

Article

Design of a Soft Sensor Based on Long Short-Term Memory Artificial Neural Network (LSTM) for Wastewater Treatment Plants

Roxana Recio-Colmenares ¹, Elizabeth León Becerril ^{1,*} , Kelly Joel Gurubel Tun ^{2,*}  and Robin F. Conchas ³

¹ Environmental Technology Department, Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco, A.C., Av. Normalistas 800, Colinas de la Normal, Guadalajara 44270, Jalisco, Mexico; roxana.recio@academicos.udg.mx

² School of Engineering and Technological Innovation, University of Guadalajara, Campus Tonalá, Tonalá 45425, Jalisco, Mexico

³ Electrical Engineering Department, Research Center and Advanced Studies of Instituto Politécnico Nacional (CINVESTAV), Unidad Guadalajara, Av. del Bosque 1145, El Bajío, Zapopan 45017, Jalisco, Mexico; robin.conchas@cinvestav.mx

* Correspondence: eleon@ciatej.mx (E.L.B.); joel.gurubel@academicos.udg.mx (K.J.G.T.)

Abstract: Assessment of wastewater effluent quality in terms of physicochemical and microbial parameters is a difficult task; therefore, an online method which combines the variables and represents a final value as the quality index could be used as a useful management tool for decision makers. However, conventional measurement methods often have limitations, such as time-consuming processes and high associated costs, which hinder efficient and practical monitoring. Therefore, this study presents an approach that underscores the importance of using both short- and long-term memory networks (LSTM) to enhance monitoring capabilities within wastewater treatment plants (WWTPs). The use of LSTM networks for soft sensor design is presented as a promising solution for accurate variable estimation to quantify effluent quality using the total chemical oxygen demand (TCOD) quality index. For the realization of this work, we first generated a dataset that describes the behavior of the activated sludge system in discrete time. Then, we developed a deep LSTM network structure as a basis for formulating the LSTM-based soft sensor model. The results demonstrate that this structure produces high-precision predictions for the concentrations of soluble X_1 and solid X_2 substrates in the wastewater treatment system. After hyperparameter optimization, the predictive capacity of the proposed model is optimized, with average values of performance metrics, mean square error (MSE), coefficient of determination (R^2), and mean absolute percentage error (MAPE), of 23.38, 0.97, and 1.31 for X_1 , and 9.74, 0.93, and 1.89 for X_2 , respectively. According to the results, the proposed LSTM-based soft sensor can be a valuable tool for determining effluent quality index in wastewater treatment systems.



Citation: Recio-Colmenares, R.; León Becerril, E.; Gurubel Tun, K.J.; Conchas, R.F. Design of a Soft Sensor Based on Long Short-Term Memory Artificial Neural Network (LSTM) for Wastewater Treatment Plants. *Sensors* **2023**, *23*, 9236. <https://doi.org/10.3390/s23229236>

Academic Editors: Chun-Lien Su, Chung-Hong Lee, Mahmoud Elsisí and Chien-Chang Chen

Received: 20 October 2023

Revised: 11 November 2023

Accepted: 13 November 2023

Published: 17 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: wastewater treatment; soft sensors; LSTM; quality index; organic substrates

1. Introduction

In nonlinear systems, such as biological ones, complex variables crucial for determining the quality of wastewater often prove challenging to measure in real time due to the presence of external disturbances and the nonlinear phenomena of these processes. Within this context, the importance lies in the design of digital sensors aimed at identifying variables hard to measure in biological processes, with a specific focus on wastewater treatment plants [1]. This approach plays an essential role in decision making for optimal operation of the process, offering practical and cost-effective alternatives to expensive or impractical conventional measurement devices. The implementation of these sensors not only brings economic benefits but also has a positive impact on the environment. In contrast to hardware sensors, digital detection techniques offer notable advantages, including delay-free estimation, low cost, simple maintenance, and high resistance to interferences [2]. Taking into account the modeling

methodologies, digital sensor models can be classified into three groups: first principles models, data-based models, and hybrid models. First principles models are complex and require significant computational resources, making data-based models the preferred option. The latter incorporates a variety of techniques such as support vector regression (SVR), artificial neural networks (ANN), Bayesian regression learning (BRL), gaussian process regression (GPR), kernel ridge regression (KRR), Kalman filters (KF), partial least squares regression (PLS), and ensemble learning [3–6]. While they have proven effective in modeling complex processes in digital sensors [7,8], the main challenge remains of the handling of unlabeled data and model generalization. Conventional modeling methods for digital sensors are not ideal for addressing large datasets, unlabeled data, and extensive industrial samples, making it difficult to obtain stable and reliable results [9]. In recent years, the widespread use of deep learning has been crucial in various fields, such as speech recognition, computer vision, natural language processing, and bioinformatics. Pretrained deep neural networks have proven to be a promising solution in extracting latent variables, significantly improving adaptability compared to traditional methods [10,11]. Furthermore, the computational efficiency of digital sensors is crucial for their successful implementation in industrial environments [12,13]. In this context, LSTM neural network models are a promising approach for time series forecasting and prediction compared to other deep neural network structures, since LSTMs are specifically designed to handle sequences of data, making them suitable for time series prediction. They can capture long-term dependencies in the data, allowing them to model relationships over extended time horizons, which is often a challenge for traditional feed-forward neural networks (FFNNs). LSTMs can process sequences of varying lengths, adapting to the specific context of each sequence, while some other deep neural network models require fixed-length input. In contrast to traditional recurrent neural networks (RNNs) which can face the vanishing gradient problem, making it difficult for them to capture long-term dependencies, LSTMs are designed to mitigate this issue through their gating mechanisms, allowing for more stable training and improved long-term performance [14]. Additionally, LSTMs can effectively handle noisy data and are robust to variations in data quality, making them suitable for real-world scenarios. LSTMs have been successfully applied to a wide range of time series forecasting tasks, including weather forecasting [14], financial forecasting [15], stock price and energy consumption predictions [16]. Their versatility and performance have made them a popular choice in these domains. The importance of using LSTM neural networks compared to traditional techniques lies in their ability to effectively model and predict relevant states in complex systems. By leveraging their ability to capture long-term dependencies in the data and handle sequences of varying lengths, LSTMs overcome the limitations of traditional techniques, especially in the context of time series prediction in bioprocesses. The use of LSTMs offers a promising and robust solution for online prediction variables, which has significant implications for improving efficiency and performance in a wide range of industrial and wastewater treatment applications.

In this work, a LSTM-based soft sensor approach to predict substrate concentrations for evaluating the effluent quality in wastewater treatment plants is proposed. The selection of deep LSTM network architecture and the configuration of hyperparameters results from a systematic exploration of parameter values, guided by empirical experimentation and prior research in the field. It represents a trade-off between the solution quality of prediction generated by the LSTM model and computational efficiency tailored to our specific problem context.

