

# Water Quality Prediction using LSTM: A Deep Learning Approach at Wat Makhm Station, Chao Phraya River, Thailand

*Nugroho Budi Wicaksono*<sup>1,2</sup>, *Sukma Meganova Effendi*<sup>1,3\*</sup>, *Dechrit Maneetham*<sup>1</sup>, and *Padma Nyoman Crisnapati*<sup>1</sup>

<sup>1</sup> Department of Mechatronics, Faculty of Technical Education, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand

<sup>2</sup> Department of Electromedical Technology, Faculty of Vocational Study, Sanata Dharma University, Indonesia

<sup>3</sup> Department of Mechatronics, Faculty of Vocational Study, Sanata Dharma University, Indonesia

**Abstract.** This study develops a Long Short-Term Memory (LSTM) neural network for forecasting water quality parameters at the Wat Makhm Station based on data collected from the Chao Phraya River, Thailand, for nine months. The study used IoT sensors to collect real-time values for ten water quality indicators: Turbidity (TURB\_NTU), Optical Dissolved Oxygen (HDO), Dissolved Oxygen Saturation (HDO\_SAT), Spatial Conductivity (SPCOND), Acidity/Basicity (pH), Total Dissolved Solids (TDS), Salinity (SALINITY), Temperature (TEMP), Chlorophyll (CHL), and Depth (DEPTH). The study identified water quality indicators through the implementation of an LSTM model following application of data cleansing techniques, using mainly the Interquartile Range (IQR) method for outlier detection. The results confirm that prediction accuracy varied across parameters. For stable indicators, very high prediction accuracy was achieved: for pH, MSE = 0.0064, MAPE = 0.89%, RMSE = 0.0800, and RMSPE = 1.12%; for salinity, MSE = 0.0006, MAPE = 10.55%, RMSE = 0.0246, and RMSPE = 41.14%. Temperatures were predicted with high confidence also: MAPE = 2.59% and RMSPE = 3.24%. In contrast, highly volatile parameters were difficult to predict; Turbidity MAPE = 32.87% and RMSPE = 109.22%; Chlorophyll MAPE = 38.64% and RMSPE = 190.15%.

## 1 Introduction

The Chao Phraya River stretches approximately 372 kilometers, beginning in the central plains of Thailand, flowing through the Bangkok metropolitan area, and finally reaching the Gulf of Thailand [1]. The river supports the ecological, social, and economic well-being of roughly 11.5 million people as a minimum. It is the main source of raw water for clean water production distributed to different areas under the responsibility of the Metropolitan Waterworks Authority (MWA) [2]. The issue of water quality and its long-term sustainability

---

\* Corresponding author : [sukma@usd.ac.id](mailto:sukma@usd.ac.id)

have grown more difficult, particularly in light of progress in industrial waste, agricultural activities, and city development.

All humans share a fundamental universal need: access to water, which is the most essential requirement for all living beings [3,4]. A report published by World Health Organization (WHO) points out that consumption of contaminated water causes the transmission of several diseases including diarrhea, cholera and typhoid, which are very fatal [3]. There are still about 2 billion people that could not access potable water, mostly in the developing world [5]. To achieve the United Nations Sustainable Development Goal 6 (SDG 6) of ensuring clean water and proper sanitation, it is essential that water quality evaluations are both efficient and effective in protecting public health and the environment. Conventional water quality assessment techniques are time-consuming and costly, also lack the capability for continuous or real-time data collection [6], which is essential for rapid detection and management of pollution incidents. To address previous issues, recent progress has focused on employing cutting-edge technologies like the Internet of Things (IoT) and machine learning (ML) to enhance the effectiveness of water quality monitoring systems [7]. IoT equipment's make it possible to collect pollution data continuously while prediction and change detection of pollution patterns are done by machine learning models in order to facilitate quick response and proactive water management strategies.

IoT devices combined with artificial intelligence (AI) have recently become significant in water quality monitoring. Deep learning approaches such as Long Short-Term Memory (LSTM) networks are highly effective in predicting water-quality parameters and reducing monitoring costs [9, 10]. Nguyen *et al.* [10] examined the predictive accuracy of ML models for indices of water quality in Vietnam and therefore recommended the further use of such techniques in the Southeast Asia region. Krohkaw *et al.* [2] employed an LSTM model for raw-water data from stations along the Chao Phraya River (Sam Lae, Rangsit Siphon, Wat Phai Lom, Wat Makham, Wat Pho Taeng Nuea, and Bangkhen Water Treatment Plant), whereas the validation only utilized data from the Sam Lae station. Other ML approaches, such as ANNs, have also been used successfully by Al-Adhaileh and Alsaade [11] for improvements of monitoring systems.

