ICA2025OSLO
14th IWA Specialised Conference on
Instrumentation, Control and Automation

# Explainable AI for Aquatic Environmental Intelligence: a SHAP-Enhanced LSTM Approach Using High-Frequency Water Quality Data in a River System

Sait Mutlu Karahan[1,2], Jan Verwaeren[1], Janelcy Alferes[2], Wouter Vandenbruwaene[2]

[1]Department of Data Analysis and Mathematical Modelling, Ghent University, Belgium

[2]Vlaamse Instelling voor Technologisch Onderzoek (VITO), Mol, Belgium

(E-mail: saitmutlu.karahan@ugent.be)

(E-mail: jan.verwaeren@ugent.be)

(E-mail: janelcy.alferescastano@vito.be)

(E-mail: wouter.vandenbruwaene@vito.be)

**Abstract:** High-frequency IoT (Internet of Things) sensor data plays a crucial role in monitoring water quality, and electrical conductivity (EC) can serve as a key indicator of salinity levels in river systems. Especially, where water is withdrawn from the river to a reservoir for supply purposes, it is essential to understand the future status of the water quality. Therefore, employing machine learning forecasting techniques can offer valuable insights into the future conditions of water quality. However, utilizing these models, such as Long Short-Term Memory (LSTM) networks, challenges explaining the details of predictions due to their "black-box" nature. To address this, Explainable Artificial Intelligence (XAI) techniques, such as Shapley Additive Explanations (SHAP), is essential for enhancing model interpretability and ensuring that predictions align with internal and external processes. This study employs an LSTM model to forecast peak points (where the EC values reach a local maximum) in EC levels in river systems, for the summer season when salinity fluctuations are most pronounced. SHAP analysis is integrated to interpret the LSTM model's predictions, identifying the key factors driving salinization behaviour. Special attention is given to the salinity dynamics at both upstream and downstream regions of the river system, offering insights into the spatial variability and its underlying drivers. The results demonstrate the LSTM model's capacity to accurately predict EC readings while SHAP effectively elucidates the influence of features on these predictions. The study underscores the importance of combining high-frequency data, LSTM networks, and interpretability techniques to address critical environmental challenges.
**Keywords**: Explainable Artificial Intelligence; Shapley Additive Explanations; IoT sensor data; LSTM

## 1. Introduction

Water quality monitoring in a river system is a vital task for the protection of the ecosystem, ensuring the quality limitation of the water in the system, and supporting sustainable development in all aspects of life (Anh et al., 2023; Tickner et al., 2017). Numerous issues with water quality are initiated by both anthropogenic and natural forces, and the salinization problem in a river system is one of the compelling issues. Increased sea levels, low upstream discharges, surges from storms, impacts from human activities, and excessive groundwater withdrawals due to climate change are some of the triggers of the salinization problem in rivers (Mazhar et al., 2022). Therefore, it is essential to monitor and address the hotspots of salinization within a river system to mitigate the adverse effects associated with this challenge.

Effective, accurate, and on-time water quality assessments, particularly on salinization, are essential to investigate the potential reasons for this problem and predict future water quality trends (Park et al., 2020). Obtaining such assessments relies comprehensively on sensors, which play a crucial role due to their implementation across various fields. Their high-frequency nature enables

the immediate production of new information, and these measurements can provide valuable insights into such a complex and dynamic river system to identify and control salinization (Rode et al., 2016).

The vast and multidimensional datasets generated by modern sensor networks have become indispensable sources for the prediction of the future status of water quality conditions by using machine learning modelling approaches (Chen et al., 2020; Wai et al., 2022). From there, LSTM networks have found prevalent application in the domain of water quality prediction due to their capacity to capture temporal dependencies and manage sequential data despite their nonlinear behaviours, and LSTM models have proved their prediction accuracy and robustness (Gao & Li, 2014; Liang et al., 2020).