2. Materials and Methods

2.1. Wastewater Treatment Plant Description

The treatment process is realized in a real small-sized plant consisting of an aeration tank with 2000 m³ as the working volume, mechanical aerators which provide oxygen ($k_L a = 4.5 \text{ h}^{-1}$) and mix the incoming wastewater, and a settler for either solids to be recirculated to the aeration tank (D_r) or extracted from the system (ϵD). The influent average flow D is about 3000 m³/day, the average chemical oxygen demand input (COD_{in}) is 320 mg/L, and the total nitrogen input (TN_{in}) is 30 mg/L after pretreatment. The operational conditions used in this process are based

on those given by [17]. The treatment plant is schematically demonstrated in Figure 1. The Activated Sludge Model (ASM1) is used to describe the biochemical transformation processes in the suspended-growth treatment reactor for chemical oxygen demand (COD) removal [18]. A reduced model is represented by Equations (1)–(6), composed of ordinary differential equations and nonlinear kinetic functions which bear resemblance to those explored in the studies referenced in [17,18]. The characterization of wastewater and estimation of parameter values were made according to [18], and the reduced model was validated in a previous work [19]. The fitted model provides a satisfactory understanding of the transformation process leading to COD removal. In this work, the data needed to train and test the LSTM network architectures used in the soft sensor model were generated by simulating the ASM1 reduced model. The main objective of wastewater treatment plants is to improve the effluent quality. Therefore, we quantify effluent quality using TCOD as the performance index. For example, for urban wastewater, the maximum specified concentration of COD leaving a small-sized wastewater treatment plant is $COD_{max} = 150 \text{ mg/L}$ [19]. The TCOD is given by Equation (1), composed of the easily biodegradable soluble substrate $X_{1,k}$, the slowly decomposing solid substrate component $X_{2,k}$, and the inert organic material I_s . The latter reflects the constant value of the inflow.

$$TCOD = X_{1,k} + X_{2,k} + I_s \quad (1)$$

$$X_{1,k+1} = D(X_{1,in} - X_{1,k}) - \frac{1}{Y_H} \cdot \mu_1 \cdot \mu_{max,H} (\mu_3 + \mu_4 \cdot \mu_6 \cdot \eta_g) X_{3,k} + \mu_7 \cdot k_h (\mu_3 + \mu_4 \cdot \mu_6 \cdot \eta_h) X_{3,k} \quad (2)$$

$$X_{2,k+1} = D(X_{2,in} - X_{2,k}) + D_r(b-1)X_{2,k} + (1-f_p)(b_H \cdot X_{3,k} + b_A \cdot X_{B_A}) - \mu_7 \cdot k_h (\mu_3 + \mu_4 \cdot \mu_6 \cdot \eta_h) X_{3,k} \quad (3)$$

$$X_{3,k+1} = D(X_{3,in} - X_{3,k}) + D_r(b-1)X_{3,k} + \mu_1 \cdot \mu_3 \cdot \mu_{max,H} \cdot X_{3,k} + \mu_1 \cdot \mu_4 \cdot \mu_6 \cdot \mu_{max,H} \cdot \eta_g \cdot X_{3,k} - b_H \cdot X_{3,k} \quad (4)$$

$$X_{4,k+1} = D(X_{4,in} - X_{4,k}) + D_r(b-1)X_{4,k} + \mu_2 \cdot \mu_5 \cdot \mu_{max,A} \cdot X_{4,k} - b_A X_{4,k} \quad (5)$$

$$X_{5,k+1} = D(X_{5,in} - X_{5,k}) + KLA(X_{5,max} - X_{5,k}) - \left(\frac{1-Y_H}{Y_H}\right) \cdot \mu_1 \cdot \mu_3 \cdot \mu_{max,H} \cdot X_{3,k} - \left(\frac{4.57-Y_A}{Y_A}\right) \mu_2 \cdot \mu_5 \cdot \mu_{max,A} \cdot X_{4,k} \quad (6)$$

where $X_{3,k}$ is the active heterotrophic particulate biomass, $X_{4,k}$ is the active autotrophic particulate biomass, and $X_{5,k}$ is soluble oxygen. The kinetic and stoichiometric parameters are detailed in Appendix A, Tables A1 and A2, respectively.

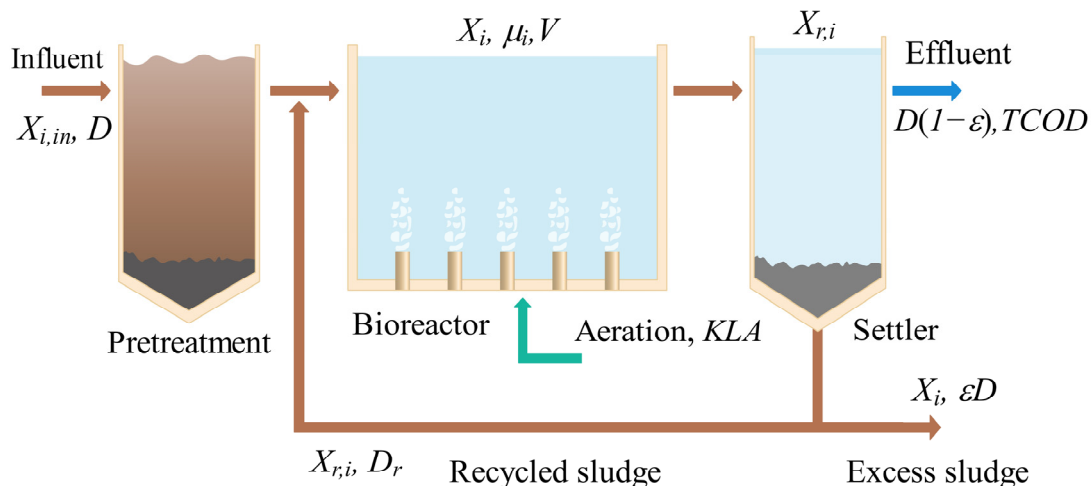


Figure 1. Wastewater plant configuration.

2.2. LSTM Network Architecture

The LSTM is a type of recurrent neural network initially introduced in the field of deep learning by Hochreiter and Schmidhuber [20] to address the issue of gradient explosion in RNNs during backpropagation. The LSTM model is widely recognized as an influential architecture for learning from sequential data due to its ability to capture long-term dependencies and effectively learn from sequences of varying lengths. A schematic of the LSTM model unit is presented in Figure 2.

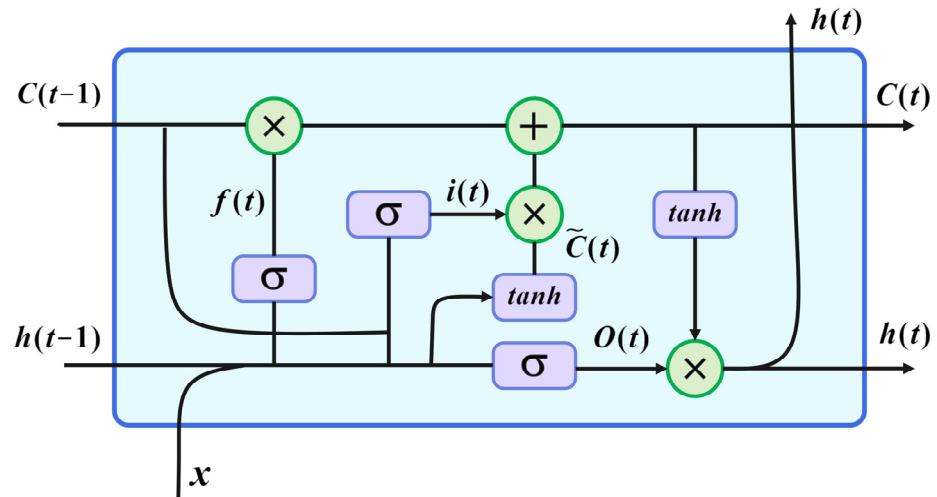


Figure 2. LSTM cell.

The LSTM unit comprises three gates responsible for controlling the flow of information: (i) the input gate, which determines the significance of input information to be remembered; (ii) the forget gate, which decides whether to retain or discard the input value; and (iii) the output gate, which governs the output of the LSTM unit. LSTM is implemented through Equations (7)–(12).

