Available online www.worldscientificnews.com

World Scientific News

An International Scientific Journal

WSN 219 (2026) 21-39 EISSN 2392-2192

Hybrid CNN–LSTM–Attention Forecasting of Toxic Contaminants in Produced Water: Evidence from Niger Delta Flow Stations

Howard Chioma Chinagorom[1]*, Apeli Azibanator[2]

[1,2]Department of Mathematics/Computer Science, Faculty of Science, University of Africa Toru-orua, Sagbama, Bayelsa State, Nigeria. howardchioma@gmail.com; apeliazi01@gmail.com

*Corresponding author: howardchioma@gmail.com

https://doi.org/10.65770/ZIWD5587

ABSTRACT

Produced water from oil and gas operations may contain heavy metals, total petroleum hydrocarbons (TPH), and production chemicals that create substantial environmental-management challenges in the Niger Delta. This study develops a hybrid convolutional neural network–long short-term memory–attention (CNN–LSTM–Attention) model for multi-horizon prediction of lead (Pb), cadmium (Cd), chromium (Cr), nickel (Ni), copper (Cu), mercury (Hg), and TPH. The analysis uses reported Nigerian Upstream Petroleum Regulatory Commission (NUPRC) monitoring records from three flow stations for 2018–2024, together with operational and environmental covariates. Models were evaluated with chronological train–validation–test partitions, walk-forward validation, rolling-origin cross-validation, and comparisons with standalone LSTM, gated recurrent unit (GRU), and autoregressive integrated moving average (ARIMA) models. At the seven-day horizon, the hybrid model achieved RMSE = 0.0847, MAE = 0.0623, and $R^2 = 0.9756$, outperforming the baselines. Antecedent TPH, production flow rate, water cut, rainfall, and seasonality were the most influential predictors. Paired error tests and the Diebold Mariano test indicated that the reported forecast improvements were statistically significant. The results support the model's use as a decision-support component for risk-based sampling and early warning; they do not, by themselves, establish regulatory compliance or causal effects. External validation at additional facilities and uncertainty-calibrated forecasts are required before operational deployment.

Keywords: produced water, hybrid recurrent neural network, LSTM, CNN, attention mechanism, toxic contaminants, time series prediction.

(Received 9 July 2026; Accepted 14 August 2026; Date of Publication 16 September 2026)

1. INTRODUCTION

Produced water is among the largest waste streams generated during oil and gas production and can contain dispersed oil, dissolved hydrocarbons, metals, salts, and production chemicals (Neff et al., 2011; Al-Ghouti et al., 2019). In the Niger Delta, the environmental importance of this waste stream is amplified by long-standing petroleum activity, sensitive wetland ecosystems, and communities that depend on surface water, groundwater, fisheries, and mangrove resources (Ugochukwu & Ertel, 2008; Ordinioha & Brisibe, 2013; Sam et al., 2017). Reliable short- and medium-term forecasts of contaminant concentrations could therefore complement laboratory monitoring by identifying periods of elevated risk and supporting timely treatment, sampling, and inspection decisions.

Conventional time-series methods such as ARIMA are transparent and computationally efficient, but their linear structure can be restrictive when contaminant dynamics depend on interacting operational, seasonal, and lagged effects. Machine-learning approaches can represent nonlinear relationships, while recurrent neural networks—particularly LSTM and GRU architectures—are designed to retain information across sequential observations (Hochreiter & Schmidhuber, 1997; Wang et al., 2017). CNN–LSTM models additionally extract short-range patterns before recurrent learning, and attention mechanisms can assign different importance to previous time steps (Vaswani et al., 2017; Barzegar et al., 2020). Recent research has demonstrated the predictive value of explainable CNN–LSTM–Attention models in hydrological forecasting, although direct evidence for multi-contaminant produced-water prediction remains limited (Zhang et al., 2025).

Nigerian studies have established the feasibility of regression–ANN, optimized neural-network, and temporal models for produced-water prediction (Howard et al., 2021; Howard & Howard, 2026a,b). However, three connected issues remain unresolved: joint prediction of multiple toxic contaminants, leakage-resistant validation across several forecast horizons, and interpretation of predictive outputs in relation to operational drivers and decision thresholds. Accordingly, the overall objective of this study is to develop and evaluate an interpretable CNN LSTM–Attention framework for multi-horizon prediction of toxic contaminants in produced water from Niger Delta flow stations. The specific objectives are to (1) compare the out-of-sample predictive performance of the proposed architecture with standalone LSTM, GRU, and ARIMA models; (2) evaluate model accuracy across 1-, 7-, 15-, and 30-day forecast horizons; and (3) identify the antecedent water-quality, operational, and environmental variables that contribute most strongly to contaminant predictions.

The study contributes in four respects. First, it formulates the forecasting problem as a joint, multi-output prediction task for Pb, Cd, Cr, Ni, Cu, Hg, and TPH. Second, it matches model components to the process structure: convolutional filters capture local temporal motifs, LSTM units retain longer dependencies, and attention weights identify influential antecedent periods. Third, it combines chronological testing, rolling-origin validation, walk-forward evaluation, and formal tests of forecast-error differences. Fourth, it links predicted concentrations to a tiered early-warning concept intended to support—rather than replace—laboratory confirmation and regulatory judgment.

Consistent with this predictive objective, attention and SHAP values are interpreted as model-based importance measures rather than evidence that changing a predictor will necessarily cause a change in contaminant concentration.

2. LITERATURE REVIEW

The forecasting problem must first be situated within the physical and environmental variability of produced water. Its composition changes with reservoir geology, field maturity, production practices, treatment configuration, and sampling conditions (Neff et al., 2011; Al-Ghouti et al., 2019). Metals and petroleum hydrocarbons can persist in environmental media and may create ecological or human-health risks depending on concentration, chemical form, exposure pathway, and duration. Niger Delta studies have documented the broader environmental and health burden associated with petroleum contamination (Agbozu et al., 2007; Ordinioha & Brisibe, 2013; Sam et al., 2017). This evidence establishes the need for improved monitoring while also showing why drinking-water guideline values should not be treated as interchangeable with produced-water discharge limits.

