Supplementary Information

Optimized Deep Learning Models for Effluent Prediction in Wastewater Treatment Processes

Canyun Yang[a], Zhuoyue Guo[a], Yun Geng[a], Fengshan Zhang[c], Wenguang Wei[c], Hongbin Liu[a,b,c,*]

[a] Jiangsu Co-Innovation Center of Efficient Processing and Utilization of Forest Resources, Nanjing Forestry University, Nanjing 210037, China

[b] Guangxi Key Laboratory of Clean Pulp & Papermaking and Pollution Control, College of Light Industry and Food Engineering, Guangxi University, Nanning 530004, China

[c] Laboratory for Comprehensive Utilization of Paper Waste of Shandong Province, Shandong Huatai Paper Co. Ltd., Dongying 257335, China

Corresponding Author:

*H.L. Tel.: +86-25-85427620; Fax: +86-25-85428793;

E-mail: hongbinliu@njfu.edu.cn

Submitted to

Environmental Science - Water Research & Technology

March 1, 2024

Figure S1 displays scatter plots of the 8 predictor variables versus the response variable, COD-S.

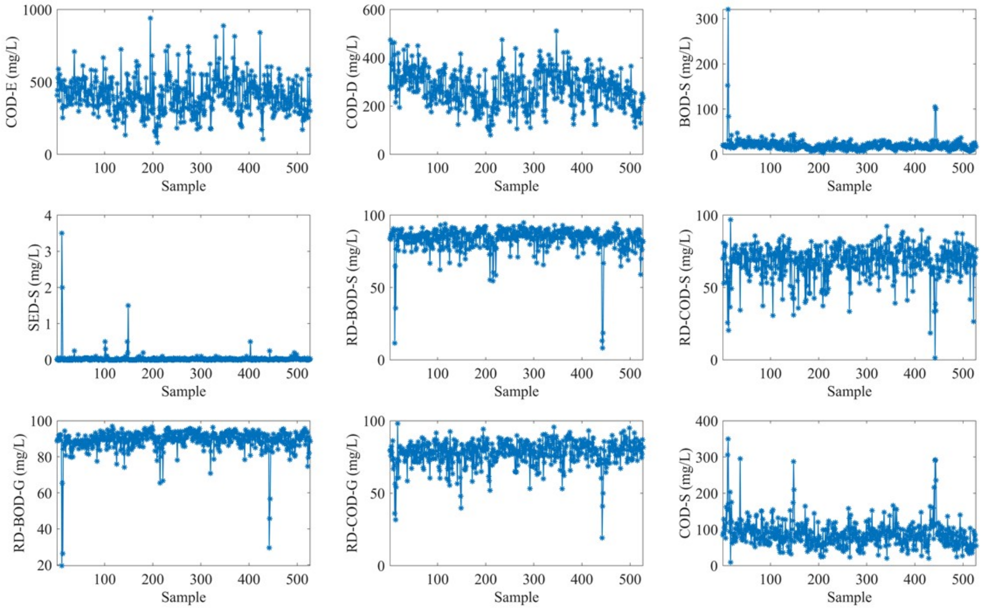

Figure S1. Predictor variables and response variable COD-S

Scatter plots of the nine predictor variables versus the response variable SS-S are shown in Figure S2.