The input gate ($i(t)$):

$$i(t) = \sigma(W_{xi}x(t) + W_{hi}h(t-1) + W_{ci}C(t-1) + b_i) \quad (7)$$

The forget gate ($f(t)$):

$$f(t) = \sigma(W_{xf}x(t) + W_{hf}h(t-1) + W_{cf}C(t-1) + b_f) \quad (8)$$

The state candidates ($\tilde{C}(t)$):

$$\tilde{C}(t) = \tanh \tanh (W_{xc}x(t) + W_{hc}h(t-1) + b_c) \quad (9)$$

The activation cell ($C(t)$):

$$C(t) = f(t)C(t-1) + i(t)\tilde{C}(t) \quad (10)$$

The output gate ($o(t)$):

$$o(t) = \sigma(W_{xo}x(t) + W_{ho}h(t-1) + W_{co}C(t) + b_o) \quad (11)$$

The hidden state ($h(t)$):

$$h(t) = o(t)\tanh(C(t)) \quad (12)$$

Regarding the components entailed in the mathematical depiction of the LSTM cell, W_{ci} , W_{cf} , and W_{co} represent the weights establishing connections between the activation

cell and the input gate, the forget gate, and the output gate, respectively. W_{hi} , W_{hf} , W_{hc} , and W_{ho} denote the weights linking the hidden layer to the input gate, the forget gate, the activation cell, and the output gate [13]. Additionally, W_{xi} , W_{xf} , W_{xc} , and W_{xo} correspond to the weight matrices connecting the input layer to the input gate, whereas b_i , b_f , b_c , and b_o refer to the respective biases. Ultimately, the values are rescaled within the range of -1 to 1 using the $\tanh$ activation function.

2.3. LSTM-Based Soft Sensor Model

In recent years, soft sensors, which estimate process variables using measured data from other sensors, have become increasingly popular due to their ability to provide accurate and reliable predictions. In this context, ANNs have emerged as a prominent approach for developing soft sensors due to their ability to handle complex nonlinear relationships and their capability to learn from data [21,22]. In this work, a deep LSTM network is chosen for modeling the temporal behavior and dependencies between WWTP inputs and outputs due to its capability for time series prediction and handling time-dependent values [23,24]. Thus, the proposed LSTM-based soft sensor model is responsible for predicting the X_1 and X_2 states to quantify effluent quality using the TCOD as the quality index. As shown in Figure 3, the model operates in three stages:

- Data preprocessing: this step includes data normalization and implementing a sliding window into the dataset.
- Data processing: this step comprises the selection, training, and testing of the deep LSTM network to predict X_1 and X_2 .
- Data postprocessing: this step consists of the denormalization of data and the evaluation of the model's performance, resulting in the predictions of X_1 and X_2 , denoted as $\hat{X}_1$ and $\hat{X}_2$, respectively.

The X_3 , X_4 , and X_5 states are the input data measurements because of their role in the biotransformation of organic micropollutants (OMPs). Table 1 presents the input and output measurements of the proposed LSTM soft sensor.

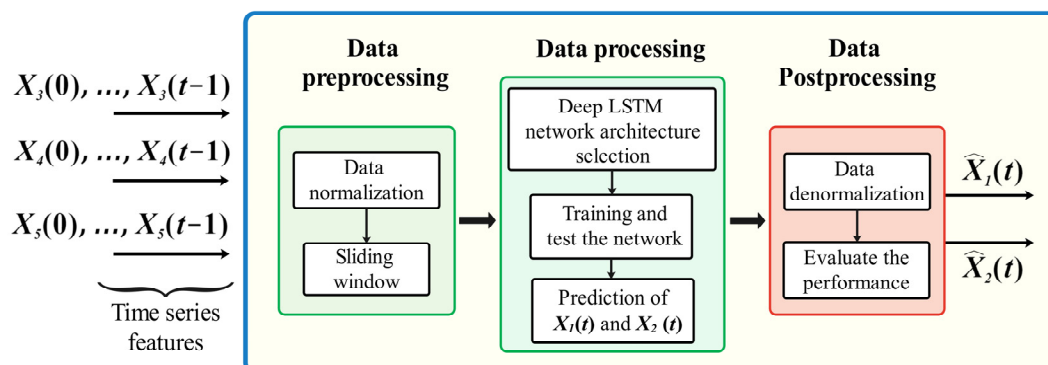


Figure 3. LSTM-based soft sensor.

Table 1. Input and output measurements of LSTM soft sensor.

Input Measurements	
Measurement	Description
$X_3(\text{mgCOD/L})$	Active heterotrophic particulate biomass
$X_4(\text{mgCOD/L})$	Active autotrophic particulate biomass
$X_5(\text{mg/L})$	Soluble oxygen
Output measurements	
$X_1(\text{mgCOD/L})$	Easily biodegradable soluble substrate
$X_2(\text{mgCOD/L})$	Slowly biodegradable particulate substrate

2.4. Dataset and Data Processing

Preparing data before feeding it into a model is a crucial step in machine learning techniques. LSTM networks require sufficient historical information to predict future outcomes and enhance system performance. In this study, wastewater dynamic states X_3 , X_4 , and X_5 are considered as input parameters. Input parameters are assumed to be available for data acquisition and they are directly related to the substrates degradation and oxidation, so they are suitable for the identification of organic substrates. Simultaneously, the output parameters X_1 and X_2 are predicted by the LSTM-based soft sensor model to determine the TCOD quality index for wastewater effluent assessment. Based on the simplified WWTP model described by Equations (1)–(6), a dataset comprising 5020 samples corresponding to 120 h of the process (5 days) was generated. The first 4500 rows corresponding to the first 108 h of the process were used for training and validating the LSTM networks. After adjusting the hyperparameters and attaining the optimal results, the remaining 520 sets of data (4501–5020) from the dataset, representing the final 12 h of the process, were used as unseen data to forecast the levels of X_1 and X_2 .

During the training phase, the model underwent supervised learning with predefined target outcomes. In the testing phase, the developed model was applied to predict the targeted substances based on the training data. Figure 4 visually demonstrates the 4500 data points generated for X_3 , X_4 , and X_5 . Statistics of parameters of the variables in the dataset generated experimentally by employing the model described by Equations (1)–(6) are presented in Table 2. It is important to note that for all kinds of data-driven models (e.g., artificial intelligence-based models), a low standard deviation of data indicates that the data points are closely clustered around the mean, which implies a smaller degree of variability or dispersion in the data; thus, it is expected to get less biased outputs from the models [25].

Table 2. Basic statistics of parameters in the generated dataset.

Parameters	Minimum	Maximum	Mean	Std. Deviation
X_1 (mgCOD/L)	189.13	393.60	290.82	31.77
X_2 (mgCOD/L)	66.06	145.53	107.80	11.91
X_3 (mgCOD/L)	23.10	57.54	40.13	5.84
X_4 (mgCOD/L)	0.70	2.80	1.73	0.35
X_5 (mg/L)	1.24	3.93	2.49	0.43

Studies have suggested that LSTM networks are responsive to dataset randomization, particularly when utilizing nonlinear activation functions. A widely adopted strategy to address this challenge is normalizing the dataset within the 0 to 1 range [24]. Consequently, we standardized both the input and target datasets using systematic weight initialization to expedite the learning process, leading to quicker convergence. The final normalized input data used for training the LSTM networks is illustrated in Figure 5.

2.5. Hyperparameter Selection for Proposed LSTM Architecture

The adequate selection of the deep LSTM network architecture, which is the core of the LSTM-based soft sensor model presented in Figure 6, involves utilizing various tools and methodologies. The optimal number of LSTM units in the hidden layer is determined through systematic experimentation, ranging from 2 to 200 cell units. Each topology is tested using a loss function as the error metric, with the process repeated thrice to ensure result consistency. After careful experimentation, the deep LSTM network architecture, depicted in Figure 6, exhibited the best training and validation accuracy results.