Besides the application fields and predictive capabilities, these machine-learning models are frequently criticized for their black-box character, which hinders their interpretability, and how predictions are derived (Hassija et al., 2024; Samek et al., 2017). In response to this struggle, techniques for improving model interpretability have drawn considerable attention. SHAP for instance is one of the tools for revealing the contributions of each feature and assess which features have the most importance on the model's result (Lundberg & Lee, 2017). By clarifying the fundamental processes of machine learning models, Shapley Additive Explanations (SHAP) improves confidence and accessibility in essential applications such as water quality control (Hellen & Marvin, 2022).

This research simulates the conductivity measurements in a river system by applying an LSTM model. Simulations are done for both the upstream and downstream part of the river, since both parts have distinct salinity dynamics. To evaluate the contribution of the internal (EC) and external (precipitation, discharge, temperature, etc.) features of the LSTM model, the SHAP technique is applied to find the contributions of the different features to the prediction of the peak points during a summer period. By using the SHAP, the interpretability of the LSTM network is improved, and in this way makes the methodology more accessible to waterway managers.

## 2. Study Site and Sensor Data

In this research, the Yser River is selected as the study site. It originates in northern France, runs into the West Flanders region of Belgium, and discharges into the North Sea. The river is predominantly contained inside Belgium's boundaries (44 km). The Yser River is a form of a rain-fed river system, where its average discharge is around 3.33 $m^3/s$ and peak flows reach up to 100 $m^3/s$ during extreme events. Salinization has been identified as one of the primary environmental challenges. Due to tidal movements originating from the North Sea. As a result of these tidal movements, diverse salinization patterns are observed in the upstream and downstream sections of the river.

Specifically, this study utilized data from 14 sensors, as shown in Figure 2.1, and the input variables have been categorized into internal and external features to reflect their origin and functional role within the river system. The internal features are represented by electrical conductivity measurements, which are directly obtained from sensor locations within the river and are considered intrinsic to the system's salinity dynamics. In contrast, the external features contain the remaining features where they exert significant influence on the salinization dynamics. Six sensor locations were selected for EC measurements, with three representing upstream and three representing downstream regions to capture salinity dynamics. For each group, the EC value of the centrally located sensor was estimated separately, based on internal and external features. For clarity, EC sensor locations are indicated by letters, and their positions from upstream to downstream are specified by numbers, and the letters "U" and "D" stand for "upstream" and "downstream," respectively. The remaining eight sensors were designated to capture external features, including temperature, wind speed, upstream and downstream water levels, precipitation, and discharge.

