Optimizing refinery water reuse: High-precision prediction of complex water quality enabling a novel proactive warning index

Jie Xu [a], Shaoze Xiao [a], Jiaying Ma [a], Duanyang Shangguan [a], Huaqiang Chu [a, b *], Jiacai Xie [c], Xuefei Zhou [a, b], Yalei Zhang [a, b, d]

a State Key Laboratory of Water Pollution Control and Green Resource Recycling, College of Environmental Science and Engineering, Tongji University, Shanghai 200092, China

b Shanghai Institute of Pollution Control and Ecological Security, Tongji University, Shanghai 200092, China

c CNPC Research Institute of Safety&Environment Technology, Beijing 102206, China

d College of Environment & Safety Engineering, Fuzhou University, Fuzhou 350116, Fujian, China

*Corresponding authors: chuhuaqiang@tongji.edu.cn (H. Chu)

Text 1

$$WWQI = \omega_1 \cdot [NH_3 - N] + \omega_2 \cdot [COD] + \omega_3 \cdot [S^{2-}] + \omega_4 \cdot [Cl^-] + \omega_5 \cdot [pH] \quad (1)$$

where ω_i is the relative contribution weight of each influent feature obtained from the SHAP analysis. The average SHAP value is calculated through weighted percentage to obtain the ω_i value. [Feature] is the warning concentration limit value for the input water characteristic value.

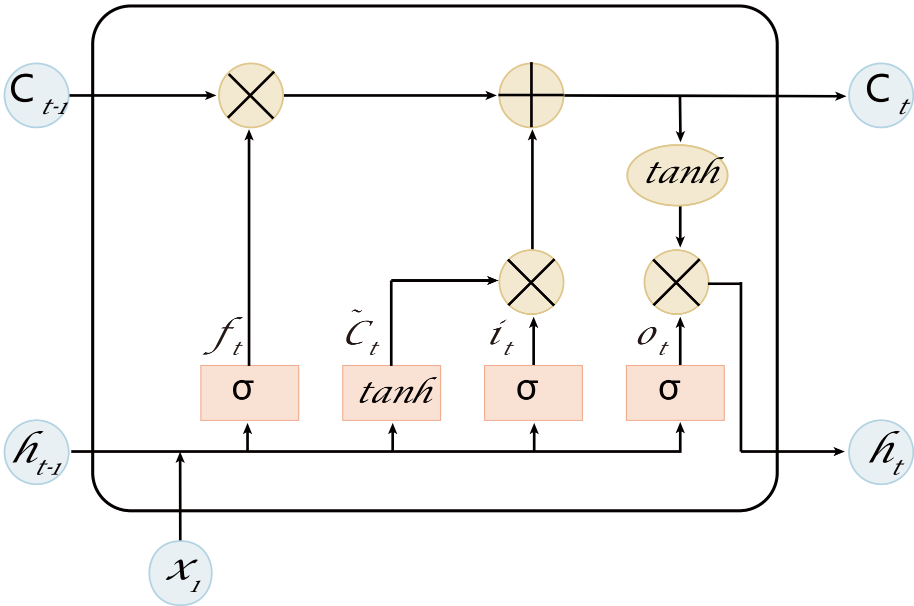

Fig. S1. LSTM cell structure.