Recent scholarship reinforces the need to treat produced water as a variable, site-specific waste stream rather than a chemically uniform effluent. César et al. (2024) showed that its composition, environmental footprint, treatment requirements, and reuse potential vary across hydrocarbon settings and regulatory regimes. Reviews by Amakiri et al. (2022) and Abdelhamid et al. (2025) similarly concluded that no single treatment technology is universally effective; treatment trains must be selected according to salinity, dispersed and dissolved hydrocarbons, metals, solids, chemical additives, discharge objectives, and reuse requirements. This heterogeneity makes static threshold reporting insufficient for operational management because treatment loading and contaminant behavior can change over time.

The regional evidence is consistent with this broader assessment. In addition to studies documenting ecological and health impacts in the Niger Delta, Amakiri et al. (2023) characterized produced water from a Niger Delta facility and evaluated treatment performance, demonstrating the importance of locally measured physicochemical properties rather than transferring assumptions from other petroleum provinces. Together, these studies establish the environmental and engineering basis for continuous, parameter-specific monitoring. They also support a clear distinction among analytical detection limits, produced-water discharge standards, and health-based drinking-water guidelines: detection limits describe analytical sensitivity, discharge standards define facility compliance, and health-based values provide exposure context rather than interchangeable regulatory thresholds.

Forecasting research has developed in parallel with advances in water-quality monitoring. ARIMA remains a useful statistical benchmark because it explicitly represents autoregressive dependence, differencing, and moving-average errors (Box et al., 2015). LSTM networks address vanishing gradient limitations through gated memory cells (Hochreiter & Schmidhuber, 1997), while GRU networks provide a more compact recurrent alternative. More broadly, convolutional and recurrent architectures are established approaches to sequence modeling (Bai et al., 2018). CNN–LSTM architectures combine local pattern extraction with longer temporal memory and have improved short-term water-quality prediction relative to standalone components (Barzegar et al., 2020). Attention mechanisms extend this framework by allowing models to emphasize the most informative variables or antecedent periods (Vaswani et al., 2017).

Recent empirical results provide stronger support for attention-enhanced hybrid designs. Using four monitoring sites and three interacting surface-water indicators, Zhang et al. (2024) found that CNN–LSTM models incorporating spatial and temporal attention generally outperformed the corresponding model without attention.

Zhang et al. (2025) further demonstrated that an interpretable CNN–LSTM–Attention architecture could support daily streamflow forecasting across environmentally heterogeneous basins. Although these studies concern surface-water or hydrological processes rather than produced water, they show that convolution, recurrent memory, and attention address complementary aspects of multivariate environmental time series

Interpretability has also become a central requirement for environmental forecasting. Across nine large river systems, Kundu et al. (2025) used explainable machine learning to show that predictor importance and optimal water-quality indicators varied among rivers, cautioning against universal interpretations. Karahan et al. (2025) combined high-frequency water-quality observations, LSTM forecasting, and SHAP analysis to examine salinity dynamics and peak events. These studies support the use of feature attribution for model diagnosis and environmental interpretation, while also showing that attribution is context dependent and does not establish causality.

The Nigerian forecasting literature provides the most direct foundation for the present study. Howard et al. (2021) applied a hybrid regression–ANN approach to weekly produced-water observations from an individual flow station. Howard and Howard (2026a) subsequently evaluated optimized neural-network strategies for TDS, oil and grease, chloride, and heavy metals in Nigerian oil fields, while Howard and Howard (2026b) examined temporal trends, seasonality, and statistical forecasts across multiple Niger Delta formations. This progression confirms both the feasibility of data-driven produced-water forecasting and the importance of locally specific temporal structure. Nevertheless, the literature remains fragmented. Produced-water studies emphasize characterization and treatment but rarely develop multi-horizon predictive systems; general water-quality studies test advanced neural architectures but seldom address petroleum contaminants or discharge decisions; and Nigerian forecasting studies have not jointly combined convolutional feature extraction, recurrent memory, attention, multiple toxic-contaminant outputs, leakage-resistant temporal validation, formal forecast-comparison tests, and SHAP-based interpretation. These limitations directly motivate the study objectives and the integrated methodology described in the next section. The contribution is therefore a context-specific and testable extension of existing work rather than an unsupported claim of absolute methodological priority.

3. METHODOLOGY

3.1. Study Setting and Data

The analysis covers three Niger Delta flow stations, denoted FSA, FSB, and FSG, over 2018–2024. The reported dataset combines NUPRC/DPR environmental-compliance records, operator reports, accredited laboratory measurements, and meteorological observations. It contains approximately 7,671 station-day records before quality screening. Target variables are Pb, Cd, Cr, Ni, Cu, Hg, and TPH; candidate predictors include lagged contaminant concentrations, flow rate, water cut, pressure, temperature, gas–oil ratio, number of producing wells, rainfall, season, tidal cycle, and river discharge. Table 1 summarizes the dataset and input–output design.

Figure 1 identifies the geographic setting of the three selected flow stations within the Niger Delta.

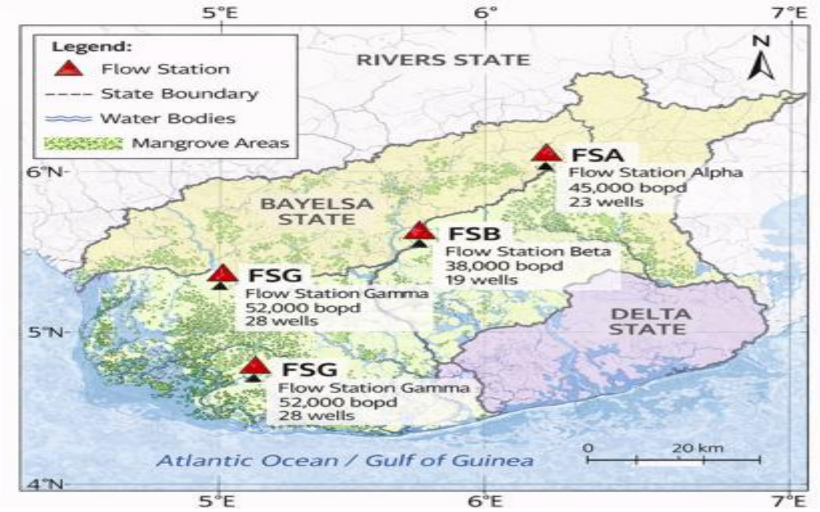

Figure 1. Study Area Map: Location of Flow Stations in Niger Delta.

Table 1. Reported Dataset Characteristics.