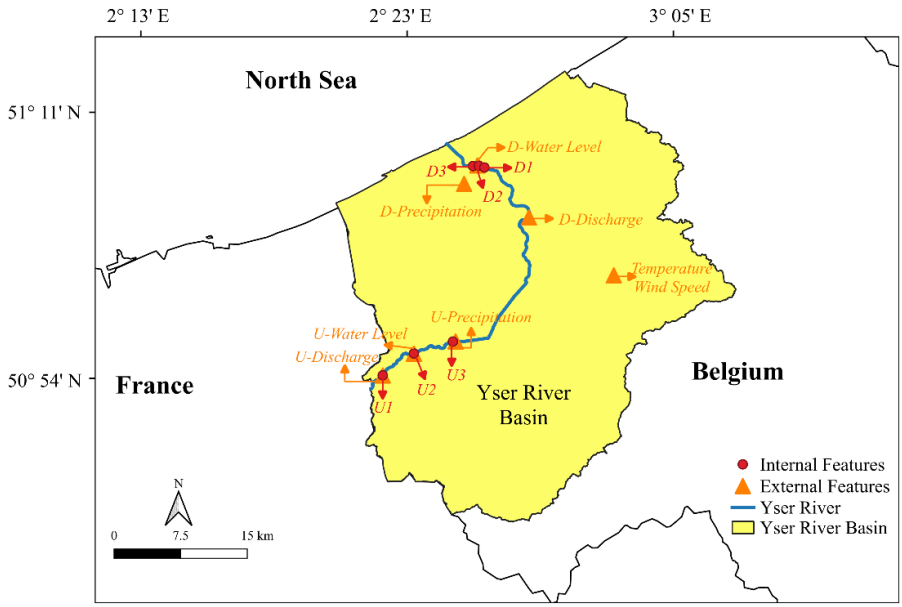


**Figure 2.1** Study Region and Sensor Locations (U: Upstream, D: Downstream)

## 3. Materials and Methods

### 3.1. Outlier Detection and Determination of Peak Points

The datasets employed in this study were subjected to preprocessing procedures prior to the implementation of the LSTM model. The data consist of measurements recorded at 15-minute intervals, covering the period from 20th September 2022 to 13th October 2023. As a first step, an outlier identification technique was used on the EC measurements to reduce the possibility of inaccuracy from sensor data. The following techniques were used to identify outliers: the exponentially weighted moving average (EWMA), z-score, and interquartile range (IQR). Measurements with a z-score outside the range of [-3, 3] were identified as outliers. For the EWMA method, one-day moving average windows were applied, and values exceeding three standard deviations from the window's mean were classified as outliers and subsequently interpolated using linear interpolation. For linear interpolation, a constraint was applied whereby data gaps exceeding one hour were not considered eligible for interpolation. Following these procedures, the statistical characteristics of the sensor data were presented in Table 3.1.1.
From the data presented in Table 3.1.1, two key observations can be noted. First, higher electrical conductivity (EC) values were recorded at downstream sensor locations, which can be attributed to saltwater intrusion from the North Sea. Second, downstream discharge values were represented as negative, indicating reverse flow conditions whereby saltwater advances inland along the river.

The peak points were identified for the central-located EC sensor from the groups of three, as the contribution of each feature to the prediction of these points was intended to be investigated. To be able to obtain those peak points daily changes are first calculated to capture short-term fluctuations for the summer period. Following this, the average magnitude of these daily changes is computed to represent the typical variability within the data. Based on this average, thresholds of 0.1 for U2 and 0.9 for D2 were established to distinguish normal fluctuations from significant deviations. Changes that exceed this defined threshold are subsequently identified and labelled as peak points, indicating moments of unusually high activity or variation within the dataset. Consequently, prominent peak points in the target time series were identified based on this threshold, as illustrated in Figure 3.1.

**Table 3.1.1** Statistical Characteristics of the Input Data (U: Upstream, D: Downstream)

| Feature Types | Sensor Name | Unit | Mean | Min | Max | Standard Deviation |
|---|---|---|---|---|---|---|
| **Internal Features** | U1 | ms/cm | 0.677 | 0.238 | 1.077 | 0.153 |
| | U2 | | 0.772 | 0.231 | 1.472 | 0.242 |
| | U3 | | 0.861 | 0.224 | 2.919 | 0.495 |
| | D1 | | 2.597 | 0.372 | 9.186 | 1.606 |
| | D2 | | 3.037 | 0.313 | 11.696 | 2.214 |
| | D3 | | 3.170 | 0.003 | 11.489 | 2.022 |
| **External Features** | U-Precipitation | mm | 0.007 | 0 | 5.440 | 0.057 |
| | U-Discharge | $m^3/s$ | 2.166 | 0.020 | 59.550 | 5.346 |
| | U-Water Level | m | 3.225 | 2.940 | 5.080 | 0.265 |
| | Temperature | °C | 12.673 | -5.530 | 32.630 | 6.661 |
| | Wind Speed | m/s | 2.825 | 0 | 14.760 | 2.382 |
| | D-Precipitation | mm | 0.023 | 0 | 8.310 | 0.154 |
| | D-Discharge | $m^3/s$ | 5.523 | -9.400 | 57.500 | 10.361 |
| | D-Water Level | m | 3.062 | 1.990 | 3.780 | 0.164 |