The purpose of the study is to develop and analyze predictive models for water quality monitoring of the Chao Phraya River using deep learning at Wat Makham Station utilizes the data given in [12]. Thus, while Krohkaew *et al.* [2] only evaluated LSTM performance at Sam Lae station, predictive studies for the other station remain unexplored. This study is conducted with support from the Metropolitan Waterworks Authority (MWA), which is responsible for managing drinking water supply for the city of Bangkok. Essentially, the investigation provides a comparatively low-cost method of assessing trends in water quality without the total reliance on laboratory testing. In particular, this study aims to explain the effectiveness of an LSTM-based model in the prediction of water quality parameters using time-series data. The insights given in this study provide insight for the applicability of a generalized framework, which can be extended to other monitoring stations for future studies. The summary of related research is presented in Table 1.

**Table 1.** Water quality monitoring-related research.

<b>Ref.</b>	<b>Contribution</b>	<b>Water Parameter</b>	<b>Country</b>	<b>Result</b>
Ref. [8]	Discusses computational approaches for estimating WQIs (Water Quality Indexes)	Cd, Cr, BOD <sub>5</sub> , DO, FC, FL, FOG, Hg, NH <sub>3</sub> , NO <sub>3</sub> , Pb, pH, TSS, SULF, TDS, TEMP, Zn	Mexico	Reduced costs in monitoring

Ref. [9]	Assesses water quality for agricultural and domestic use	DO, BOD, TCB, FCB, NH <sub>3</sub>	Thailand	Standard method for monitoring
Ref. [10]	Evaluates ML models for water quality classification	TEMP, pH, DO, BOD <sub>5</sub> , COD, N-NH <sub>4</sub> , N-NO <sub>3</sub> , N-NO <sub>2</sub> , P-PO <sub>4</sub> , FCB, TCB	Vietnam	Effective prediction of river water quality
Ref. [11]	Implementing ANNs to predict water quality changes	DO, pH, Conductivity, BOD, NO <sub>3</sub> , FCB, TCB	India	Enhanced monitoring efficiency

## 2 Method

### 2.1 Data acquisition

The dataset from [12] employed the Eureka Water Probe Manta+35 sensors, as shown in Figure 1, that were pre-deployed at Wat Makham station; enabling useful, quick, and real time studies of river water quality. Measurements of each Manta+35 were set for several important and representative water quality factors: Turbidity (TURB\_NTU), Optical Dissolved Oxygen (HDO), Dissolved Oxygen Saturation (HDO\_SAT), Spatial Conductivity (SPCOND), Acidity/Basicity (pH), Total Dissolved Solids (TDS), Salinity (SALINITY), Temperature (TEMP), Chlorophyll (CHL), and Depth (DEPTH). Utilizing IoT technology, these sensors allowed for the implementation of data collection in an unattended mode using buoys which made it possible to record water parameters every 10 minutes and upload them into the MySQL database. The entire process was conducted over nine-month period, from June 2022 to February 2023, and was authorized by local authorities.



**Fig. 1.** Manta+35 sensor probe: (a) side view, (b) bottom view, capable of holding up to 14 sensors.

### 2.2 Dataset preprocessing

Meticulous preparation was needed for deep learning applications to handle a very large and comprehensive water-quality dataset analyzed at Wat Makham Station with over 35,185 data points collected in nine months. This preparation began with a thorough statistical analysis of the dataset, as detailed in Table 2. Subsequently, we performed data-cleaning tasks: duplications and improper null value entries in the rows are addressed. Following this, outright outliers were identified and depending on their nature, were appropriately dealt with using the Interquartile Range method. The interquartile range method is robust for the detection of outliers in a dataset. This was accomplished by calculating the interquartile range, which were the range between the first quartile (25th percentile) of the data and the third quartile (75th percentile) of the data for a given parameter. If a data point falls below  $Q_1 - (1.5 \times IQR)$  or above  $Q_3 + (1.5 \times IQR)$  it is considered a likely outlier.