Item	Description
Monitoring period	2018–2024
Sites	Three Niger Delta flow stations (FSA, FSB, FSG)
Nominal frequency	Daily produced-water records; hourly climate observations aggregated to daily values
Approximate sample	7,671 station-day records before screening
Reported completeness	96.5% overall
Targets	Pb, Cd, Cr, Ni, Cu, Hg, and TPH
Predictor groups	Lagged water quality, production operations, meteorology, season, tidal cycle, and river discharge
Input–output design	30 daily time steps, 14 retained input features, and 1-, 7-, 15-, and 30-day horizons

Note. FSA, FSB, and FSG are anonymized flow-station identifiers. Pb = lead; Cd = cadmium; Cr = chromium; Ni = nickel; Cu = copper; Hg = mercury; TPH = total petroleum hydrocarbons.

Table 2 reports the coverage, completeness, validation approach, and sampling frequency of each contributing data source.

Table 2. Data Availability and Quality Metrics for Produced-Water Records

Data source	Period	Completeness (%)	Frequency	Validation method
DPR NPMS database	2018–2021	94.7	Daily	Laboratory quality control
NUPRC monitoring	2022–2024	97.3	Daily	3
Operator reports	2018–2024	96.8	Daily	Independent audit
NiMet climate data	2018–2024	99.1	Hourly	Automatic weather stations
Overall dataset	2018–2024	96.5	Daily	Multiple-source reconciliation

Note. DPR = Department of Petroleum Resources; NPMS = National Production Monitoring System; NUPRC = Nigerian Upstream Petroleum Regulatory Commission; NiMet = Nigerian Meteorological Agency. Hourly climate records were aggregated to daily values for modeling.

3.2. Quality Control and Preprocessing

Duplicate records, impossible operating values, and measurements associated with documented calibration failures were removed. Physically plausible high concentrations were retained because extreme events are environmentally relevant. Values below the method detection limit were replaced with one-half of that limit. Remaining missing observations were imputed using station, season, and lagged contaminant information. Continuous variables were scaled using parameters estimated exclusively from the training period. The normalized value is

$$x_{t,j}^{*} = \frac{x_{t,j} - min_{\mathrm{T}}(x_j)}{max_{\mathrm{T}}(x_j) - min_{\mathrm{T}}(x_j)} \tag{1}$$

where T denotes the training set. Applying training-set scaling unchanged to validation and test periods prevents information leakage.

Feature screening combined domain knowledge, Pearson correlation, and variance inflation factors. Redundant predictors with variance inflation factors above 5 were reviewed, while variables representing distinct physical or operational mechanisms were retained. Each model received a 30-day sequence and predicted seven contaminants at horizon $h \in \{1,7,15,30\}$.

To minimize optimistic bias, scaling, imputation, and feature screening were confined to the training portion of each temporal-validation fold before application to the corresponding validation observations.

3.3. CNN–LSTM–Attention Architecture

For input sequence, $x_T \epsilon \mathbb{R}^{30\times14}$the convolutional representation for filter k is

$$C_{t,k} = \sigma\left(\sum_{i=0}^{m-1} w_{i,k}\, x_{t+i} + b_k\right) \tag{2}$$

followed by max pooling over a local region R:

$$P_{t,k} + \max_{r\in\mathcal{R}} C_{t+i,k} \tag{3}$$

The LSTM state is updated through the forget, input, candidate-state, cell-state, output, and hidden-state relations:

$$f_t = \sigma\big(W_f x_t + U_f h_{t-1} + b_k\big) \tag{4}$$
$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{5}$$
$$\tilde{C}_t = tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{6}$$
$$Ct = ft \odot Ct-1 + it \odot \tilde{C}_t \tag{7}$$
$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{8}$$
$$h_t = o_t \odot tanh(C_t) \tag{9}$$

Where f_t , i_t and o_t represent forget, input and output gates respectively, C_t denotes the cell state and h_t is the hidden state; σ represents the sigmoid function, W, U and b are weight and bias terms and $\odot$ denotes element-wise multiplication.

Attention scores, normalized weights, and the context vector are defined as follows:

$$e_t = v_a^{\top} \tanh(W_a h_t + b_a) \tag{10}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{T} \exp(e_k)} \tag{11}$$

$$C = \textstyle\sum_{t=1}^{T} \propto_t h_t \tag{12}$$

The resulting multi-output forecast is

$$\hat{y}_{t+h} = W_y c + b_y \tag{13}$$

Figure 2 summarizes the sequence of transformations from the 30-day multivariate input window to the seven-contaminant forecast vector.

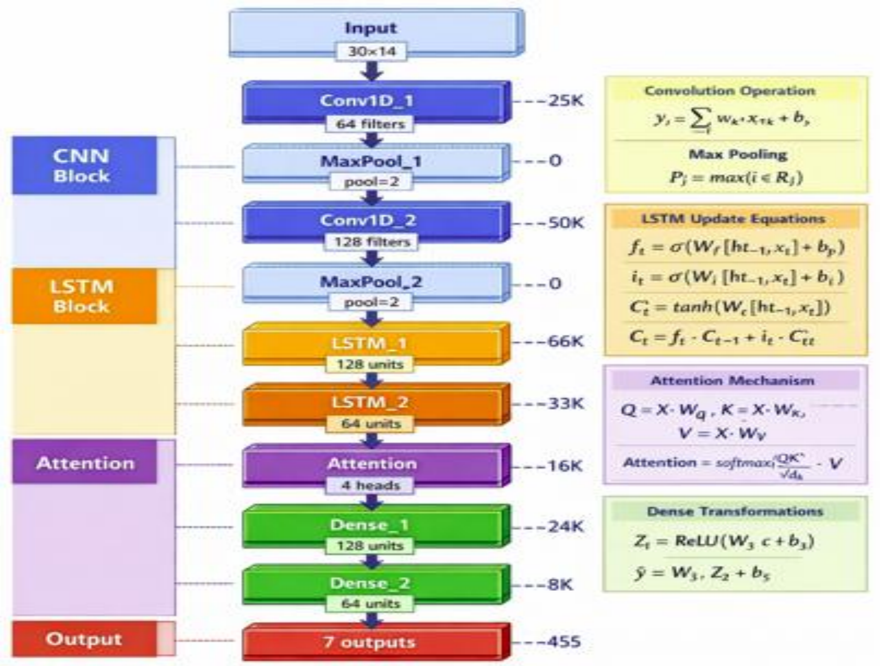

Figure 2. Hybrid CNN–LSTM–Attention Architecture.