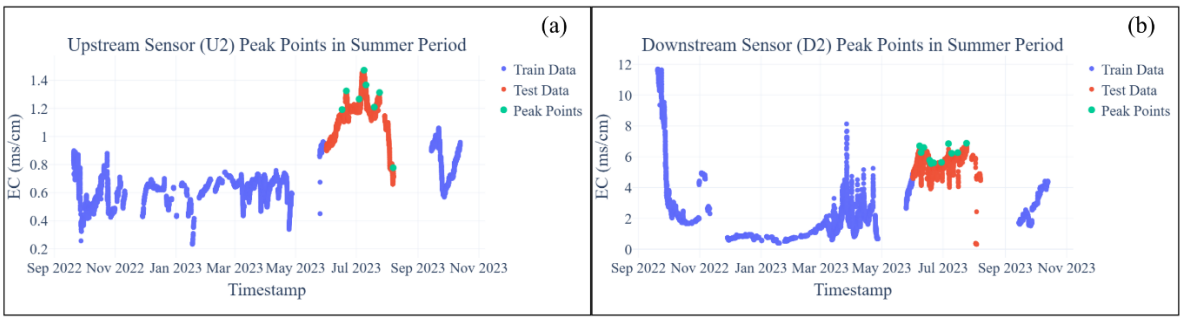


**Figure 3.1** Peak Points of Sensors during the summer period: (a) Upstream, (b) Downstream

### 3.2. LSTM Model and Hyperparameter Tuning

The LSTM networks are a particular kind of recurrent neural network (RNN) that is intended to overcome the drawbacks of avoiding long-term dependence issues in RNN (Hochreiter & Schmidhuber, 1997). Due to their highly parameterized nature, LSTM networks are capable of capturing temporal dependencies and uncovering the nonlinear dynamics inherent in time series data. As previously stated, the objective of forecasting is to predict the values at the centrally located sensor, which is excluded from the input data and used solely as the target variable. Accordingly, as shown in Table 3.2.1, two separate models were developed based on the spatial grouping of EC sensors into upstream and downstream regions.

**Table 3.2.1** Models and their input data (U: Upstream, D: Downstream)

| | **Input Data** | |
|---|---|---|
| **Models** | Internal Features | External Features |
| **U2** | U1, U3 | Temperature, Wind Speed, U-Water Level, U-Precipitation, U-Discharge |
| **D2** | D1, D3 | Temperature, Wind Speed, D-Water Level, D-Precipitation, D-Discharge |

In this study, two separate input layers were employed in the LSTM model based on the categorization of input features. The first input layer was used to process internal features, while the second was chosen for external features. In addition to temperature and wind speed—common external variables—location-specific features were incorporated separately for upstream and

downstream regions as given in Table 3.2.1. Each input layer was connected to a separate LSTM layer to process its respective feature set. The outputs of these LSTM layers were subsequently merged through a concatenation layer to integrate the internal and external information. A dense output layer was then applied to generate predictions at the identified peak points, as illustrated in Figure 3.2.1.

To forecast peak points for sensors U2 and D2, two days of historical data were utilized to predict the entire summer period at 15-minute intervals. Given the spatial variability in salinization dynamics across the river system, two separate hyperparameter optimization processes were conducted by using the Optuna optimization framework (Akiba et al., 2019). Optuna was employed to minimize the mean squared error (MSE) during the cross-validation process, and time series split method with five folds was applied to generate sequential training and validation datasets (Pedregosa et al., 2011). For each model, 30 optimization trials were conducted. An early stopping mechanism was incorporated, whereby the optimization process was terminated if the validation loss failed to decrease after five consecutive steps. The hyperparameters and their respective optimization ranges are presented in Table 3.2.2. From the optimization trials, the optimal hyperparameters were identified and subsequently refined according to sensor location, as shown in the Model-U2 and Model-D2 columns of the same table.