**Table 2.** Statistical data characteristics from data acquisition of Wat Makham Station.

	TURB_NTU	HDO	HDO_SAT	SPCOND	pH	TDS	SALINITY	TEMP	CHL	DEPTH
<b>count</b>	35,185	35,185	35,185	35,185	35,185	35,185	35,185	35,185	35,185	35,185
<b>mean</b>	30.7696	1.6181	21.4624	331.5826	7.2151	206.2113	0.1502	29.4965	15.3314	1.5790
<b>std</b>	58.958	0.8276	11.0702	1,840.6545	0.2624	55.6580	0.0416	1.8642	26.6046	0.4902
<b>min</b>	-2.460	0.01	0.2	0	-19.18	0	0.01	25.110	-0.03	0.15
<b>25%</b>	12.84	0.91	11.9	231	7.13	148.4	0.11	28.67	2.29	1.21
<b>50%</b>	19.81	1.63	21.8	326	7.22	209	0.15	30.09	3.82	1.45
<b>75%</b>	43.45	2.25	29.8	402	7.30	257.3	0.19	30.69	6.81	2.13
<b>max</b>	5,189	7.84	105.2	345,200	41.44	321.1	0.24	36.89	177.53	4.95

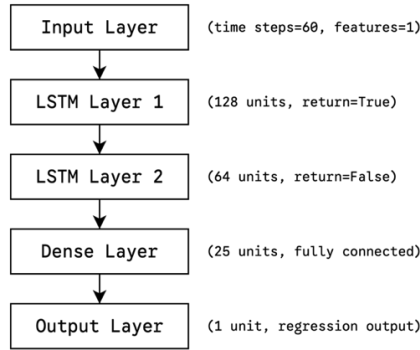
Since some variables (for example: turbidity, TURB\_NTU, or chlorophyll, CHL) might already exist with skewed distributions, the IQR calculations serve best as indicators for spotting truly exceptional data points. It is also necessary for the values of zeros to be removed from the dataset. This is because their presence can lead to significant differences in calculations for Mean Squared Error (MSE) and Root Mean Square Percentage Error (RMSPE); that is, calculations might include division by zero or infinity ( $\infty$ ) values, in turn, making it hard to derive meaningful measurements of error.

### 2.3 LSTM model development

LSTM is used to evaluate the time-series water quality dataset from Wat Makham Station. LSTM networks are a class of RNNs [13,14]. LSTM can learn to predict water quality parameters over a period of time because their architecture is designed to capture long-term dependencies, as well as ease in handling sequential data. One example of LSTM is vanilla LSTM which uses gates (forget gate, input gate, and output gate) to limit information flow through the data sequence. The gates are nothing more than very specific “neurons” that act as mini neural networks, monitoring what information comes into the system and controlling the flow of information inside the LSTM structure.

In this study, the data were represented in a format of NumPy array. This dataset was divided into a training set and a test set, utilizing 80 % for training. We used MinMaxScaler to normalize the data between the range [0, 1], to make it suitable for input in the LSTM model. At the training time, which sequences of 60-time steps are used as input features, and another step is used as the target variable. These sequences were then reshaped to align with the three-dimensional input form expected by the LSTM.

The model comprised two LSTM layers (128-unit and 64-unit) followed by two stacked dense layers for regression output as shown in the Figure 2. The model was compiled with the Adam optimizer and trained for five epochs using Mean Squared Error (MSE) loss, as it effectively reduces prediction error. The configuration was selected based on preliminary analyses and standard procedures in time-series forecasting. Having an 80:20 split ensures sufficient training data and retains some unseen samples for validation. Adam was chosen for being adaptive to the learning rate and efficient for use with recurrent networks. The model was trained for five epochs since the loss curves indicated convergence within this number of iterations, whereas further training caused overfitting in volatile parameters. The size of the batch was 32 because it gave a good balance between convergence speed and stability of the model. Following training, the testing dataset was scaled to facilitate predictions, which were subsequently inverse transformed to their original scale for evaluation. The LSTM model was implemented using Python in Google Colab (Keras/TensorFlow Sequential API), while the data preprocessing was done using scikit-learn-based tools and performance visualization was done in Matplotlib.



**Fig. 2.** Sequential LSTM model architecture used in this study.

Performance metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Percentage Error (RMSPE), were calculated to assess the model’s accuracy. The formulas for MSE, RMSE, MAPE, and RMSPE respectively use the formulas in (1), (2), (3) and (4).

