Front.* **2021**, *12*, 469–477. [CrossRef]

68. Parsa, A.B.; Movahedi, A.; Taghipour, H.; Derrible, S.; Mohammadian, A. (Kouros) Toward Safer Highways, Application of XGBoost and SHAP for Real-Time Accident Detection and Feature Analysis. *Accid. Anal. Prev.* **2020**, *136*, 105405. [CrossRef]

69. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2018**, arXiv:1710.10903.

70. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2017**, arXiv:1706.03762.

71. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

72. Chollet, F. *Deep Learning with Python*, 2nd ed.; Simon and Schuster: New York, NY, USA, 2021; ISBN 978-1-63835-009-5.

73. Zhang, S.; Tong, H.; Xu, J.; Maciejewski, R. Graph Convolutional Networks: A Comprehensive Review. *Comput. Soc. Netw.* **2019**, *6*, 11. [CrossRef] [PubMed]

74. Wilson, T.; Tan, P.-N.; Luo, L. A Low Rank Weighted Graph Convolutional Approach to Weather Prediction. In Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 17–20 November 2018; pp. 627–636.

75. Hamilton, W.; Ying, Z.; Leskovec, J. Inductive Representation Learning on Large Graphs. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.

76. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 4–24. [CrossRef] [PubMed]

77. Labonne, M. *Hands-On Graph Neural Networks Using Python*; Packt: Birmingham, UK, 2023.

78. Maas, A.L.; Hannun, A.Y.; Ng, A.Y. Rectifier Nonlinearities Improve Neural Network Acoustic Models. In Proceedings of the ICML Workshop on Deep Learning for Audio, Speech and Language Processing, Atlanta, GA, USA, 16 June 2013.
79. Liew, S.S.; Khalil-Hani, M.; Bakhteri, R. Bounded Activation Functions for Enhanced Training Stability of Deep Neural Networks on Visual Pattern Recognition Problems. *Neurocomputing* **2016**, *216*, 718–734. [CrossRef]
80. Rudin, C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. [CrossRef] [PubMed]
81. McGovern, A.; Lagerquist, R.; John Gagne, D.; Jergensen, G.E.; Elmore, K.L.; Homeyer, C.R.; Smith, T. Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning. *Bull. Am. Meteorol. Soc.* **2019**, *100*, 2175–2199. [CrossRef]
82. Gu, Z.; Liu, Y.; Hughes, D.J.; Ye, J.; Hou, X. A Parametric Study of Adhesive Bonded Joints with Composite Material Using Black-Box and Grey-Box Machine Learning Methods: Deep Neuron Networks and Genetic Programming. *Compos. Part B Eng.* **2021**, *217*, 108894. [CrossRef]
83. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
84. Nevo, S.; Morin, E.; Gerzi Rosenthal, A.; Metzger, A.; Barshai, C.; Weitzner, D.; Voloshin, D.; Kratzert, F.; Elidan, G.; Dror, G.; et al. Flood Forecasting with Machine Learning Models in an Operational Framework. *Hydrol. Earth Syst. Sci.* **2022**, *26*, 4013–4032. [CrossRef]
85. Yang, D.; Kleissl, J.; Gueymard, C.A.; Pedro, H.T.C.; Coimbra, C.F.M. History and Trends in Solar Irradiance and PV Power Forecasting: A Preliminary Assessment and Review Using Text Mining. *Sol. Energy* **2018**, *168*, 60–101. [CrossRef]
86. Weisberg, S. *Applied Linear Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2005; ISBN 978-0-471-70408-9.
87. Ding, Y.; Zhu, Y.; Feng, J.; Zhang, P.; Cheng, Z. Interpretable Spatio-Temporal Attention LSTM Model for Flood Forecasting. *Neurocomputing* **2020**, *403*, 348–359. [CrossRef]
88. Dazzi, S.; Vacondio, R.; Mignosa, P. Flood Stage Forecasting Using Machine-Learning Methods: A Case Study on the Parma River (Italy). *Water* **2021**, *13*, 1612. [CrossRef]
89. Wang, H.; Hu, Y.; Guo, Y.; Wu, Z.; Yan, D. Urban Flood Forecasting Based on the Coupling of Numerical Weather Model and Stormwater Model: A Case Study of Zhengzhou City. *J. Hydrol. Reg. Stud.* **2022**, *39*, 100985. [CrossRef]
90. Ziyad Sami, B.F.; Latif, S.D.; Ahmed, A.N.; Chow, M.F.; Murti, M.A.; Suhendi, A.; Ziyad Sami, B.H.; Wong, J.K.; Birima, A.H.; El-Shafie, A. Machine Learning Algorithm as a Sustainable Tool for Dissolved Oxygen Prediction: A Case Study of Feitsui Reservoir, Taiwan. *Sci. Rep.* **2022**, *12*, 3649. [CrossRef]
91. Curbani, F.E.; Lacerda, K.C.; Curbani, F.; Barreto, F.T.C.; Tadokoro, C.E.; Chacaltana, J.T.A. Numerical Study of Physical and Biogeochemical Processes Controlling Dissolved Oxygen in an Urbanized Subtropical Estuary: Vitória Island Estuarine System, Brazil. *Environ. Model. Assess.* **2022**, *27*, 233–249. [CrossRef]
92. Kogekar, A.P.; Nayak, R.; Pati, U.C. Forecasting of Water Quality for the River Ganga Using Univariate Time-Series Models. In Proceedings of the 2021 8th International Conference on Smart Computing and Communications (ICSCC), Kochi, India, 1–3 July 2021; pp. 52–57.
93. Yaseen, Z.; Ehteram, M.; Sharafati, A.; Shahid, S.; Al-Ansari, N.; El-Shafie, A. The Integration of Nature-Inspired Algorithms with Least Square Support Vector Regression Models: Application to Modeling River Dissolved Oxygen Concentration. *Water* **2018**, *10*, 1124. [CrossRef]
94. Facebook's Core Data Science Team Forecasting at Scale. Available online: http://facebook.github.io/prophet/ (accessed on 20 September 2023).
95. Taylor, S.J.; Letham, B. *Forecasting at Scale*; PeerJ Preprints: London, UK, 2017.