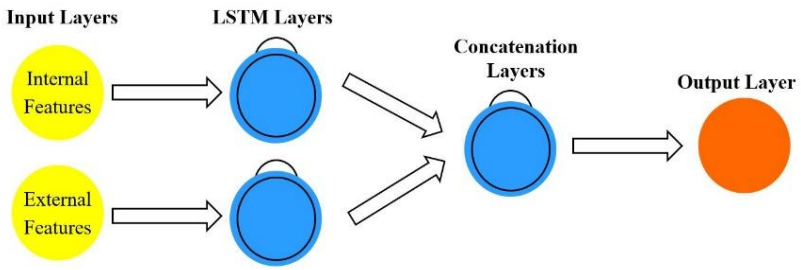


**Figure 3.2.1** The representative configuration of LSTM Model

**Table 3.2.2.** The Value Ranges of Hyperparameters and the Optimal Model Hyperparameters

| Hyperparameters | Value Ranges | Model-U2 | Model-D2 |
|---|---|---|---|
| **number of units** | {1, 2, 3, ..., 100} | 81 | 99 |
| **dropout rate** | [0, 0.5] | 0.18 | 0.36 |
| **learning rate** | [0.00001, 0.01] | 0.0008 | 0.0006 |
| **epochs** | {10, 11, 12, ... , 100} | 98 | 39 |
| **batch size** | {16, 32, 64, 128, 256, 512} | 256 | 16 |
| **activation** | {'linear', 'tanh', 'relu'} | relu | tanh |
| **kernel initializer** | {'glorot_uniform', 'glorot_normal', 'he_normal'} | he_normal | glorot_uniform |

### 3.3. Model interpretability with SHAP

Despite the high predictive accuracy of machine learning models, the interpretability of these models remains a critical requirement. In this study, SHAP, a widely recognized explainable artificial intelligence technique, was employed. Based on cooperative game theory, SHAP provides a consistent and theoretically grounded framework for quantifying the contribution of each feature to the model's predictions (Lundberg et al., 2018). This approach enhances the transparency and reliability of the predictive model by offering insights into the relationships between input variables and predicted outcomes.

SHAP was applied to interpret the predictions of the LSTM model for peak EC values during the summer period, a time characterized by increased saltwater intrusion and elevated EC measurements. Following the training and validation of the model, SHAP values were computed for the test dataset to assess the influence of both internal and external features. These values represent the individual contribution of each feature to the model's output, thereby enabling a detailed understanding of the factors influencing salinization within the river system.

## 4. Results

The results of this study encompass both the predictive performance of the LSTM model, and the interpretability insights derived from SHAP analysis for peak EC values during the summer period. To evaluate the performance of the LSTM model, a comparative analysis was conducted against a Linear Regression Model using key evaluation metrics, including mean squared error (MSE), mean absolute error (MAE), and normalized root mean squared error (NRMSE), calculated. The linear regression model served as a baseline, while the LSTM model demonstrated ability to capture temporal dependencies as given in **Table 4.1**. The results confirmed that LSTM offers improved accuracy and adaptability in dynamic environments. As expected, sensor D2 exhibited higher error levels compared to sensor U2, which can be attributed to its increased exposure to external influences and saltwater intrusion, expose it more sensitive to environmental variability.

**Table 4.1. LSTM Model Performance in terms of evaluation metrics**

| | LSTM | | | Linear Regression | | |
|---|---|---|---|---|---|---|
| | **MSE** | **MAE** | **NRMSE** | **MSE** | **MAE** | **NRMSE** |
| **U2** | 0.22 | 0.44 | 0.64 | 0.39 | 0.59 | 0.85 |
| **D2** | 8.36 | 2.72 | 2.06 | 16.35 | 4.02 | 3.04 |

### 4.1. SHAP Explanations of Peak Points of Upstream Located Sensor (U2)

The SHAP explanation was employed for sensor location U2, utilizing the SHAP summary plot (Figure 4.2a) and the contributions of features for all identified peak points (Figure 4.2b). To ensure clear visualization and understanding, the mean of the time steps was used for visualization purposes. As illustrated in the summary plot, the contributions of features for determining the peak points of sensor U2, sensor U3 was the most significant feature. An increase in the sensor readings of U2 for the predictions of peak points positively influenced from its neighbouring sensors of U1 and U3. Additionally, it is evident that an increase in the upstream discharge value leads to negative contributions to peak points due to flushing and dilution effects. As shown in Figure 4.2b, the mean values of each feature's contribution for all peak points were calculated. From this, it was determined that sensor readings from U1, U3, temperature, and wind speed positively contribute to our predictions, while the remaining features contribute negatively.