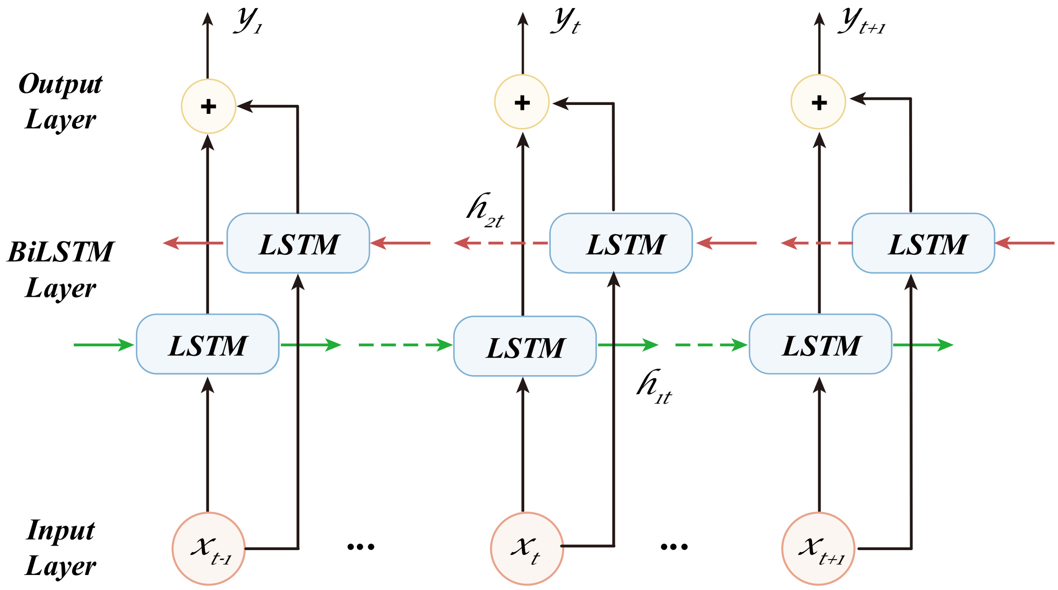

Fig. S2 BiLSTM structure

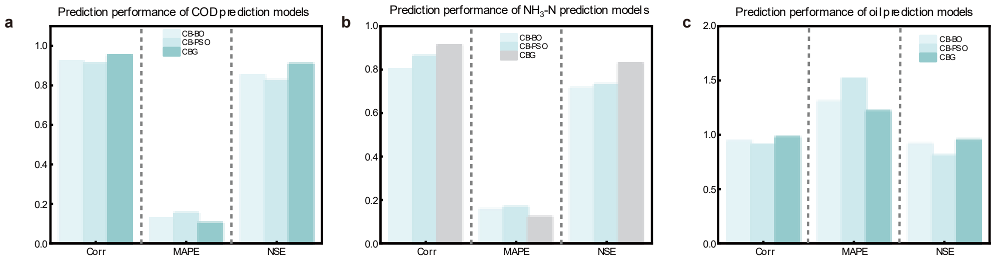

Fig. S3. Bar chart comparing the prediction performance metrics of CNN-BiLSTM models using the grey wolf optimization (GWO) algorithm, bayesian optimization (BO) algorithm, and Particle Swarm optimization (PSO) algorithm.