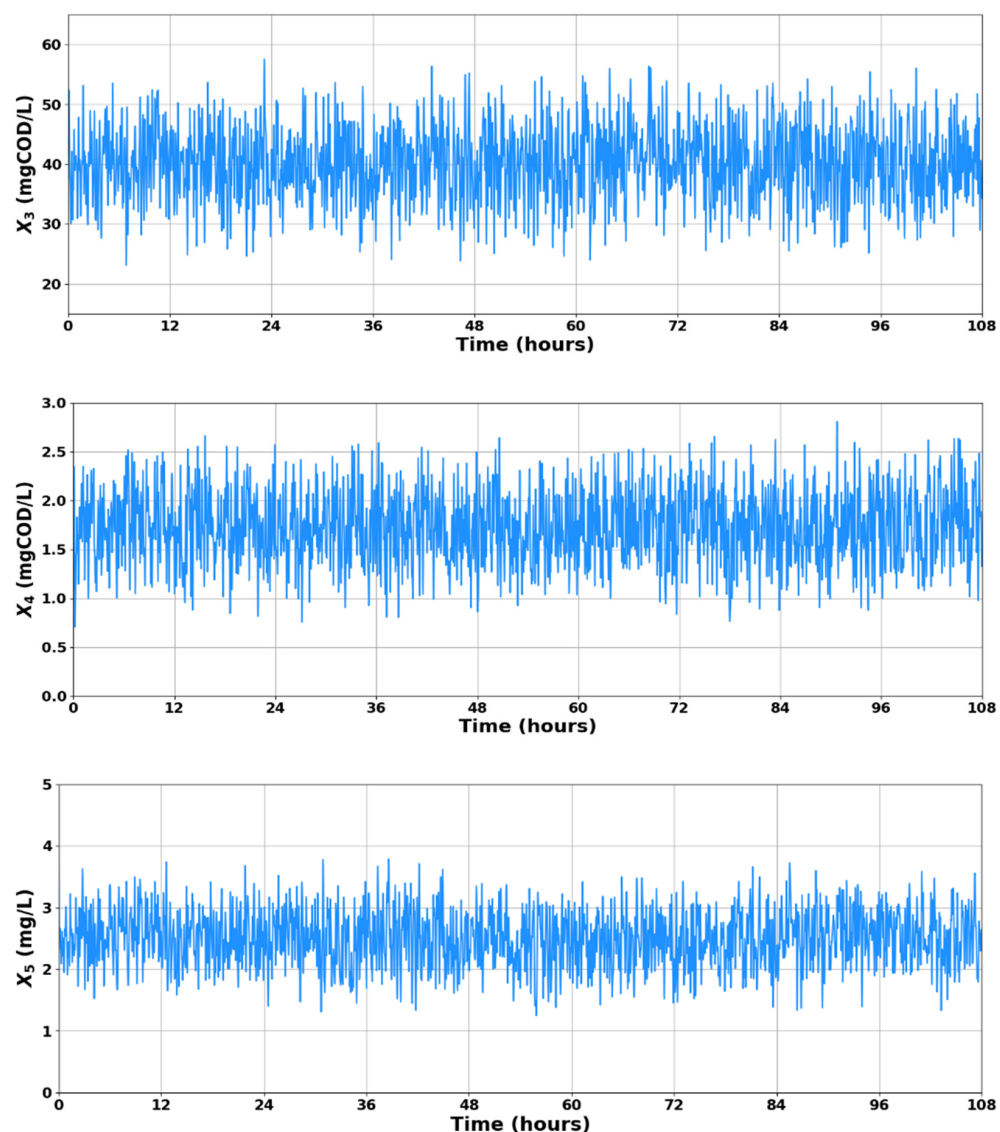
The training process involved using the seven most recent past measurements to perform the prediction of substrates. Notable minor hyperparameters of the selected configuration are presented in Table 3. The proposed LSTM network was implemented in Python 3.10.12 software, utilizing the Keras library with TensorFlow as its backend framework. Table 4 lists various available open-source libraries employed in this study.

Table 3. Hyperparameters selected for the deep LSTM architecture.

Hyperparameters	Selected Values
Batch size	128
Previous time steps	7
Optimizer	Adam
Epoch size	700
Dropout rate	0.1
Optimizer learning rate	0.001
Number LSTM layers	2
Number of LSTM cells per layer	L_1 128 and L_2 64
Activation function	RELU

Table 4. Libraries of Python 3.10.12 employed in this work.

Library	Purpose	Version
Numpy [25]	Data processing	1.23.5
Pandas [26]	Data management	1.5.3
Matplotlib [27]	Graphic Generation	3.7.1
Tensorflow [28]	Neural network implementation	2.13.0

**Figure 4.** Dataset of input measurements X_3 , X_4 , and X_5 .

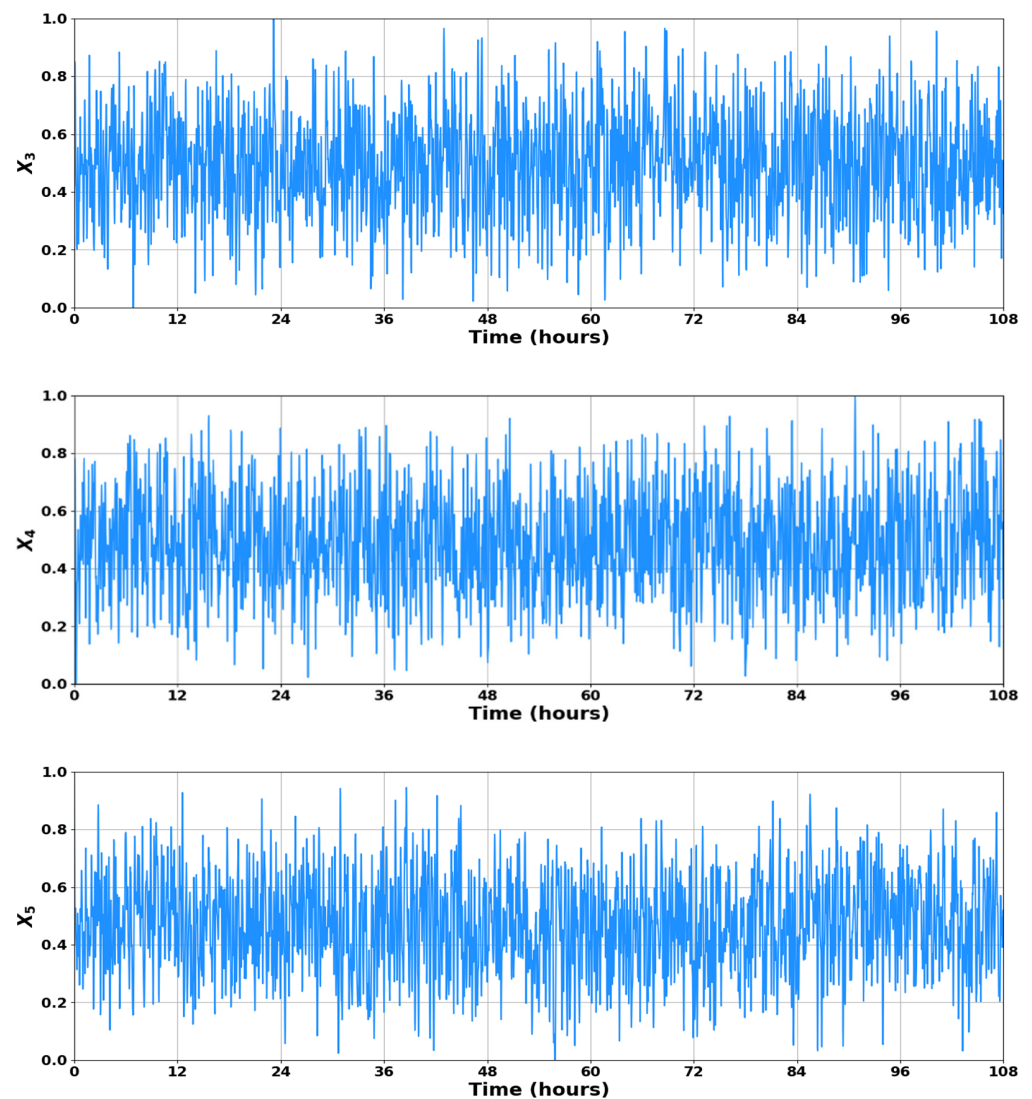


Figure 5. Normalized dataset of input measurements X_3 , X_4 , and X_5 .