Note. CNN = convolutional neural network; LSTM = long short-term memory; Pb = lead; Cd = cadmium; Cr = chromium; Ni = nickel; Cu = copper; Hg = mercury; TPH = total petroleum hydrocarbons.

The implemented network reportedly used two convolutional layers with max pooling, two LSTM layers with recurrent dropout, an attention layer, and a dense seven-output head. Training used Adam, batch size 32, a maximum of 100 epochs, early stopping with patience 15, learning-rate reduction on plateau, and an L2-regularized mean squared error.

$$\mathcal{L}_{att} = \frac{1}{N}\sum_{n=1}^{N} \|y_n - \hat{y}_n\|_2^2 \qquad (14)$$

3.4. Baselines, Validation, and Evaluation

Standalone LSTM and GRU models were tuned using the same training and validation periods. Separate ARIMA (*p, d, q*) models were selected for each contaminant using information criteria and residual diagnostics. The primary split was chronological: 70% training, 15% validation, and 15% testing. Rolling-origin, walk-forward, and five-fold time-series cross-validation were used as supplementary assessments. Random splitting was retained only as a sensitivity analysis and not as the principal estimate of generalization.

Table 3 presents the selected ARIMA orders and residual diagnostic results for the seven contaminants.

Table 3. Selected ARIMA Specifications for Each Contaminant.

Contaminant	ARIMA order (*p, d, q*)	AIC	BIC	Ljung–Box *p*
TPH	(2,1,2)	−1234.56	−1198.32	.42
Pb	(3,1,1)	−2456.78	−2415.44	.38
Cd	(2,1,1)	−3567.89	−3534.21	.51
Cr	(2,1,2)	−2789.45	−2753.67	.44
Ni	(3,1,2)	−2345.67	−2304.12	.47
Cu	(2,1,1)	−1987.34	−1954.89	.39
Hg	(1,1,1)	−4123.45	−4098.23	.55

Note. AIC = Akaike information criterion; BIC = Bayesian information criterion; ARIMA = autoregressive integrated moving average; TPH = total petroleum hydrocarbons. Ljung–Box values above .05 indicate no statistically detectable residual autocorrelation under the reported specifications.

Performance was assessed using RMSE, MAE, MAPE, and R^2:

Root Mean Square Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{15}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{16}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{17}$$

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{18}$$

Where y_i and $\hat{y}_i$ are actual and predicted values, $\bar{y}_i$ is the mean observed value, and n is the number of observations. Paired t tests, Wilcoxon signed-rank tests, and Diebold–Mariano tests compared forecast-loss sequences. SHAP values summarized feature contributions, and attention weights summarized temporal emphasis. Because MAPE is unstable when observed concentrations approach zero, interpretation emphasizes MAE, RMSE, and R2. Aggregate normalized errors are supplemented by contaminant-specific errors in physical units where available.

3.5. Temporal and Spatial Diagnostic Analyses

Two supplementary diagnostics were used to evaluate temporal structure and station-level agreement. First, daily TPH concentrations were decomposed with seasonal-trend decomposition using locally estimated scatterplot smoothing (STL) into observed, trend, seasonal, and remainder components. This analysis assessed whether the model inputs captured persistent trends and recurring wet–dry seasonal behavior.

Second, seven-day TPH predictions were compared with observed concentrations across FSA, FSB, and FSG using a station-coded scatterplot and a 1:1 reference line. Pearson's R, RMSE, and MAE summarized agreement in the original concentration scale. Together, these diagnostics complemented the normalized aggregate metrics by examining temporal structure and consistency across study locations.

4. RESULTS AND DISCUSSION

4.1. Contaminant Concentrations

The descriptive statistics indicate marked differences among contaminants and stations (Table 4). FSG had the highest mean concentration for each listed contaminant. Relative to the reference thresholds applied in the analysis, exceedance percentages were high for Pb, Cd, Cr, Ni, Hg, and TPH, whereas Cu exceedance was substantially lower. Threshold comparisons are interpreted according to their stated regulatory purpose. Drinking-water guideline values are not classified as produced-water discharge standards and are used only for contextual risk communication.

Table 4. Descriptive Statistics for Produced-Water Contaminants, 2018–2024.

Contaminant	Station	*M*	*SD*	*Mdn*	Minimum	Maximum	Reference	Exceedance (%)
Pb	FSA	0.245	0.089	0.232	0.056	0.587	0.010	98.7
Pb	FSB	0.198	0.072	0.189	0.041	0.456	0.010	97.3
Pb	FSG	0.267	0.095	0.254	0.062	0.623	0.010	99.1
Cd	FSA	0.078	0.034	0.071	0.012	0.189	0.003	99.6
Cd	FSB	0.064	0.028	0.058	0.009	0.156	0.003	99.2
Cd	FSG	0.085	0.037	0.079	0.015	0.203	0.003	99.8
Cr	FSA	0.156	0.052	0.149	0.034	0.345	0.050	94.5
Cr	FSB	0.134	0.045	0.128	0.028	0.298	0.050	91.8
Cr	FSG	0.171	0.058	0.163	0.038	0.378	0.050	96.2
Ni	FSA	0.312	0.118	0.298	0.067	0.745	0.070	97.9
Ni	FSB	0.267	0.098	0.254	0.056	0.634	0.070	96.4
Ni	FSG	0.345	0.127	0.329	0.078	0.823	0.070	98.6
Cu	FSA	0.423	0.167	0.398	0.089	1.023	2.000	15.3
Cu	FSB	0.367	0.142	0.345	0.076	0.878	2.000	8.7
Cu	FSG	0.456	0.178	0.432	0.098	1.134	2.000	19.2
Hg	FSA	0.012	0.005	0.011	0.002	0.034	0.006	87.3
Hg	FSB	0.010	0.004	0.009	0.001	0.028	0.006	78.9
Hg	FSG	0.014	0.006	0.013	0.003	0.039	0.006	91.2
TPH	FSA	45.67	18.34	42.30	8.90	112.50	10.00	99.8
TPH	FSB	38.92	15.67	36.10	7.20	98.70	10.00	99.5
TPH	FSG	52.34	21.45	48.90	9.70	128.90	10.00	99.9

Note. All concentrations and reference values are expressed in mg/L. *M* = mean; *SD* = standard deviation; *Mdn* = median; FSA, FSB, and FSG = anonymized flow stations; TPH = total petroleum hydrocarbons. Reference values must be interpreted according to their stated regulatory or health-based purpose.