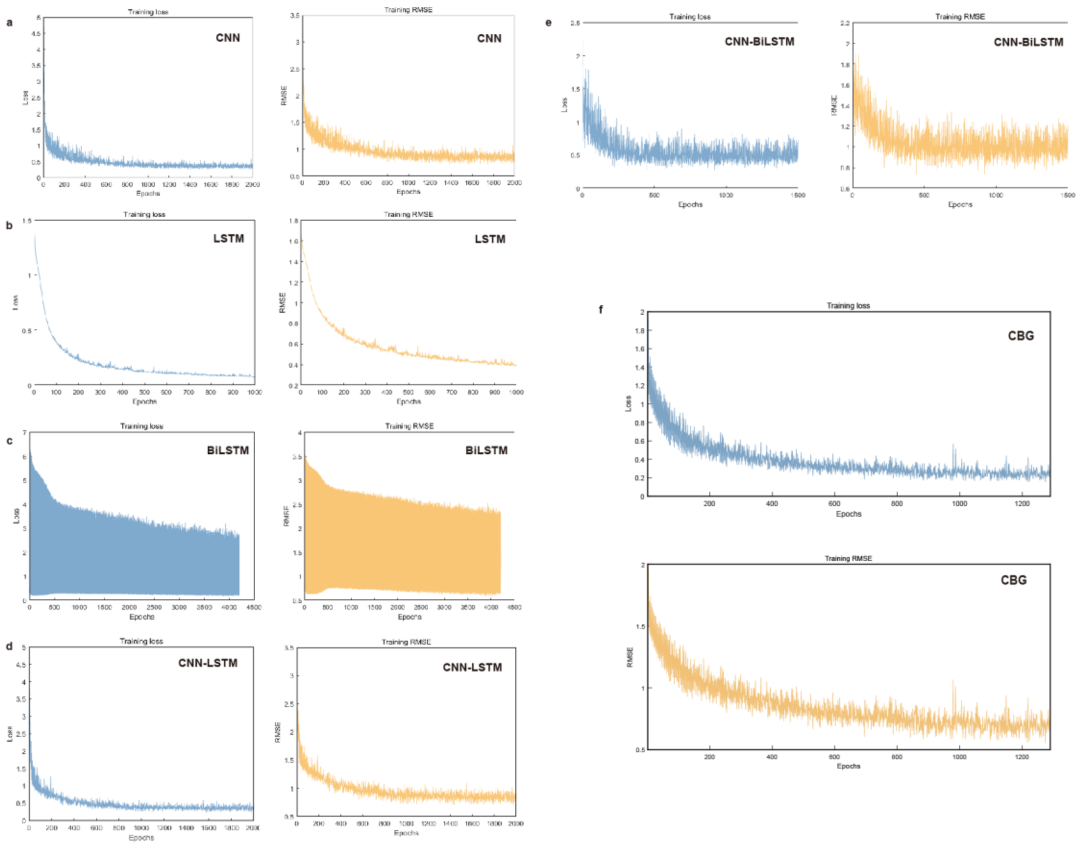

Fig. S4. Loss and error curves of each model's training process.

Table S1 Predictive performance evaluation of different deep learning models on training and test sets.

Variable	Metric	Training Set						Testing Set					
		CNN-BiLSTM	CNN	LSTM	CNN-LSTM	BiLSTM	CBG	CNN-BiLSTM	CNN	LSTM	CNN-LSTM	BiLSTM	CBG
COD	RMSE (mg/L)	499.17	747.98	135.42	382.96	674.82	374.62	671.71	765.42	1084.18	799.02	745.74	547.99
	MAE (mg/L)	272.71	370.02	86.13	198.62	375.31	181.09	349.58	393.39	602.84	355.03	420.77	253.99
	MAPE (%)	9.85	22.46	6.23	13.21	24.56	12.61	4.49	23.90	38.52	22.71	27.30	15.67
	Corr	0.17	0.51	0.94	0.89	0.59	0.89	0.10	0.38	0.47	0.52	0.46	0.73
	NSE	0.64	0.19	0.92	0.79	0.34	0.95	0.32	0.08	0.18	0.23	0.13	0.86
oil	RMSE (mg/L)	376.20	688.82	130.16	294.80	549.41	363.40	594.64	878.05	1096.63	784.08	721.25	460.18
	MAE (mg/L)	179.53	245.83	58.44	107.93	271.01	108.42	255.19	292.84	416.46	249.05	336.50	181.87
	MAPE (%)	50.93	255.75	74.52	121.69	405.71	133.19	72.07	304.77	524.79	231.39	487.65	195.76
	Corr	0.19	0.77	0.94	0.95	0.81	0.93	0.18	0.43	0.37	0.54	0.56	0.80

	NSE	0.84	0.47	0.93	0.90	0.66	0.85	0.58	0.09	0.14	0.14	0.22	0.77
	RMSE (mg/L)	4.40	6.29	1.91	3.88	5.25	3.41	5.15	6.62	7.89	5.61	5.75	4.46
	MAE (mg/L)	3.28	5.03	1.40	2.82	4.07	2.36	3.88	5.16	5.82	4.07	4.47	3.96
NH_3-N	MAPE (%)	4.06	31.10	8.69	17.20	24.82	14.33	4.48	31.97	33.30	23.65	27.09	16.26
	Corr	0.17	0.66	0.92	0.87	0.74	0.90	0.15	0.60	0.88	0.69	0.68	0.80
	NSE	0.68	0.34	0.89	0.75	0.54	0.90	0.55	0.26	0.55	0.46	0.44	0.78