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

$$RMSE = \sqrt{MSE} \tag{2}$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{3}$$

$$RMSPE = \sqrt{\frac{100}{n} \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{y_i} \right)^2} \tag{4}$$

### 3 Result and discussion

#### 3.1 Result of data preprocessing

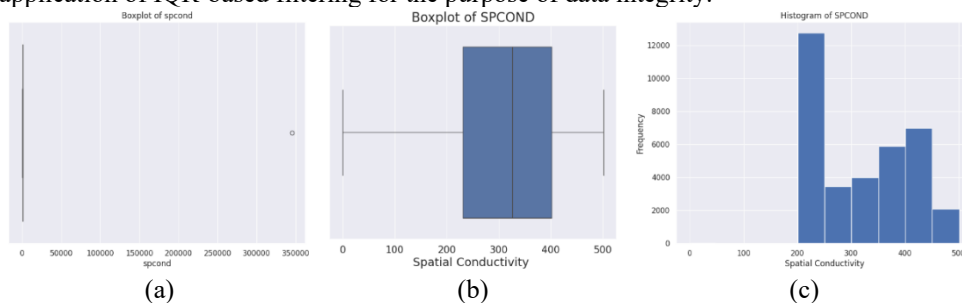
Applying the Interquartile Range (IQR) method for outlier elimination, a high ratio of data retention is achieved in multiple parameters as described in Table 3. Turbidity (TURB\_NTU) still had 98.97% of the original dataset retained after filtering with 34,821 data points. Similarly high retention rates were found for other parameters, *i.e.*, dissolved oxygen (HDO) and pH, with 99.93% and 99.94%, respectively. Chlorophyll (CHL) showed the lowest retention rate of 81.20%, asserting that more data for it were classified to be as outliers compared to other data in a set. This demonstrates the value of using IQR not only for detecting outliers but also to keep data meaningful, particularly when dealing with parameters that are expected to be skewed.

**Table 3.** The quantity of data remaining for each parameter after IQR removal.

Parameter	Data After IQR Removal	Percentage of Filtered Data (%)
TURB NTU	34,821	98.97
HDO	35,162	99.93

HDO SAT	35,161	99.93
SPCOND	35,184	99.99
pH	35,163	99.94
TDS	35,185	100.00
SALINITY	35,185	100.00
TEMP	34,971	99.39
CHL	28,570	81.20
DEPTH	35,184	100.00

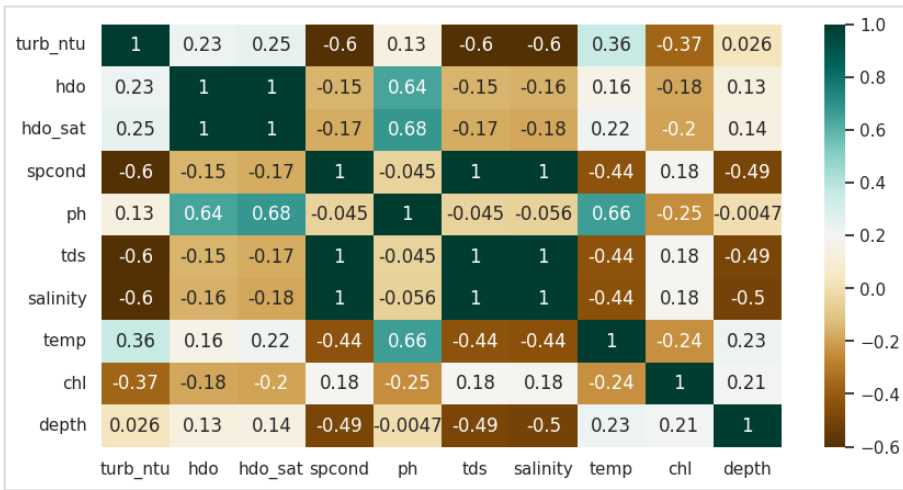
In this context, the focus has been placed on assessing Spatial Conductivity (SPCOND) because it determines the water's electrical conductivity and its relationship with ion levels. An isolated extreme anomaly was detected, prompting the need for a visual illustration owing to its large deviation from the norm. In Figure 3 panel (a) shows an outlier that skews the boxplot's scale, while panel (b), which uses the IQR method, shows the SPCOND range of about 200–400  $\mu\text{S}/\text{cm}$  more precisely. Just one data point was removed, reducing the dataset from 35,185 to 35,184, demonstrating that the outlier had little effect on the overall data pattern. The histogram in panel (c) reveals that the distribution after the outlier's removal mostly lies within the range of 200 to 300  $\mu\text{S}/\text{cm}$ , with a gradual fall off toward 500  $\mu\text{S}/\text{cm}$ . The existence of this one outlier implies that SPCOND is fairly consistent and justifies the application of IQR-based filtering for the purpose of data integrity.