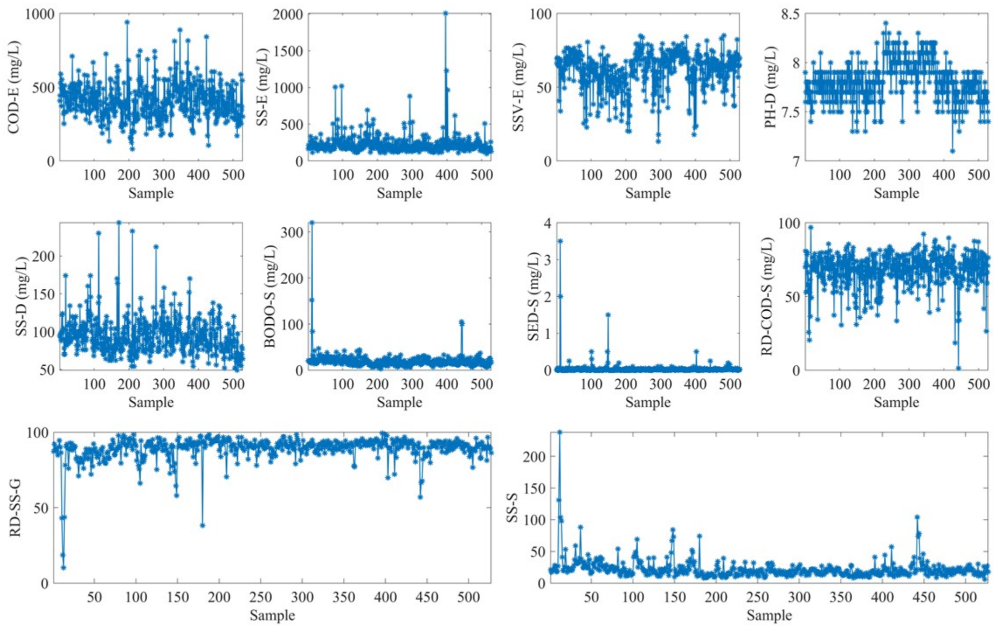

Figure S2. Predictor variables and response variable SS-S

Specific variable names and corresponding variable contents of the data used in the paper are shown in Table S1.

Table S1 Variable names and meanings in urban sewage data

Variable name	Implication
COD-E	Input chemical oxygen demand
COD-D	Oxygen input to secondary settler Chemical Oxygen Demand
BOD-S	Output Biological Oxygen Demand
COD-S	Output Chemical Oxygen Demand
SS-S	Output of suspended solids
SED-S	Exporting sediments
RD-BOD-S	Secondary settler input BOD
RD-COD-S	Oxygen input to secondary settler Chemical Oxygen Demand
RD-BOD-G	Global input of biological oxygen demand
RD-COD-G	Global Energy Input Chemical Oxygen Demand
SS-E	Input suspended matter
SSV-E	Input volatile suspended solids
PH-D	Secondary settler input pH
SS-D	Secondary settler input suspended solids
RD-SS-G	Global ability to enter suspensions

Table S2 shows the mean value, standard deviation, minimum value and maximum value of some data in this data set.

Table S2 Descriptive statistics of urban sewage data

Variable name	Standard deviation	Mean	Maximum	Minimum
COD-E	119.79	406.91	81.00	941.00
COD-D	73.72	273.99	80.00	511.00
BOD-S	17.20	20.00	3.00	320.00
COD-S	9.30	80.11	29.20	100.00
SS-S	0.19	0.04	0.00	3.50
SED-S	14.06	38.81	0.60	79.10
RD-BOD-S	6.84	88.97	19.60	97.00
RD-COD-S	8.75	77.87	19.20	98.10
RD-BOD-G	8.16	88.96	10.30	99.40
RD-COD-G	4.33	99.10	36.40	100.00
SS-E	135.82	227.38	98.00	2008.00
SSV-E	12.30	61.36	13.20	85.00
PH-D	0.20	7.81	7.10	8.40
SS-D	23.96	94.19	49.00	244.00
RD-SS-G	38.43	87.12	9.00	350.00

The structural diagram of the GRU is illustrated in Figure S3, where h denotes the internal memory state vector of GRU and h_{t-1} denotes the hidden state of the previous moment, encapsulating information from data observed by the preceding node. h_t denotes the hidden state passed to the next moment. $\tilde{h}_t$ denotes the candidate's hidden state. r_t and z_t denote the reset gate and update gate, respectively, the Sigmoid function, and the Tanh function. x_t denotes the input of the present node.

