Supporting Information

Hybrid data and knowledge driven approach for determining coagulant dosing in drinking water treatment plants

Dongsheng Wang[a*], Chuanzhuang Wang[a], Jiahao Liu[a], Yicong Yuan[a]

a College of Automation & College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

Corresponding author contact details: Dongsheng Wang
Email: wangdongsheng@njupt.edu.cn

This supporting information contains:
6 pages
2 Tables
1 Figures
3 Equations
2 Algorithms

List

Table S1 List of Abbreviations in Alphabetical Order

Description	Abbreviation
Autoencoder	AE
Artificial neural networks	ANN
Elastic weighted consolidation	EWC
Gated recurrent unit	GRU
Global attention	GA
Gated recurrent unit based GA	GRU_GA
Gated recurrent unit based LA	GRU_LA
Kolmogorov–Smirnov	KS
Local attention	LA
Long short-term memory.	LSTM
Multiple linear regression	MLR
Random forest	RF
Support vector machine	SVM

Table S2 KS test statistics for selected variables.

variables	KS	p-values
raw water turbidity (NTU)	0.427	1.305e-121
oxygen consumption (mg/L)	0.419	2.525e-117
pH	0.486	3.183e-157
conductivity (μs/cm)	0.521	9.241e-181
turbidity in plant (NTU)	0.572	7.658e-218
water flow rate (m3/h)	0.283	1.734e-53
coagulant dosage (mg/L)	0.581	7.982e-225
effluent turbidity (NTU)	0.237	1.079e-37
ammonia nitrogen (mg/L)	0.913	0
dissolved oxygen (mg/L)	0.790	0

Fig. S1 Distribution comparison of training and testing datasets.

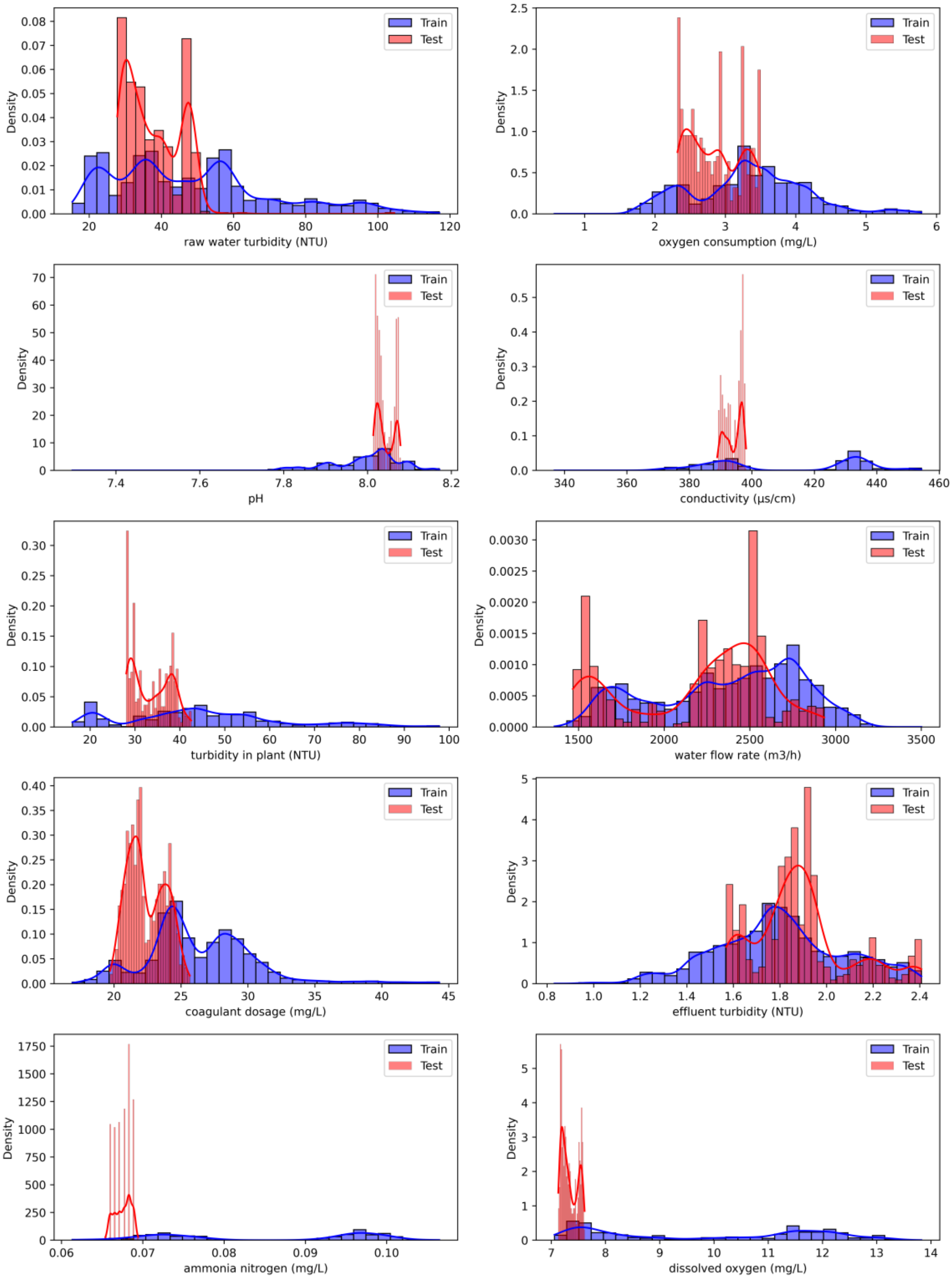

Fig S1 Distribution comparison of training and testing datasets.

Equation. S1

$$a_t = softmax(w_a h_{t-1} + b_a) \quad (S1)$$

Equation. S2

$$c_t = \sum_{i=t-L}^{t+L} a_{ti} x_i \quad (S2)$$

Equation. S3

$$h_t = GRU(c_t, h_{t-1}) \quad (S3)$$

Algorithm. S1 Reasoning Process of Expert System-based Hybrid Rule Algorithm

Algorithm. S1 Reasoning Process of Expert System-based Hybrid Rule Algorithm

1: **Input:** water quality status θ

2: **Output:** Coagulant adjustment C

3: Initialize a Hybrid Rule Base Ω

4: **if** $\Omega \neq \emptyset$ **then**

5:     **for** $k = 1$ to m **do**

6:         **if** $\Omega_k = \theta$ **then**

7:             **if** $\Omega_k \neq \emptyset$ **then**

8:                 Creating backtracking points P

9:             **else**

10:                 Adjustment of coagulant dosage C

7:             **if** $P \neq 0$ **then**

12:                 back points P

13:                 return step7

11:             **else**

13:                 return C

12:         **end for**

8:         **return** C

24: **end if**

Algorithm. S2 GRU_LA Algorithm

Algorithm. S2 GRU_LA Algorithm

Input: The water quality data series x

Output: The hidden state h_t

Initialize the weights of the GRU_LA model W_a and the dimensions of the sliding window L.

for t= $1,2,\cdots 6$ **do**

  Calculate attention weights a_t by(S1)

  Apply the attention weights a_t to the $[t-L,\ t+L]$ window of the input vector sequence x

  Calculate the local context vector c_t using(S2)

  Input the vector c_t and hidden state h_{t-1} into the GRU network model

  Calculate the next moment state h_t using(S3)

end for