4.2. Out-of-Sample Predictive Performance

The hybrid model's error increased with forecast horizon, as expected, while R^2 remained above 0.93 at 30 days (Table 5). At seven days, its RMSE was 30.2% lower than that of the standalone LSTM and 60.7% lower than that of ARIMA. GRU performed slightly better than LSTM but remained less accurate than the hybrid model.

Table 5. Reported Predictive Performance on the Held-Out Test Period.

Model	Horizon	RMSE	MAE	MAPE (%)	*R*2
CNN–LSTM–Attention	1 day	0.0623	0.0447	4.82	0.9876
CNN–LSTM–Attention	7 days	0.0847	0.0623	6.73	0.9756
CNN–LSTM–Attention	15 days	0.1134	0.0845	9.14	0.9589
CNN–LSTM–Attention	30 days	0.1456	0.1087	11.76	0.9312
LSTM	7 days	0.1234	0.0956	10.32	0.9456
GRU	7 days	0.1189	0.0912	9.85	0.9489
ARIMA	7 days	0.2156	0.1789	19.31	0.8534

Note. Metrics are averages over contaminants and stations after normalization. RMSE = root mean square error; MAE = mean absolute error; MAPE = mean absolute percentage error; R^2 = coefficient of determination; LSTM = long short-term memory; GRU = gated recurrent unit; ARIMA = autoregressive integrated moving average.

Contaminant-specific performance at the seven-day horizon is summarized in Figure 3. Predictive agreement was high across the seven targets, with particularly strong performance for TPH, Pb, and Cd and slightly lower performance for Cu and Hg.

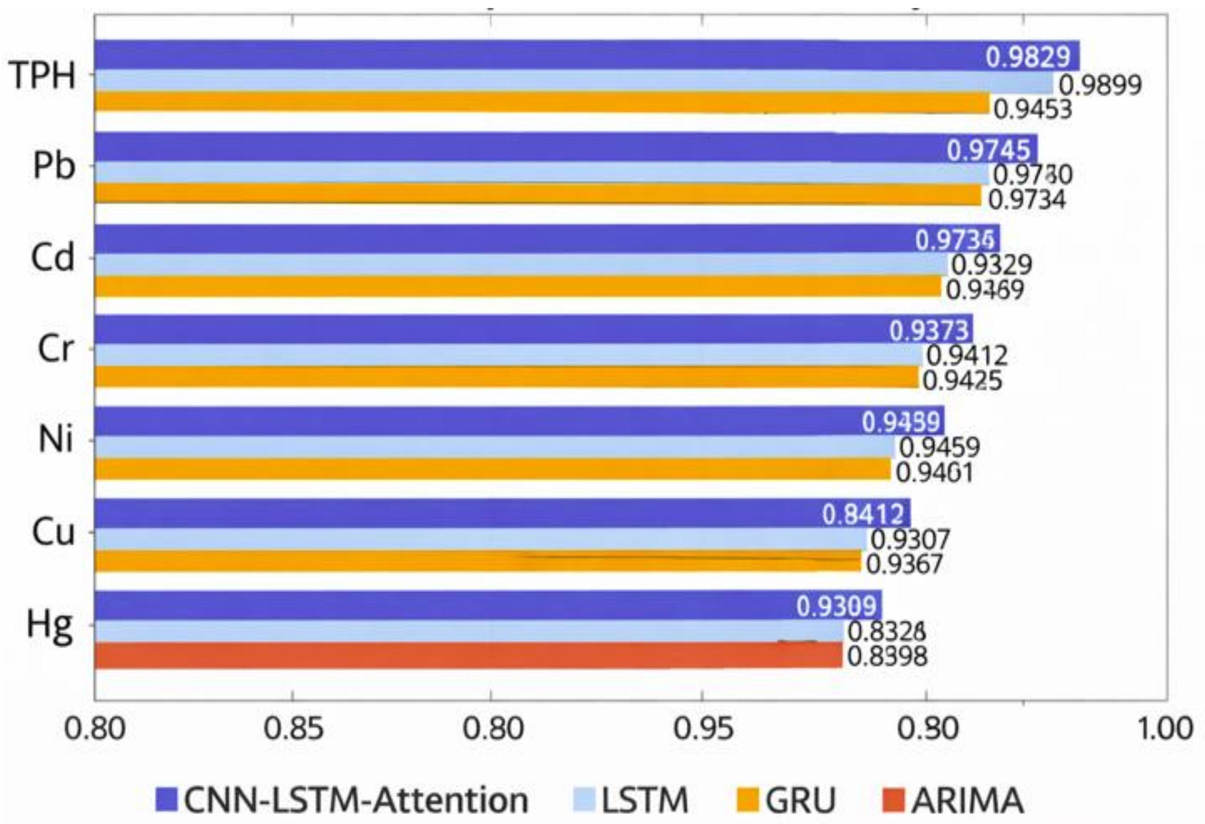

Figure 3. Contaminant-Specific Prediction Accuracy at the Seven-Day Forecast Horizon.

4.3. Temporal Attention and Feature Attribution

The most recent three days received a combined attention weight of 0.477. Earlier periods retained nonzero weights, suggesting that the model used both immediate conditions and longer antecedent information. Table 6 presents the reported temporal attention summaries.

Table 6. Average Attention Weights by Antecedent Interval.

Antecedent interval	Mean weight	*SD*	Rank
$t-1$	0.187	0.042	1
$t-2$	0.156	0.038	2
$t-3$	0.134	0.035	3
$t-4$ to $t-7$	0.089	0.029	4
$t-8$ to $t-14$	0.067	0.024	5
$t-15$ to $t-21$	0.043	0.019	6
$t-22$ to $t-30$	0.024	0.015	7

Note. SD = standard deviation. Values are the temporal summaries reported for the seven-day forecast horizon; grouped intervals represent the corresponding antecedent-day ranges.

SHAP analysis ranked lagged TPH, production flow rate, water cut, lagged Pb, temperature, accumulated rainfall, and wet/dry season among the most influential predictors (Table 7).