**Fig. 3.** Data of Spatial Conductivity (SPCOND): (a) Outlier, (b) IQR method, (c) SPCOND after the outlier's removal.

### 3.2 Data analysis

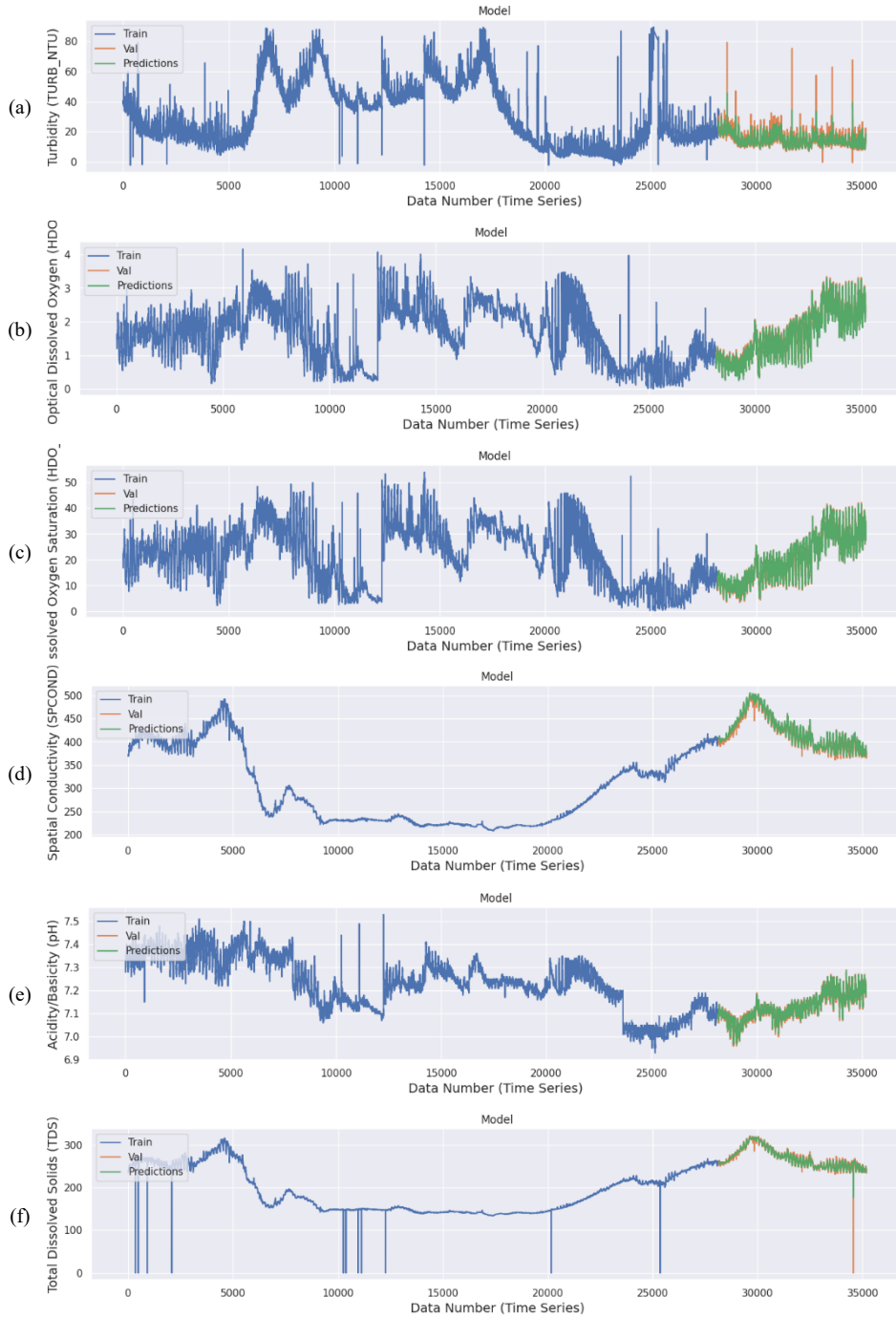
The correlation matrix displayed in Figure 4 highlights significant interactions among primary water quality parameters including Turbidity (TURB\_NTU), Optical Dissolved Oxygen (HDO), Dissolved Oxygen Saturation (HDO\_SAT), Spatial Conductivity (SPCOND), Acidity/Basicity (pH), Total Dissolved Solids (TDS), Salinity (SALINITY), Temperature (TEMP), Chlorophyll (CHL), and Depth (DEPTH). Interestingly, pH has a very strong positive correlation with HDO (0.64) and HDO\_SAT (0.68), which means more alkaline conditions are linked to higher dissolved oxygen and its saturation. In addition, Turbidity correlates negatively with Salinity (-0.6), SPCOND (-0.6), and TDS (-0.6), which suggests that the presence of suspended particles reduces the ionic or mineral content of the river. Meanwhile, Temperature has a positive correlation with pH (0.66) and a moderate one with Turbidity (0.36), which leads to the conjecture of a connection between high temperatures, alkaline pH, and increased turbidity.