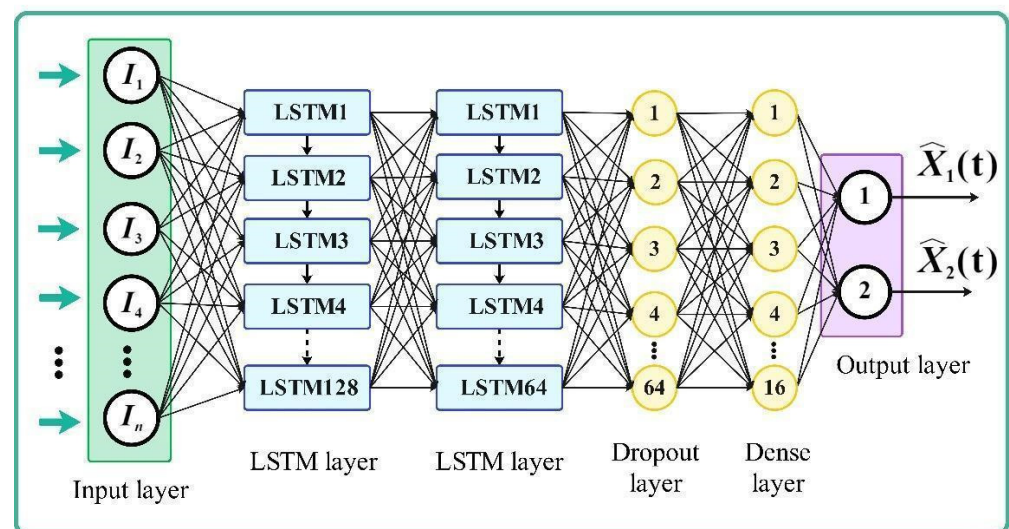


Figure 6. Proposed deep LSTM network architecture.

2.6. Model Performance Evaluation

The objective of model performance evaluation is to validate the accuracy of the proposed model and identify any errors, thus guaranteeing its reliable applicability [15]. In this study, we employ the MSE, R^2 , and MAPE as three performance metrics to evaluate the predictive capabilities of the proposed LSTM-based soft sensor model. The calculations for MSE, R^2 , and MAPE are as follows [29]:

1. MSE: it measures the average of the squares of the errors and is given by the following equation:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

2. R^2 : The coefficient of determination measures how much one variable can explain the variation in another variable when predicting the outcome of an event. The formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (14)$$

3. MAPE: it is the mean or average of the absolute percentage errors of prediction:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{\hat{y}_i} \right| \times 100 \quad (15)$$

Regarding Equations (13)–(15), n represents the number of samples, y_i corresponds to the i -th sample of the observed output data, $\hat{y}_i$ is the i -th predicted value, and $\bar{y}$ is the mean of the predicted values.

3. Prediction Results

3.1. Training and Validation Stage

Before running the proposed deep LSTM neural network for X_1 and X_2 prediction in the WWTP model described by Equations (1)–(6), the generated dataset comprising 5020 samples was divided into three groups, including data for training, data for validation, and data for testing, respectively. As mentioned in the previous section, a group of data comprised of 4500 rows corresponding to the first 108 h of the process was used for training and validating the LSTM networks, of which 80% was employed for training, and the remaining 20% of data was employed for validation. Figure 7 shows the training loss and validation loss curves in terms of MSE for the developed deep LSTM network using the hyperparameters presented in Table 3. From Figure 7, it is possible to observe that no overfitting occurs during the training and validation of the LSTM network stage.

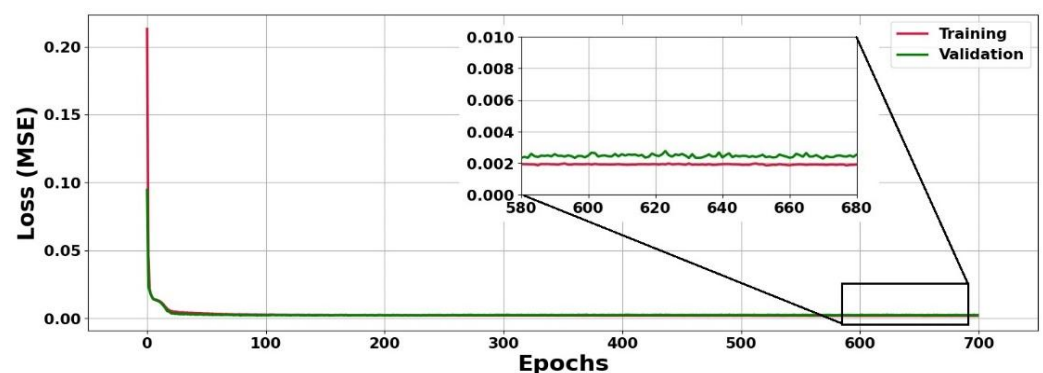


Figure 7. Loss curves of training and validation of proposed deep LSTM network.

3.2. Testing Stage

After training and validating the deep LSTM network, we applied the testing data containing 520 rows to evaluate the model's prediction reliability for unseen data during the training process. Figure 8 shows the prediction results for X_1 with the respective prediction error. The prediction results for X_2 are presented in Figure 9. The states of the system are available via the LSTM-based soft sensor model, and the TCOD quality index is calculated by Equation (1). Figure 10 displays the predicted TCOD vs the real value in the wastewater plant along 240 h. From these results, it can be appreciated that, in general, the predicted values were close to the observed values, indicating the adequate capability of the proposed strategy to predict the behavior of X_1 and X_2 for unseen data.

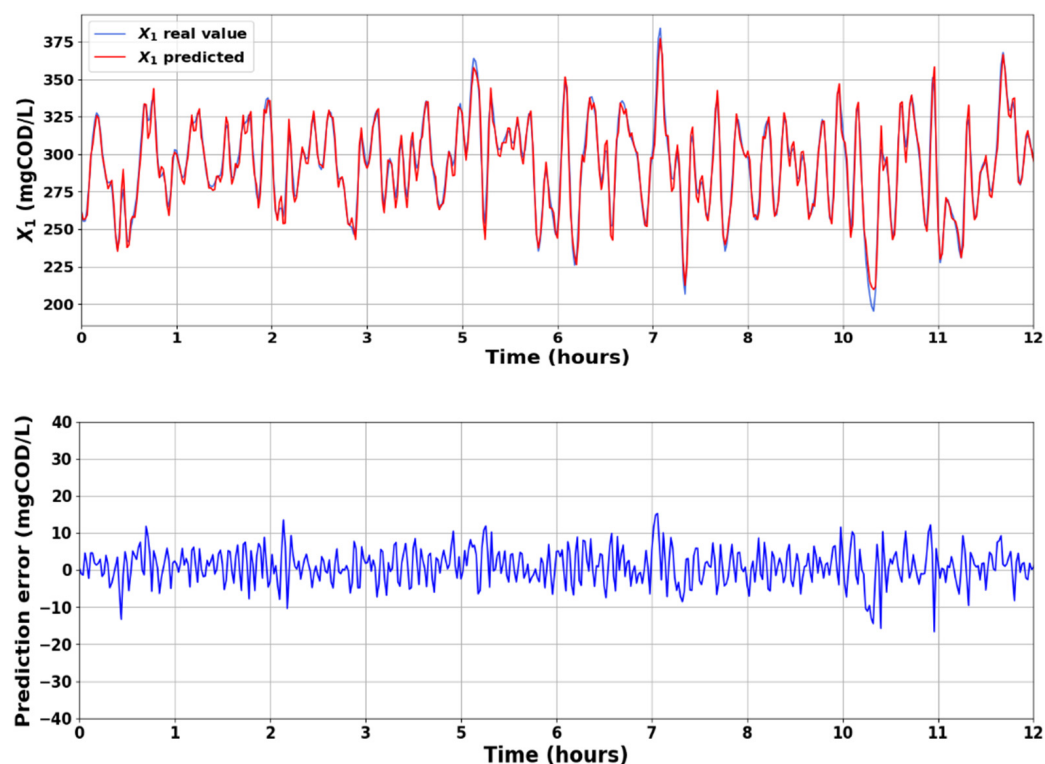


Figure 8. Observed and predicted results for X_1 .