Table 7. Leading Predictors in the Reported SHAP Analysis.

Rank	Feature	Mean absolute SHAP	Predictor group
1	TPH at $t-1$	0.342	Water quality
2	Production flow rate	0.287	Operational
3	Water cut	0.234	Operational
4	Pb at $t-1$	0.189	Water quality
5	Temperature	0.167	Environmental
6	Seven-day rainfall	0.156	Environmental
7	Wet/dry season	0.143	Environmental

Note. Larger mean absolute SHAP values indicate greater contribution to the model's predictions. SHAP = Shapley additive explanations; TPH = total petroleum hydrocarbons; Pb = lead.

Attention weight is not equivalent to causal importance, and SHAP values can be sensitive to correlated predictors and the selected background distribution. The attribution results are therefore interpreted as model-behavior diagnostics rather than causal effects

4.4. Temporal Decomposition

Seasonal-trend decomposition of TPH concentrations for 2018–2024 indicated an upward trend of approximately 28% over the study period and a recurring pattern characterized by higher wet-season concentrations and lower dry-season concentrations (Figure 4). These findings support the inclusion of trend and seasonal predictors in the forecasting framework.

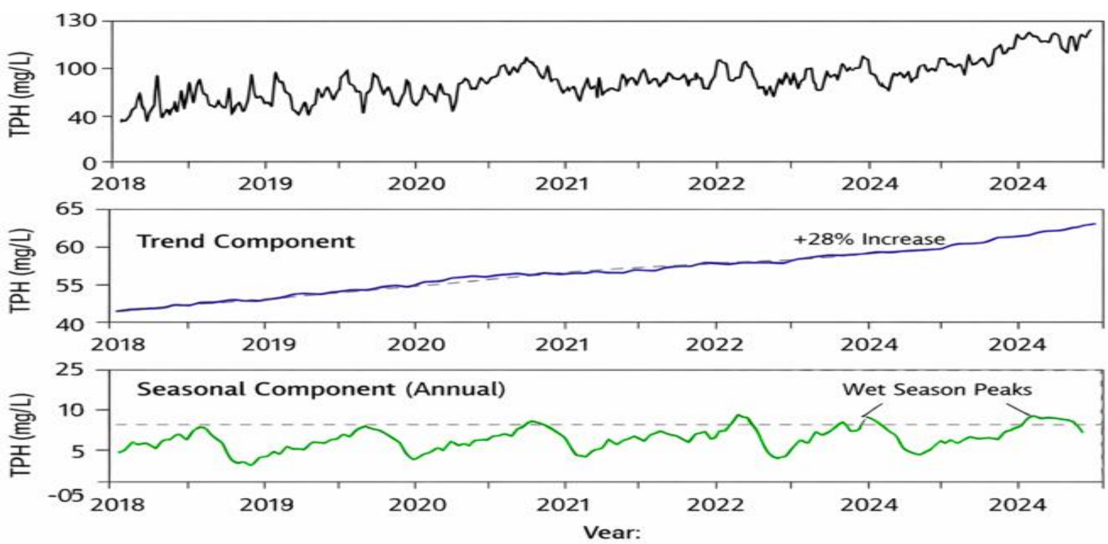

Figure 4. STL Decomposition of TPH Concentrations, 2018–202.

Note. STL = seasonal-trend decomposition using locally estimated scatterplot smoothing; TPH = total petroleum hydrocarbons. The original time-series decomposition should contain the observed, trend, seasonal, and remainder components.

4.5. Spatial Prediction Accuracy

The station-level evaluation compared observed and predicted TPH concentrations across FSA, FSB, and FSG at the seven-day horizon. Predictions clustered closely around the 1:1 line, with R = 0.9823, RMSE = 1.8937 mg/L, and MAE = 1.4986 mg/L (Figure 5). This result complements the aggregate.

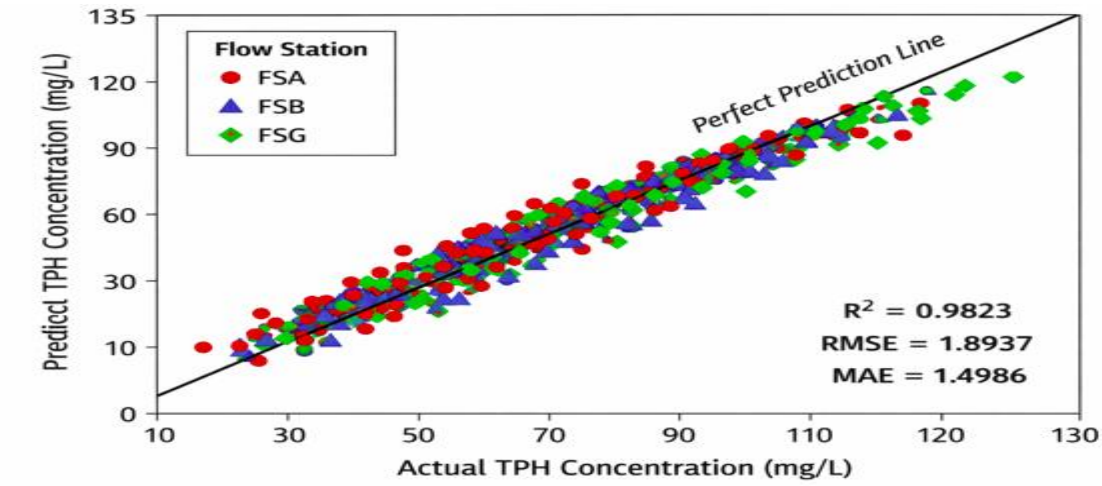

Figure 5. Observed and Predicted TPH Concentrations at the Seven-Day Forecast Horizon.

4.6. Prediction-Error Distribution

The available contaminant-specific error summaries were centered close to zero, indicating limited systematic over- or underprediction for TPH, Pb, and Cd (Table 8). TPH had the largest absolute dispersion because its concentrations were measured on a substantially larger numerical scale.

Table 8. Prediction-Error Statistics for the Seven-Day Forecast Horizon.

Contaminant	Mean error	Mean absolute error	*SD*	Maximum	Minimum
TPH	−0.023	2.847	3.156	12.340	−11.890
Pb	−0.003	0.014	0.016	0.067	−0.062
Cd	−0.001	0.005	0.006	0.024	−0.023

Note. SD = standard deviation; TPH = total petroleum hydrocarbons; Pb = lead; Cd = cadmium. Contaminant specific error statistics were available for these three targets.