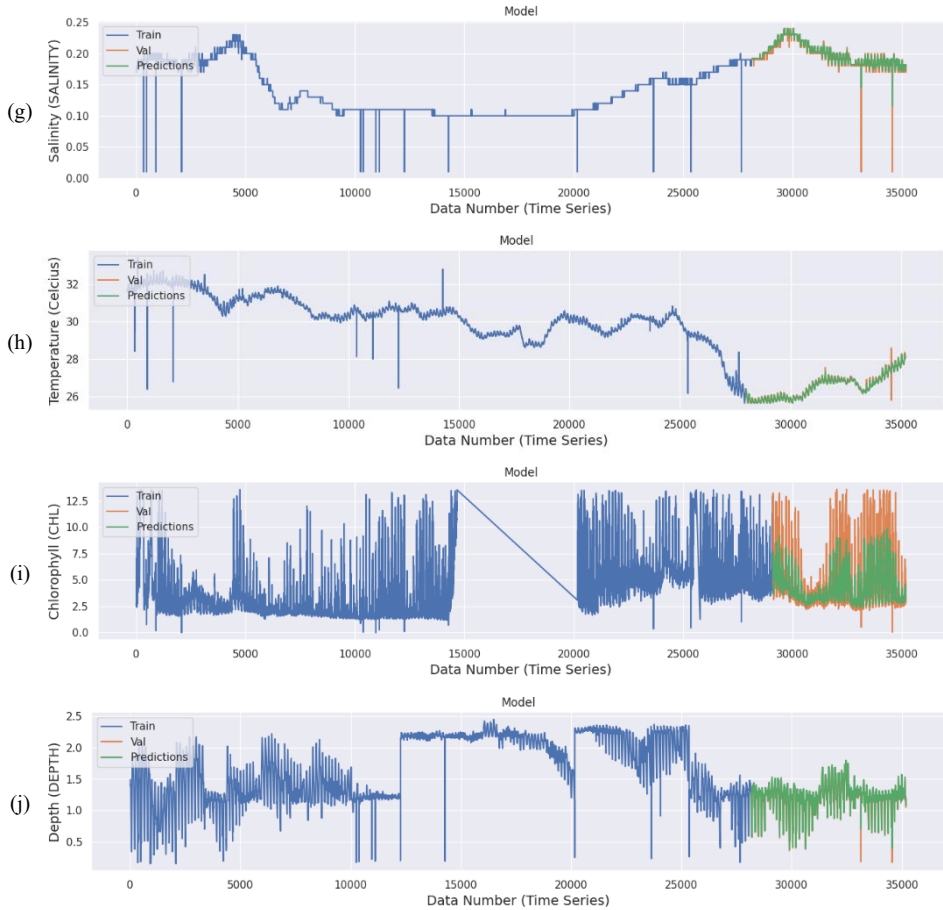


**Fig. 4.** Correlation matrix from 10 parameters of water quality.

SPCOND and Depth show a moderate negative correlation (-0.49), where it is suggested that the conductivity is greater in more shallow waters. Ions and dissolved minerals are likely to be found in higher concentrations in shallower regions due to evaporation, sediment contact, or even reduced mixing since the surface is quite close to the bottom. Additionally, Chlorophyll shows a negative correlation with Turbidity (-0.37) and Temperature (-0.24), which indicates that higher chlorophyll levels are associated with lower turbidity and cooler temperatures.