The prediction results of the proposed model were evaluated based on the performance metrics MSE, R^2 , and MAPE presented in Equations (13)–(15). Generally speaking, a good fit between the observed and predicted results is obtaining values of MSE close to zero, while obtaining values of R^2 close to 1. Table 5 presents a comparative analysis in terms of the performance metrics for the prediction results achieved by employing the proposed LSTM-based model against those achieved using the FFNN [23] technique. The comparison was conducted using the same test dataset for both techniques. The FFNN was implemented with a single hidden layer comprising 128 neurons and trained using the Levenberg–Marquardt Algorithm (LMA).

Marquardt algorithm. The average MSE, R^2 , and MAPE values obtained were 23.38, 0.97, and 1.31 for X_1 , and 9.74, 0.93, and 1.89 for X_2 , respectively. The results indicate a superior performance of the proposed approach over the results obtained using the FFNN. Figures 11 and 12 show the scatter plot of the real versus predicted values of X_1 and X_2 for the LSTM-based model and FFNN technique, respectively.

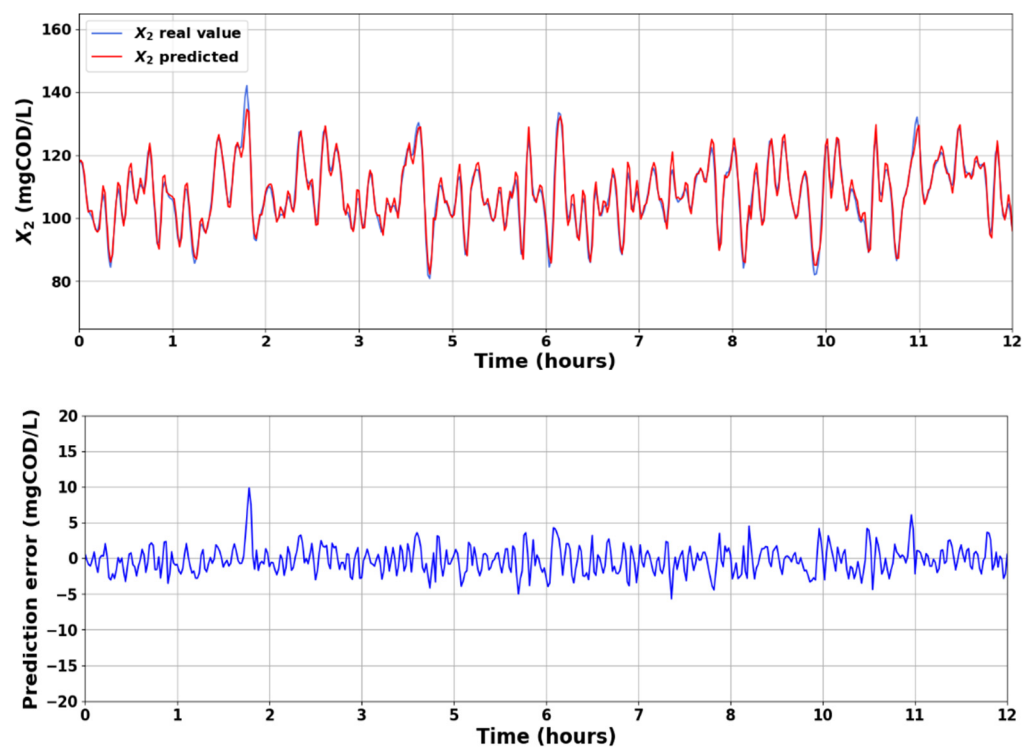


Figure 9. Observed and predicted results for X_2 .

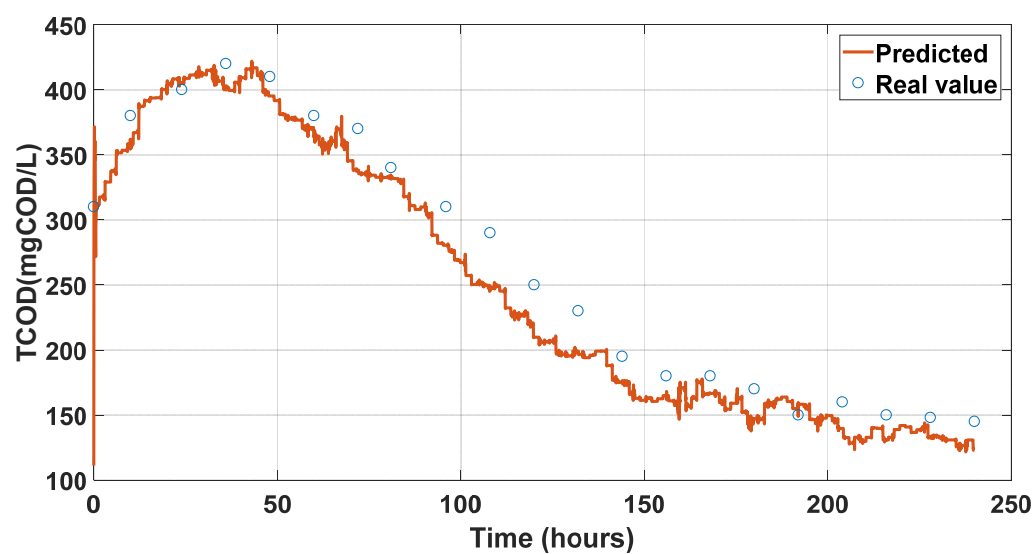


Figure 10. TCOD predicted vs. real values.

Table 5. Summary of prediction performance.

Method	Metric of Performance	Obtained Values for X_1	Obtained Values for X_2
Proposed LSTM-based approach	MSE	23.38	9.74
	R^2	0.97	0.93
	MAPE (%)	1.31	1.89
FFNN	MSE	115.52	13.70
	R^2	0.86	0.81
	MAPE (%)	3.34	2.76