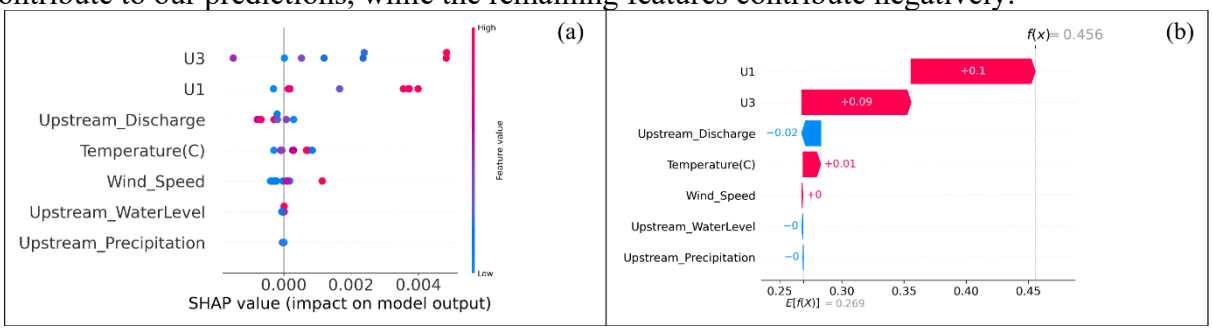


**Figure 4.2** (a) SHAP summary plot (b) contributions of features

### 4.2. SHAP Explanations of Peak Points of Downstream Located Sensor (D2)

For the peak points of sensor D2, the same SHAP-based interpretability procedure was applied. As illustrated in Figure 4.3a, the sensor positioned closer to the downstream (North Sea) was found

to employ a greater influence on the peak points predictions of sensor D2. Another notable observation, as shown in Figure 4.3b, is that downstream discharge was found to contribute positively to the predicted values. However, the summary plot revealed an interesting pattern: a decrease in downstream discharge resulted in a positive contribution to the predictions, indicating an inverse relationship. In other words, negative discharge values, which indicate saltwater intrusion, positively contribute to the prediction values. Upon analyzing the mean SHAP contributions for the predicted peak points (Figure 4.3b), sensor D3 was identified as having the highest impact, supporting its potential role as an early-warning indicator of tidal movement. Conversely, downstream precipitation was observed to contribute negatively to the model's predictions.

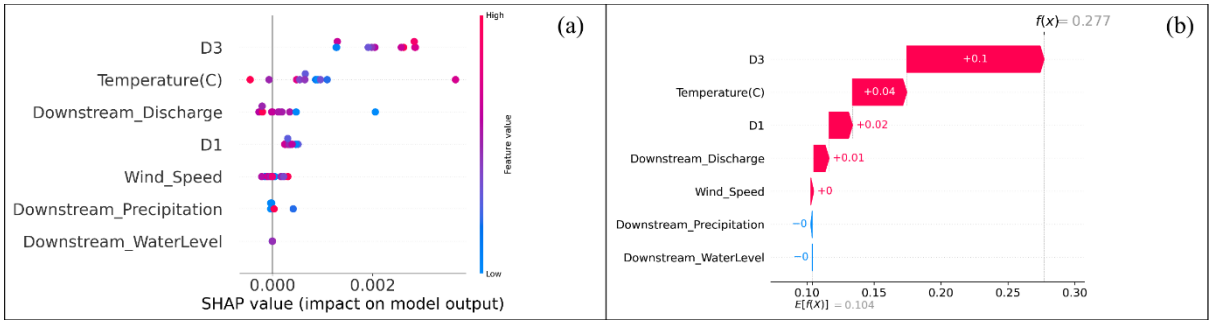


**Figure 4.3** (a) SHAP summary plot (b) contributions of features

## 5. Conclusion

This study has demonstrated the potential of integrating LSTM networks, with SHAP to predict and interpret water quality parameters in river systems, with a particular focus on salinization by using the EC measurements. By quantifying the contribution of individual features, SHAP offers a transparent understanding of how internal and external factors influence the predictions of salinity levels according to sensor locations in a river system. Based on the SHAP analysis, it was observed that the distinct features contributed with varying orders of magnitude depending on the sensor location within the river system, and it was established that neighbouring sensors and temperature plays a consistent role in enhancing salinization processes across both parts of river. The insights derived from the SHAP analysis have facilitated informed decision-making and supported the integration of interpretable LSTM models into water quality monitoring and management practices. Consequently, this study has contributed to bridging the gap between machine learning methodologies and their practical applications in environmental systems.