Figure 5 illustrates that the LSTM-generated prediction curves for water quality indicators display the training data in blue, the validation data in green, and the forecasted data in orange. It can be seen for turbidity (TURB\_NTU—panel (a)) that the LSTM model tries to fit the changes in the data but most of the peaks are smoothed out because these only occur in a short period of time. Changes in Optical Dissolved Oxygen (HDO—panel (b)) and Dissolved Oxygen Saturation (HDO\_SAT—panel (c)) show how trends are reasonably followed in case of stabilized conditions are maintained, although changes are not well adjusted either. Spatial conductivity (SPCOND) in panel (d) is able to mimic the data accurately especially for the smooth regions. For panel (e), Acidity/basicity (pH) deviations are quite low, suggesting consistent performance. For the panel (f), Total Dissolved Solids (TDS) align well with the broader trend but is greatly deviating at the peaks and troughs, indicating its inability to cope with abrupt changes. In the case of the salinity projections (panel (g)), is one of the best parameters. The behaviour of the model mimics almost exactly the validation series for the entire time span, which means the model can generalize well. Temperature (TEMP) in the panel (h) indicates that the general downward trend followed by a slow rise is well caught by the mode. For Chlorophyll (CHL)—panel (i), the model has difficulties with larger fluctuations and discrepancies become even obvious during extreme values. Lastly, Depth (panel j) predictions are well aligned with the actual data with very minor deviations along the path. To summarize, efficiency of the LSTM model was high for less changing parameters (salinity, pH, and temperature) for which gate mechanisms were able to manage long-term dependencies, and for highly dynamic parameters (*i.e.* turbidity and chlorophyll) for which memory gates due to additive noise were weak.





**Fig. 5.** Time-series predictions for 10 water quality parameters using an LSTM model, comparing the training data (in blue), validation data (in green), and predicted values (in orange). From (a) to (j), Turbidity (TURB\_NTU), Optical Dissolved Oxygen (HDO), Dissolved Oxygen Saturation (HDO\_SAT), Spatial Conductivity (SPCOND), Acidity/Basicity (pH), Total Dissolved Solids (TDS), Salinity (SALINITY), Temperature (TEMP), Chlorophyll (CHL), and Depth (DEPTH), respectively.

**Table 4.** Summary of MSE, MAPE, RMSE, and RMSPE from 10 parameters of water quality.

Parameter	MSE	MAPE	RMSE	RMSPE (%)
TURB_NTU	34.2443	32.8662	5.8519	109.2184
HDO	1.0385	65.9926	1.0191	100.5364
HDO_SAT	163.7166	67.4276	12.7952	103.9594
SPCOND	2532.6657	9.4454	50.3256	11.9373
pH	0.0064	0.8943	0.0800	1.1219
TDS	1034.5861	66.4641	32.1650	3588.0596
SALINITY	0.0006	10.5517	0.0246	41.1435
TEMP	0.7468	2.5944	0.8642	3.2416
CHL	5.2448	38.6380	2.2901	190.1467
DEPTH	0.0985	22.6641	0.3139	37.7471

Table 4 provides the overall MSE, MAPE, RMSE, and RMSPE for each of the 10 parameters (summary statistics). It supports the findings from the plots. Summary of the performance metrics for pH and Salinity shows very low error values across all metrics, with

an MSPE = 0.0064 and RMSPE = 1.1219% (for pH). This indicates the predictions are quite accurate. Temperature also has low error with MAPE (2.5944%) and RMSPE (3.2416%) values, which indicates the model would be able to predict steady state parameters very well.

The parameters Turbidity (TURB\_NTU), Optical Dissolved Oxygen (HDO), and Dissolved Oxygen Saturation (HDO\_SAT) showed significantly large errors, highlighting the model's challenge in accurately capturing sudden variations and outliers. For example, the parameters Turbidity (with a MAPE of 32.87% and RMSPE of 109.22%) and Chlorophyll (with a MAPE of 38.64% and RMSPE of 190.15%) point to the challenge of developing precise models for highly variable factors. Total Dissolved Solids (TDS) got the poorest performance (MSE = 1034.59, RMSPE = 3588.06%), which is a sign of very strong volatility and sensitivity to outliers. We should consider improving dynamic indicators by leveraging feature engineering, managing outliers appropriately, or developing more sophisticated models.

## 4 Conclusion

The application of the LSTM model for water quality prediction in the Chao Phraya River, especially in Wat Makhm station, gave an insight into what can and cannot be expected from deep learning in water quality monitoring. The model gave predictability with high accuracy for stable concentration parameters such as pH (RMSPE of 1.1219%) and temperature (MAPE of 2.5944%), thereby ensuring its suitability for giving credible values for constant water quality indices. However, the error has been found to be very large for those parameters that are volatile, especially Total Dissolved Solids (TDS) and turbidity, recording RMSPEs of 3588.0596% and 109.2184%, respectively, laying suggestions for future work in improving these time-varying water quality parameters. These findings suggest that while LSTM models hold much promise for water quality monitoring, future studies should focus on developing more robust methods for the prediction of parameters that change very fast, possibly by incorporating further feature engineering or more sophisticated model architectures. Despite such limitations, the study has been able to show the potentiality of the IoT with deep learning for water quality management systems, especially in considering water sources of high importance such as the Chao Phraya River.