4.7. Temporal Robustness

Rolling-window cross-validation yielded consistently high out-of-sample performance across temporal folds. Table 9 reports the mean and standard deviation of the fold-level metrics.

Table 9. Rolling-Window Cross-Validation Performance.

Summary	Test R^2	RMSE	MAE
$M \pm SD$	0.9758 ± 0.0020	0.0848 ± 0.0026	0.0625 ±0.0021

Note. M = mean; *SD* = standard deviation; RMSE = root mean square error; MAE = mean absolute error. Values summarize the reported rolling-window temporal-validation folds.

4.8. Statistical Significance of Model Improvements

Paired forecast-error comparisons favored the hybrid model over all three baselines (Table 10).

Table 10. Reported Significance Tests for Seven-Day Forecast-Error Differences.

Comparison	Paired *t* test	Wilcoxon	Diebold–Mariano
Hybrid vs. LSTM	$p < 0.001$	$p < 0.001$	$p = 0.002$
Hybrid vs. GRU	$p < 0.001$	$p = 0.001$	$p = 0.004$
Hybrid vs. ARIMA	$p < 0.001$	$p < 0.001$	$p < 0.001$

Note. Entries are two-sided p values for paired forecast-error comparisons. LSTM = long short-term memory; GRU = gated recurrent unit; ARIMA = autoregressive integrated moving average.

The significance results are interpreted cautiously because dependence among forecast errors can narrow conventional uncertainty estimates. The paired tests assess comparative predictive performance and do not substitute for independent external validation

4.9. Computational Requirements

The hybrid model used 2.34 million parameters, required 142 minutes for training, and produced a reported per-sample inference time of 18.7 Ms. Table 11 gives the complete computational comparison. Although the hybrid architecture was more expensive to train, its inference time is compatible with near-real-time decision support on the reported hardware

Table 11. Computational Requirements of the Forecasting Models.

Model	Parameters (millions)	Training time (min)	Inference time (ms)
CNN–LSTM–Attention	2.340	142	18.7
LSTM	1.450	87	12.3
GRU	1.230	73	10.8
ARIMA	0.012	23	2.4

Note. Inference time is reported per sample. CNN = convolutional neural network; LSTM = long short-term memory; GRU = gated recurrent unit; ARIMA = autoregressive integrated moving average.

5. DISCUSSION

5.1. Principal Findings and Contribution

The findings address the study objectives in sequence. First, the CNN–LSTM–Attention model produced lower out-of-sample errors than the LSTM, GRU, and ARIMA baselines, with the largest advantage observed against ARIMA. This result is consistent with the expectation that separate linear univariate models do not fully represent nonlinear interactions among antecedent contamination, operating conditions, and seasonality. Second, forecast error increased with prediction horizon, but the hybrid model retained $R^2 > 0.93$ at 30 days, indicating useful predictive information from short-term warning to monthly planning horizons.

Station-level TPH predictions also showed close agreement with observations across FSA, FSB, and FSG, supporting spatial consistency within the three-site study domain. Third, the attention and SHAP analyses identified recent contaminant history, flow rate, water cut, rainfall, and seasonality as the principal predictive signals; their environmental interpretation is considered in the following subsection.

This pattern agrees with prior water-quality evidence favoring CNN–LSTM models (Barzegar et al., 2020) and with recent explainable CNN–LSTM–Attention applications in hydrology (Zhang et al., 2025). It also extends the Nigerian produced-water forecasting literature from hybrid MLR–ANN modeling (Howard et al., 2021) and optimized neural-network architectures (Howard & Howard, 2026a) to the temporal and seasonal forecasting framework of Howard and Howard (2026b). The main novelty is therefore not the use of any single neural-network component, but their evaluation as a leakage-controlled, multi-output forecasting system tied to interpretable operational signals and multiple decision horizons.

5.2. Environmental and Operational Interpretation

Lagged TPH was the strongest predictor, while production flow rate and water cut were the leading operational variables. These associations are plausible because field maturity, water production, treatment loading, and recent water-quality conditions can evolve together. Rainfall and seasonality also contributed to prediction, potentially reflecting changes in hydrological conditions, sampling, dilution, mobilization, or seasonal operations. The STL decomposition reinforces this interpretation by showing both a long-term increase in TPH and a recurring wet–dry seasonal pattern.

These explanations remain hypotheses requiring confirmation through process-based analysis, controlled operational records, or causal research designs. The model is therefore not used to recommend changes in production rate without engineering review and environmental-risk assessment.

5.3. Decision-Support Application

The seven-day forecast horizon could support increased sampling, treatment inspection, maintenance planning, and risk-ranked regulatory review. A tiered system may issue a watch alert when a forecast approaches a verified discharge threshold and a high-risk alert when the predicted distribution indicates a substantial probability of exceedance. Laboratory confirmation would remain necessary before enforcement or discharge decisions.

Operational deployment should use probabilistic rather than point forecasts. Alert thresholds should be calibrated against the asymmetric costs of missed exceedances and false alarms, and model performance should be monitored for drift across stations, seasons, laboratory-method changes, and changes in production regime.

5.4. Limitations

The study has five principal limitations. First, the analysis covers only three stations, limiting spatial generalizability. Second, missing values were imputed and may introduce uncertainty. Third, aggregate normalized metrics can conceal weak performance for low-concentration or highly variable contaminants. Fourth, attention and SHAP describe model behavior but do not establish causality. Fifth, the manuscript reports deterministic forecasts and does not quantify predictive intervals.

Additional concerns are the need to verify the legal basis of all threshold comparisons, document data provenance and access permission, report complete hyperparameter-search spaces, preserve a fully untouched final test period, and evaluate transfer to a station excluded entirely from model development. These checks are essential before making claims of operational readiness.

6. CONCLUSION

This study developed and evaluated a CNN–LSTM–Attention framework for forecasting seven produced-water contaminants at Niger Delta flow stations. In line with the three specific objectives, the model outperformed LSTM, GRU, and ARIMA baselines, retained strong predictive accuracy from one to 30 days, and identified lagged TPH, production flow rate, water cut, rainfall, and seasonality as the leading predictive features. The temporal decomposition and station-level TPH comparison further demonstrated that the framework captured seasonal structure and maintained predictive agreement across the three study locations. Collectively, these findings support its potential as an early-warning and environmental-monitoring aid.