## REFERENCES

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Anh, N. T., Can, L. D., Nhan, N. T., Schmalz, B., & Luu, T. Le. (2023). Influences of key factors on river water quality in urban and rural areas: A review. *Case Studies in Chemical and Environmental Engineering*, *8*. https://doi.org/10.1016/j.cscee.2023.100424

Chen, Y., Song, L., Liu, Y., Yang, L., & Li, D. (2020). A review of the artificial neural network models for water quality prediction. In *Applied Sciences (Switzerland)* (Vol. 10, Issue 17). MDPI AG. https://doi.org/10.3390/app10175776

Gao, L., & Li, D. (2014). A review of hydrological/water-quality models. In *Frontiers of Agricultural Science and Engineering* (Vol. 1, Issue 4, pp. 267–276). Higher Education Press Limited Company. https://doi.org/10.15302/J-FASE-2014041

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. In *Cognitive Computation* (Vol. 16, Issue 1, pp. 45–74). Springer. https://doi.org/10.1007/s12559-023-10179-8

Hellen, N., & Marvin, G. (2022). Explainable AI for Safe Water Evaluation for Public Health in Urban Settings. *2022 International Conference on Innovations in Science, Engineering and Technology, ICISET 2022*, 227–232. https://doi.org/10.1109/ICISET54810.2022.9775912

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Liang, Z., Zou, R., Chen, X., Ren, T., Su, H., & Liu, Y. (2020). Simulate the forecast capacity of a complicated water quality model using the long short-term memory approach. *Journal of Hydrology*, *581*. https://doi.org/10.1016/j.jhydrol.2019.124432

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. http://arxiv.org/abs/1705.07874

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2018). *Consistent Individualized Feature Attribution for Tree Ensembles*. http://arxiv.org/abs/1802.03888

Mazhar, S., Pellegrini, E., Contin, M., Bravo, C., & De Nobili, M. (2022). Impacts of salinization caused by sea level rise on the biological processes of coastal soils - A review. In *Frontiers in Environmental Science* (Vol. 10). Frontiers Media S.A. https://doi.org/10.3389/fenvs.2022.909415

Park, J., Kim, K. T., & Lee, W. H. (2020). Recent advances in information and communications technology (ICT) and sensor technology for monitoring water quality. *Water (Switzerland)*, *12*(2). https://doi.org/10.3390/w12020510

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2825–2830. http://arxiv.org/abs/1201.0490

Rode, M., Wade, A. J., Cohen, M. J., Hensley, R. T., Bowes, M. J., Kirchner, J. W., Arhonditsis, G. B., Jordan, P., Kronvang, B., Halliday, S. J., Skeffington, R. A., Rozemeijer, J. C., Aubert, A. H., Rinke, K., & Jomaa, S. (2016). Sensors in the Stream: The High-Frequency Wave of the Present. *Environmental Science and Technology*, *50*(19), 10297–10307. https://doi.org/10.1021/acs.est.6b02155

Samek, W., Wiegand, T., & Müller, K.-R. (2017). *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models*. http://arxiv.org/abs/1708.08296

Tickner, D., Parker, H., Moncrieff, C. R., Oates, N. E. M., Ludi, E., & Acreman, M. (2017). Managing rivers for multiple benefits-A coherent approach to research, policy and planning. Frontiers in Environmental Science, 5. https://doi.org/10.3389/fenvs.2017.00004

Wai, K. P., Chia, M. Y., Koo, C. H., Huang, Y. F., & Chong, W. C. (2022). Applications of deep learning in water quality management: A state-of-the-art review. In *Journal of Hydrology* (Vol. 613). Elsevier B.V. https://doi.org/10.1016/j.jhydrol.2022.128332