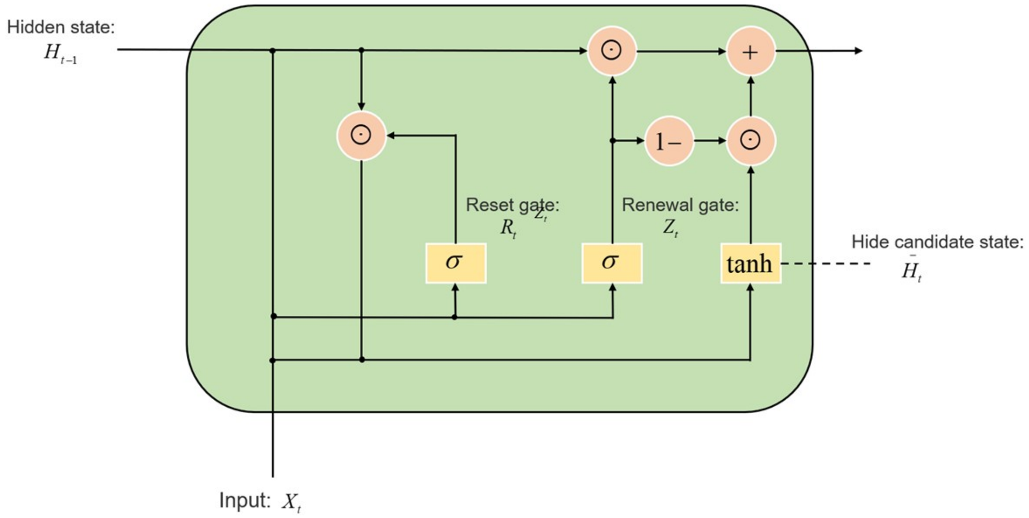

Figure S3. GRU Network Structure Diagram

The results of the coefficient paths obtained by Lasso regression are displayed in Figure S4.

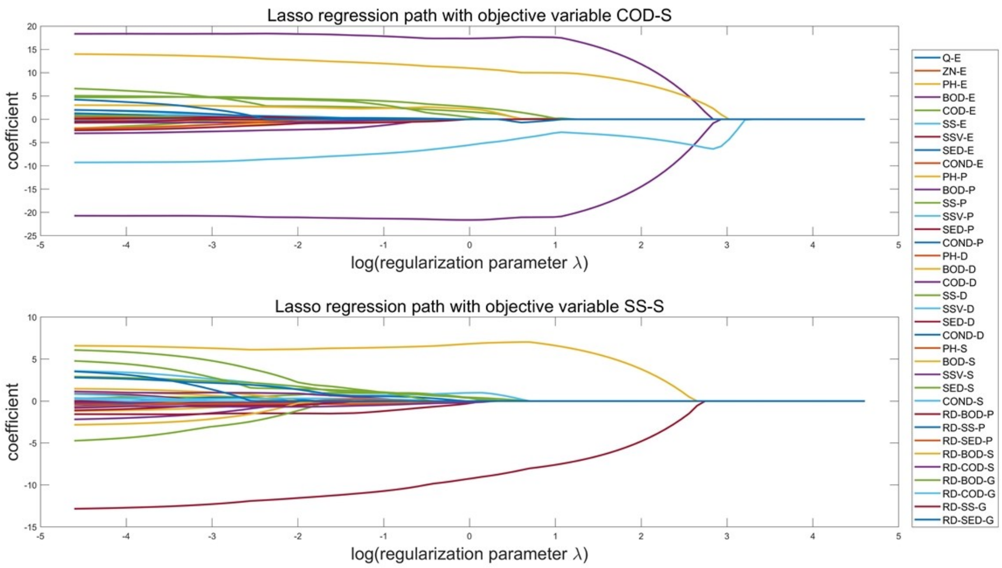

Figure S4. Path diagram of coefficients of Lasso regression

The specific numbers of R^2, RMSE, and MAPE statistics when different models predict COD are shown in Table S3.

Table S3. Comparison of different modeling methods for prediction of COD-S

Methods	Training set				Test set			
	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
Lasso-RF	0.91	11.34	4.95	0.07	0.86	14.30	7.72	0.11
Lasso-LSTM	0.97	6.18	5.29	0.05	0.91	11.06	5.29	0.07
Lasso-CNN	0.94	9.22	6.44	0.10	0.84	15.39	8.09	0.13
Lasso-GRU	0.92	10.78	8.88	0.14	0.87	13.99	11.02	0.18
Lasso-CNN-GRU	0.97	3.06	6.18	0.04	0.91	4.12	11.94	0.05
BO-CNN-GRU	**0.99**	**3.47**	**1.70**	**0.02**	**0.97**	**6.49**	**2.50**	**0.03**

The specific numerical statistics of R^2, RMSE, and MAPE when different models predict SS are listed in Table S4.

Table S4. Comparison of different modeling methods for prediction of SS-S

Methods	Training set				Test set			
	R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
Lasso-RF	0.76	8.66	3.62	0.15	0.62	7.27	5.00	0.27
Lasso-LSTM	0.76	8.52	5.14	0.23	0.71	6.39	4.48	0.24
Lasso-CNN	0.97	3.32	2.42	0.12	0.86	4.45	2.69	0.14
Lasso-GRU	0.95	3.81	2.84	0.14	0.80	5.30	3.47	0.19
Lasso-CNN-GRU	0.93	3.15	4.81	0.16	0.89	3.97	4.17	0.17
BO-CNN-GRU	**0.96**	**3.52**	**1.90**	**0.09**	**0.90**	**3.61**	**2.36**	**0.12**