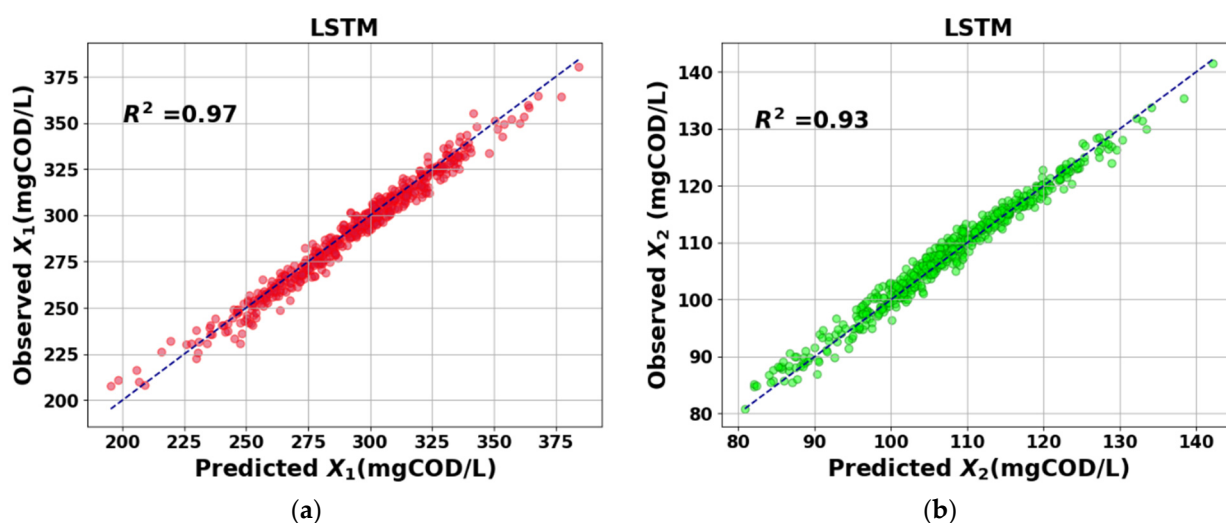


Figure 11. Plot of observed versus predicted values and the fitted regression line using the LSTM-based soft sensor model. (a) Red circles indicate X_1 positive correlation and (b) green circles indicate X_2 positive correlation.

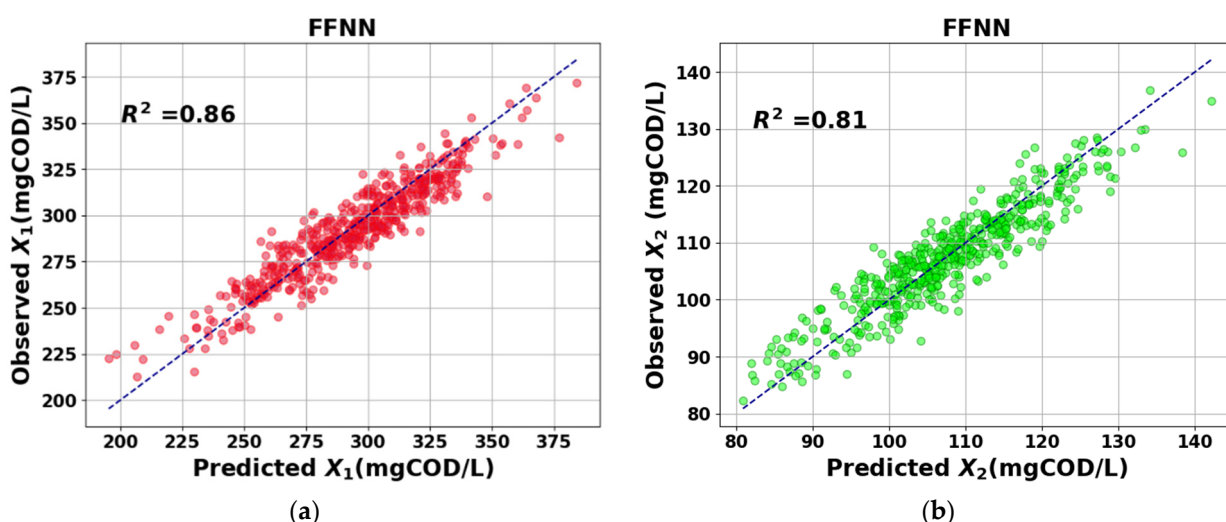


Figure 12. Plot of observed versus predicted values and the fitted regression line using the FFNN. (a) Red circles indicate X_1 positive correlation and (b) green circles indicate X_2 positive correlation.

4. Discussion

Based on the results presented in Table 5, it can be observed that the MSE for X_2 prediction was comparatively better than that for X_1 , which can be attributed to some large punctual deviations due to outliers in the dataset. Hence, some of the peaks or differences shown in the plots could be attributed to these outliers, leading to substantial deviations in the consecutive results. This complication could be addressed with a more extensive preprocessing process of the dataset. On the other hand, the prediction results in terms of performance metrics R^2 and MAPE were better for X_1 in contrast to the prediction results for X_2 . These obtained results can be attributed to the fact that according to the basic statics of variables in the dataset presented in Table 3, the easily biodegradable soluble substrate X_1 presents a lower value of standard deviation, which implies a low data variability and a more stable and predictable pattern, resulting in more accurate and less biased predictions generated by the model. In general, the results demonstrate that the proposed LSTM-based soft sensor model is competent in capturing the nonlinear behavior of substrates X_1 and X_2 present in the wastewater biological process for effluent quality evaluation.

5. Conclusions

This study proposes an LSTM-based soft sensor model to predict the concentrations of two critical substrates for effluent quality determination in wastewater treatment plants. First, we generated a dataset that describes the behavior of a real small-sized WWTP, modeled by the discrete-time ASM1. Then, we developed a deep LSTM network structure as the foundation for formulating the LSTM-based soft sensor model. The results demonstrate that this structure yields high-accuracy predictions for the organic substrates. After hyperparameter fine-tuning, the predictive capability of the proposed model was optimized, with average values of the performance metrics MSE, R^2 , and MAPE of 23.38, 0.97, and 1.31 for substrate X_1 , and 9.74, 0.93, and 1.89 for substrate X_2 , respectively. According to the results, the proposed LSTM-based soft sensor can be a valuable management tool for decision making, with the aim to satisfy legislative requirements. However, it is important to note that LSTM networks still present several challenging limitations. For instance, LSTMs are prone to overfitting when dealing with small datasets. Additionally, data preparation is critical for LSTM predictions. In most cases, it is necessary to normalize or standardize the data, handle missing values, and select appropriate features to ensure that the LSTM can effectively learn from the input. Therefore, at this time, the authors are actively exploring the development of methodologies for optimally selecting the most suitable LSTM network structure and the respective hyperparameters according to the particular application. Furthermore, as future work, it is intended to investigate the application of the proposed LSTM-based soft sensor to simulate a closed-loop wastewater treatment plant system.

Author Contributions: Methodology, R.R.-C. and R.F.C.; Validation, R.R.-C. and R.F.C.; Formal analysis, E.L.B. and K.J.G.T.; Resources, E.L.B.; Writing—review & editing, K.J.G.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT; Project CF-2023-G-648).

Data Availability Statement: Data are contained within the article.

Acknowledgments: Roxana Recio-Colmenares acknowledges CONAHCYT for the Postdoctoral Research Fellowships 2256881.

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

Appendix A

Table A1. Mathematical model parameters.

Parameter	Values	Units	Description
Y_A	0.24	mg COD/mg N	Autotrophic yield coefficient
Y_H	0.67	mg CDO/mg COD	Heterotrophic yield coefficient
$\mu_{max,A}$	0.8	1/h	Maximum specific growth rate for autotrophs
$\mu_{max,H}$	6	1/h	Maximum specific growth rate for heterotrophs
μ_1	—	—	Monod kinetics for easily biodegradable soluble substrate
μ_2	—	—	Monod kinetics for the component $X_{5,k}$ as a function of $X_{4,k}$
μ_3	—	—	Monod kinetics for the component $X_{5,k}$ as a function of $X_{3,k}$
μ_4	—	—	Monod kinetics for soluble nitrate and nitrite nitrogen
μ_5	—	—	Monod kinetics for soluble ammonium nitrogen
μ_6	—	—	Inhibition kinetics for $X_{5,k}$
μ_7	—	—	Saturation kinetics
b	0.9	—	Fraction coefficient for D_r
b_A	0.05	1/h	Autotrophic decay coefficient

Table A1. Cont.