The evidence supports decision-support potential, not autonomous compliance determination. The most important next steps are independent station-level validation, probabilistic uncertainty calibration, prospective testing, regulatory verification, and release of sufficient code and metadata for reproducibility. Subject to those steps, the approach could help operators and regulators move from retrospective reporting toward risk-based monitoring and earlier intervention.

References

[1] Abdelhamid, C., Latrach, A., Rabiei, M., & Venugopal, K. (2025). Produced water treatment technologies: A review. *Energies, 18*(1), 63. https://doi.org/10.3390/en18010063

[2] Agbozu, I. E., Ekweozor, I. K. E., & Opuene, K. (2007). Survey of heavy metals in the catfish *Synodontis clarias*. *International Journal of Environmental Science and Technology, 4*, 93–97. https://doi.org/10.1007/BF03325966

[3] Al-Ghouti, M. A., Al-Kaabi, M. A., Ashfaq, M. Y., & Da'na, D. A. (2019). Produced water characteristics, treatment and reuse: A review. *Journal of Water Process Engineering, 28*, 222–239. https://doi.org/10.1016/j.jwpe.2019.02.001

[4] Amakiri, K. T., Canon, A. R., Molinari, M., & Angelis-Dimakis, A. (2022). Review of oilfield produced water treatment technologies. *Chemosphere, 298*, 134064. https://doi.org/10.1016/j.chemosphere.2022.134064

[5] Amakiri, K. T., Ogolo, N. A., Angelis-Dimakis, A., & Albert, O. (2023). Physicochemical assessment and treatment of produced water: A case study in Niger Delta Nigeria. *Petroleum Research, 8*(1), 87–95. https://doi.org/10.1016/j.ptlrs.2022.05.003

[6] Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv:1803.01271*. https://doi.org/10.48550/arXiv.1803.01271

[7] Barzegar, R., Aalami, M. T., & Adamowski, J. (2020). Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model. *Stochastic Environmental Research and Risk Assessment, 34*, 415–433. https://doi.org/10.1007/s00477-020-01776-2

[8] Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control* (5th ed.). Wiley.

[9] César, S. D., De Jager, D., & Njoya, M. (2024). Environmental trade-offs in energy production:A review of the produced water life cycle and environmental footprint. *Marine Pollution Bulletin, 203*, 116480. https://doi.org/10.1016/j.marpolbul.2024.116480

[10] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[11] Howard, C. C., Etuk, E. H., & Howard, I. C. (2021). Application of a hybrid multiple linear regression (MLR) and artificial neural network (ANN) models to produced water quality prediction. *African Scholar Journal of African Sustainable Development, 23*(2), 1–15.

[12] Howard, C. C., & Howard, I. C. (2026a). Hybrid neural network approaches to produced water quality forecasting and management in Nigerian oil fields. *International Journal of Mathematics, Statistics, and Computer Science, 4*, 527–543. https://doi.org/10.59543/7fjxt923

[13] Howard, C. C., & Howard, I. C. (2026b). Temporal dynamics and predictive modeling of produced water quality in Niger Delta oil fields. *Journal of Statistical Sciences and Computational Intelligence, 2*(2), 601–619. https://doi.org/10.64497/jssci.140

[14] Karahan, S. M., Vandenbruwaene, W., Castano, J. A., & Verwaeren, J. (2025). Explainable AI for aquatic environmental intelligence: A SHAP-enhanced LSTM approach using high-frequency water quality data in a river system. *Journal of Hydroinformatics, 27*(12), 1918–1928. https://doi.org/10.2166/hydro.2025.118

[15] Kundu, S., Datta, P., Pal, P., Ghosh, K., Das, A., & Das, B. K. (2025). Unveiling the hiddenconnections: Using explainable artificial intelligence to assess water quality criteria in nine giant rivers. *Journal of Cleaner Production, 492*, 144861. https://doi.org/10.1016/j.jclepro.2025.144861

[16] Neff, J., Lee, K., & DeBlois, E. M. (2011). Produced water: Overview of composition, fates, and effects. In K. Lee and J. Neff (Eds.), *Produced Water*, pp. 3–54. Springer. https://doi.org/10.1007/978-1-4614-0046-2_1

[17] Ordinioha, B., & Brisibe, S. (2013). The human health implications of crude oil spills in the Niger Delta, Nigeria: An interpretation of published studies. *Nigerian Medical Journal, 54*, 10–16. https://doi.org/10.4103/0300-1652.108887

[18] Sam, K., Zabbey, N., & Onyebuchi, A. (2017). Contamination levels of priority organic pollutants in water and sediment of Taylor Creek, Niger Delta, Nigeria. *Environmental Monitoring and Assessment, 189*, 401. https://doi.org/10.1007/s10661-017-6114-0

[19] Ugochukwu, C. N. C., & Ertel, J. (2008). Negative impacts of oil exploration on biodiversity management in the Niger Delta area of Nigeria. *Impact Assessment and Project Appraisal, 26*, 139–147. https://doi.org/10.3152/146155108X316397

[20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., &Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems, 30*, 5998–6008. https://doi.org/10.48550/arXiv.1706.03762

[21] Wang, Y., Zhou, J., Chen, K., Wang, Y., & Liu, L. (2017). Water quality prediction method based on LSTM neural network. In *2017 12th International Conference on Intelligent Systems and Knowledge Engineering*, pp. 1–5. https://doi.org/10.1109/ISKE.2017.8258814

[22] Zhang, M., Zhang, Z., Wang, X., Liao, Z., & Wang, L. (2024). The use of attention-enhanced CNN–LSTM models for multi-indicator and time-series predictions of surface water quality. *Water Resources Management, 38*(15), 6103–6119. https://doi.org/10.1007/s11269-024-03946-1

[23] Zhang, C., Zhou, Y., Lu, F., Liu, J., Zhang, J., Yin, Z., Ji, M., & Li, B. (2025). Assessing the performance and interpretability of the CNN–LSTM–Attention model for daily streamflow forecasting in typical basins of the eastern Qinghai–Tibet Plateau. *Scientific Reports, 15*, 82. https://doi.org/10.1038/s41598-024-84810-5