The authors would like to acknowledge support from The United Board for Christian Higher Education in Asia.

## References

1. R. Chaysiri, J. Sukruay, Sustainability analysis for the lower Chao Phraya River using system dynamics. *J. Res. Unit Sci. Tech. Environ. Learning*. **12**, 105 (2021). <https://doi.org/10.14456/jstel.2021.8>
2. J. Krohkaew, P. Nilaphruek, N. Witthayawiroj, S. Uapipatanakul, Y. Thwe, P. N. Crisnapati, Thailand raw water quality dataset analysis and evaluation. *Data (Basel)* **8**, 141 (2023). <https://doi.org/10.3390/data8090141>
3. B. X. Lee, F. Kjaerulf, S. Turner, L. Cohen, P. D. Donnelly, R. Muggah, R. Davis, A. Realini, B. Kieselbach, L. S. MacGregor, I. Waller, R. Gordon, M. Moloney-Kitts, G. Lee, J. Gilligan, Transforming our world: Implementing the 2030 agenda through sustainable development goal indicators. *J Public Health Pol* **37**(Suppl1), 13-31 (2016). <https://doi.org/10.1057/s41271-016-0002-7>
4. S. N. Zainurin, W. Z. Wan Ismail, S. N. I. Mahamud, I. Ismail, J. Jamaludin, K. N. Z. Ariffin, W. M. Wan Ahmad Kamil, Advancements in monitoring water quality based on

- various sensing methods: a systematic review. *Int J Environ Res Public Health*. **19** (21), 14080 (2022). <https://doi.org/10.3390/ijerph192114080>
5. World Health Organization, Drinking-water. World health organization [online]. (2023). <https://www.who.int/news-room/fact-sheets/detail/drinking-water>
  6. P. Jayaraman, K. K. Nagarajan, P. Partheeban, and V. Krishnamurthy, Critical review on water quality analysis using IoT and machine learning models. *IJIM Data Insights*. **4**, 100210 (2024). <https://doi.org/10.1016/j.ijime.2023.100210>
  7. M. A. Rahu, A. F. Chandio, K. Aurangzeb, S. Karim, M. Alhusein, M. S. Anwar, Toward design of internet of things and machine learning-enabled frameworks for analysis and prediction of water quality. *IEEE Access*. vol. **11**, 101055 (2023). <https://doi.org/10.1109/ACCESS.2023.3315649>
  8. A. Fernández del Castillo, C. Yebra-Montes, M. Verduzco Garibay, J. de Anda, A. Garcia-Gonzalez, and M. S. Gradilla-Hernández, Simple prediction of an ecosystem-Specific Water Quality Index and the Water Quality Classification of a Highly Polluted River through Supervised Machine Learning. *Water*. **14**, 1235 (2022). <https://doi.org/10.3390/w14081235>
  9. S. Phitaktim, C. Wangka-orm, S. Wesetsri, T. Togari, S. Sato, M. Wongaree, Assessment of surface water quality of Huai Luang River in Udon Thani, Thailand, using the water quality index. *Int. J. Environ. Sci. Dev.* **14**, 381 (2023). <https://doi.org/10.18178/ijesd.2023.14.6.1458>
  10. D. P. Nguyen, H. D. Ha, N. T. Trinh, M. T. Nguyen, Application of artificial intelligence for forecasting surface quality index of irrigation systems in the Red River Delta, Vietnam. *Environ Syst Res*. **12**, 24 (2023). <https://doi.org/10.1186/s40068-023-00307-6>
  11. M. Hmoud Al-Adhaileh, F. Waselallah Alsaade, Modelling and prediction of water quality by using artificial intelligence. *Sustainability*. **13**, 4259 (2021). <https://doi.org/10.3390/su13084259>
  12. J. Krohkaew, P. Nilaphruek, N. Witthayawiroj, S. Uapipatanakul, Y. Thwe, P. N. Crisnapati, RAW-Q: Raw Water Quality Dataset, Thailand. *IEEE Dataport* (2023). <https://doi.org/10.21227/3q8d-jw96>
  13. S. Hochreiter, J. Schmidhuber, Long Short-Term Memory. *Neural Comput.* **9**, 1735 (1997). <https://doi.org/10.1162/neco.1997.9.8.1735>
  14. K. Greff, R. K. Srivastava, J. Koutnik, B. R. Steunebrink, J. Schmidhuber, LSTM: A search space Odyssey. *IEEE Trans Neural Netw Learn Syst.* **28**, 2222 (2017). <https://doi.org/10.1109/TNNLS.2016.2582924>