Parameter	Values	Units	Description
b_H	0.22	1/h	Heterotrophic decay coefficient
f_P	0.08	—	Fraction of biomass yielding particulate products
η_g	0.8	—	Correction factor for anoxic growth of heterotrophs

Table A2. Initial conditions and additional parameters.

Parameter	Values	Units	Description
D	2	1/h	Dilution rate
D_r	1	1/h	Dilution recycle rate
KLA	—	1/h	Oxygen transfer coefficient
$X_{i,in}$	200	mgCOD/L	Initial condition of $X_{1,k}$
$X_{2,in}$	100	mgCOD/L	Initial condition of $X_{2,k}$
$X_{3,in}$	0	mgCOD/L	Initial condition of $X_{3,k}$
$X_{4,in}$	0	mgCOD/L	Initial condition of $X_{4,k}$
$X_{5,in}$	2	mgCOD/L	Initial condition of $X_{5,k}$
$X_{5,max}$	10	mg/L	Maximum concentration of soluble oxygen
V	15	L	Tank volume
I_s	5	mgCOD/L	Concentration of soluble and particulate inert organic matter
η_h	0.4	-	Correction factor for anoxic hydrolysis

References

- Kadlec, P.; Gabrys, B.; Strandt, S. Data-driven soft sensors in the process industry. *Comput. Chem. Eng.* **2009**, *33*, 795–814. [\[CrossRef\]](#)
- Alanis, A.Y.; Sanchez, E.N. Full Order Neural Observers. In *Discrete-Time Neural Observers*; Academic Press: Cambridge, MA, USA, 2017; pp. 23–74. [\[CrossRef\]](#)
- Hu, X.; Cao, D.; Egardt, B. Condition Monitoring in Advanced Battery Management Systems: Moving Horizon Estimation Using a Reduced Electrochemical Model. *IEEE/ASME Trans. Mechatron.* **2018**, *23*, 167–178. [\[CrossRef\]](#)
- Grbić, R.; Slišković, D.; Kadlec, P. Adaptive soft sensor for online prediction and process monitoring based on a mixture of Gaussian process models. *Comput. Chem. Eng.* **2013**, *58*, 84–97. [\[CrossRef\]](#)
- Lou, H.H.; Mukherjee, R.; Wang, Z.; Olsen, T.; Diwekar, U.; Lin, S. A New Area of Utilizing Industrial Internet of Things in Environmental Monitoring. *Front. Chem. Eng.* **2022**, *4*, 842514. [\[CrossRef\]](#)
- Zhang, Y.; Jin, H.; Liu, H.; Yang, B.; Dong, S. Deep Semi-Supervised Just-in-Time Learning Based Soft Sensor for Mooney Viscosity Estimation in Industrial Rubber Mixing Process. *Polymers* **2022**, *14*, 1018. [\[CrossRef\]](#)
- Wu, H.; Han, Y.; Jin, J.; Geng, Z. Novel Deep Learning Based on Data Fusion Integrating Correlation Analysis for Soft Sensor Modeling. *Ind. Eng. Chem. Res.* **2021**, *60*, 10001–10010. [\[CrossRef\]](#)
- Quan, J. Visualization and Analysis Model of Industrial Economy Status and Development Based on Knowledge Graph and Deep Neural Network. *Comput. Intell. Neurosci.* **2022**, *2022*, 7008093. [\[CrossRef\]](#)
- Yan, W.; Xu, R.; Wang, K.; Di, T.; Jiang, Z. Soft Sensor Modeling Method Based on Semisupervised Deep Learning and Its Application to Wastewater Treatment Plant. *Ind. Eng. Chem. Res.* **2020**, *59*, 4589–4601. [\[CrossRef\]](#)
- Li, Z.; Jin, H.; Dong, S.; Qian, B.; Yang, B.; Chen, X. Semi-supervised ensemble support vector regression based soft sensor for key quality variable estimation of nonlinear industrial processes with limited labeled data. *Chem. Eng. Res. Des.* **2022**, *179*, 510–526. [\[CrossRef\]](#)
- Hu, X.; Li, S.E.; Yang, Y. Advanced Machine Learning Approach for Lithium-Ion Battery State Estimation in Electric Vehicles. *IEEE Trans. Transp. Electr.* **2016**, *2*, 140–149. [\[CrossRef\]](#)
- Bakirov, R.; Gabrys, B.; Fay, D. Multiple adaptive mechanisms for data-driven soft sensors. *Comput. Chem. Eng.* **2017**, *96*, 42–54. [\[CrossRef\]](#)
- Gopakumar, V.; Tiwari, S.; Rahman, I. A deep learning based data driven soft sensor for bioprocesses. *Biochem. Eng. J.* **2018**, *136*, 28–39. [\[CrossRef\]](#)
- Venkatachalam, K.; Trojovský, P.; Pamucar, D.; Bacanin, N.; Simic, V. DWFH: An improved data-driven deep weather forecasting hybrid model using Transductive Long Short Term Memory (T-LSTM). *Expert Syst. Appl.* **2023**, *213*, 119270. [\[CrossRef\]](#)

15. Fang, Z.; Ma, X.; Pan, H.; Yang, G.; Arce, G.R. Movement forecasting of financial time series based on adaptive LSTM-BN network. *Expert Syst. Appl.* **2023**, *213*, 119207. [[CrossRef](#)]
16. Gülmez, B. Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm. *Expert Syst. Appl.* **2023**, *227*, 120346. [[CrossRef](#)]
17. Yoon, S.-H.; Lee, S. Critical operational parameters for zero sludge production in biological wastewater treatment processes combined with sludge disintegration. *Water Res.* **2005**, *39*, 3738–3754. [[CrossRef](#)] [[PubMed](#)]
18. Henze, M.; Grady, C.; Gujer, W.; Marais, G.; Matsuo, T. A general model for single-sludge wastewater treatment systems. *Water Res.* **1987**, *21*, 505–515. [[CrossRef](#)]
19. Recio-Colmenares, R.; Gurubel-Tun, K.J.; Zúñiga-Grajeda, V. Optimal neural tracking control with metaheuristic parameter identification for uncertain nonlinear systems with disturbances. *Appl. Sci.* **2020**, *10*, 7073. [[CrossRef](#)]
20. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
21. Pisa, I.; Santín, I.; Vicario, J.L.; Morell, A.; Vilanova, R. ANN-based soft sensor to predict effluent violations in wastewater 275 treatment plants. *Sensors* **2019**, *19*, 1280. [[CrossRef](#)]
22. Yaqub, M.; Asif, H.; Kim, S.; Lee, W. Modeling of a full-scale sewage treatment plant to predict the nutrient removal efficiency 279 using a long short-term memory (LSTM) neural network. *J. Water Process. Eng.* **2020**, *37*, 101388. [[CrossRef](#)]
23. Silva, I.N.; Spatti, D.H.; Flauzino, R.A.; Liboni, L.; Alves, S. *Artificial Neural Networks a Practical Course*; Springer: Cham, Switzerland, 2017.
24. Li, J.; Qi, C.; Li, Y.; Wu, Z. Prediction and compensation of contour error of cnc systems based on lstm neural-288 network. *IEEE/ASME Trans.-Actions Mechatron.* **2021**, *27*, 572–581. [[CrossRef](#)]
25. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [[CrossRef](#)] [[PubMed](#)]
26. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 51–56. [[CrossRef](#)]
27. Hunter, J.D. Matplotlib A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
28. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for large-scale machine learning. In Proceedings of the Symposium on Operating Systems Design and 301 Implementation, Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
29. Moriasi, D.N.; Arnold, J.G.; van Liew, M.W.; Bingner, R.L.; Harmel, R.D.; Veith, T.L. Model Evaluation Guidelines 293 for Systematic Quantification of Accuracy in Watershed Simulations. *Trans. ASABE* **1983**, *50*, 294.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.