Supplementary Material for

Neural network and random forest-based analyses of the performance of community drinking water arsenic treatment plants

Animesh Bhattacharya[a], Saswata Sahu[b], Venkatesh Telu[c], Srimanti Duttagupta[d], Soumyajit Sarkar[a], Jayanta Bhattacharya[e], Abhijit Mukherjee[a,f,*], Partha Sarathi Ghosal[b,*]

[a]School of Environmental Science and Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721 302, India

[b]School of Water Resources, Indian Institute of Technology Kharagpur, Kharagpur 721 302, India

[c]Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721 302, India

[d] Graduate School of Public Health, San Diego State University, CA, 92182

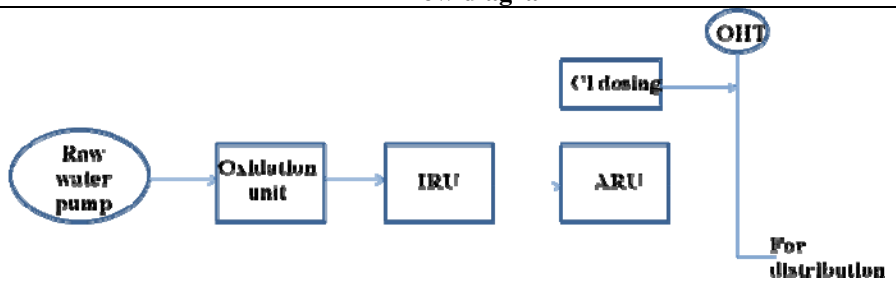
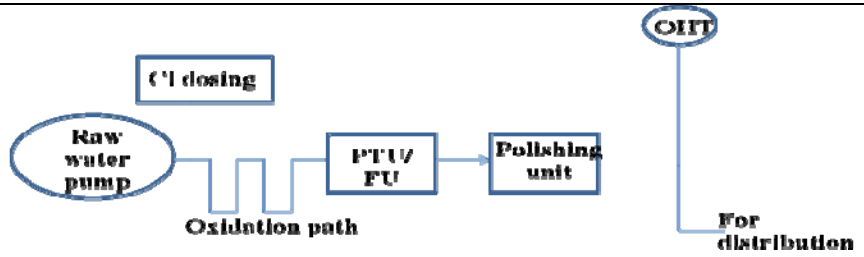
[e]Department of Mining Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721 302, India

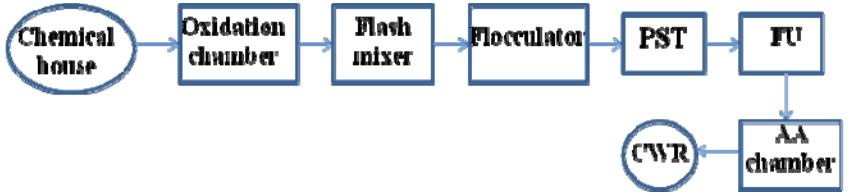
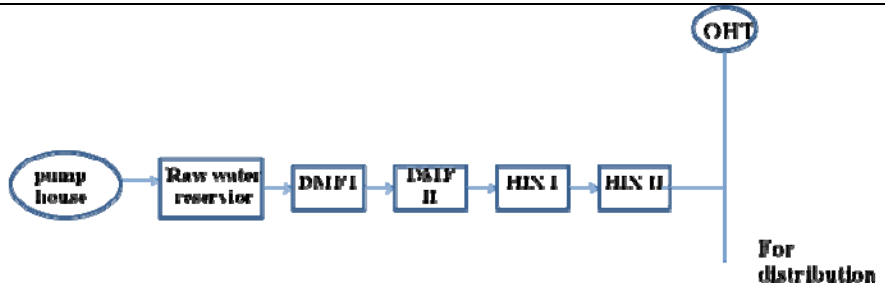
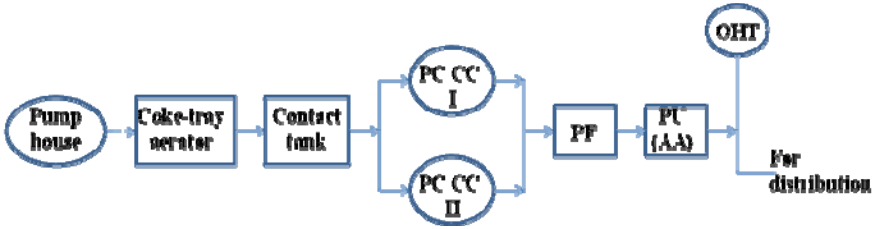
[f]Department of Geology and Geophysics, Indian Institute of Technology Kharagpur, Kharagpur 721 302, India

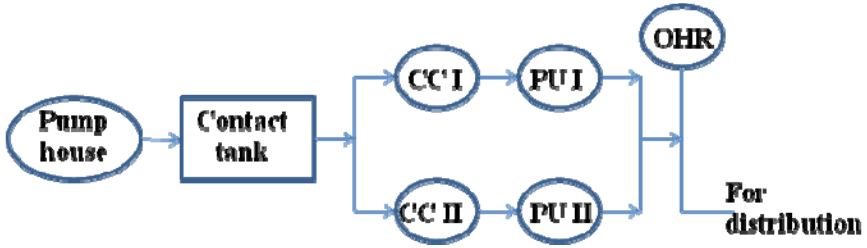
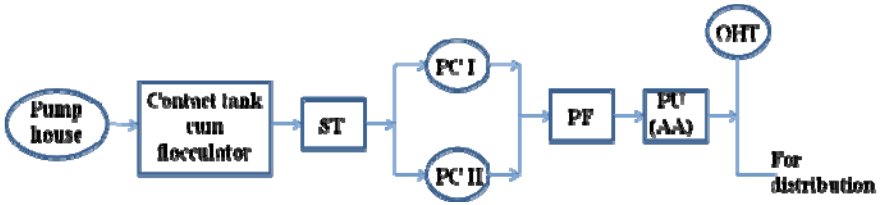
*Corresponding authors: Partha S. Ghosal(psghosal@swr.iitkgp.ac.in) and Abhijit Mukherjee (amukh2@gmail.com; abhijit@gg.iitkgp.ac.in)

Table S1:

Technical details of implemented schemes

Tech. type	Reagents used	AIRP flow diagram	scheme
8	Media-GFH Disinfectant- $Ca(ClO)_2$		(Arabpur PWSS, Doradaha PWSS, Gobindapur PWSS, Habibpur PWSS, Karimpur jalangi PWSS Zone 6-10, Narayanpur PWSS)
1	Media-AA Coagulant- $FeCl_3$ Oxidant-NaClO Disinfectant- $Ca(ClO)_2$		Goukhana Lalgola PWSS, Dier Shyampur PWSS, Komnagar PWSS, Ramchandrapur PWSS, Udaynagar PWSS, Ukhilnara WSS zone 2, Bhayna AIRP, Bansberia AIRP, Chandpur PWSS, Khanjepur WSS

1 and 2	Media-AA Oxidant- $Ca(ClO)_2$ Coagulant- $FeCl_3$ or Alum + $FeCl_3$ Disinfectant- NaClO or Cl		Barabil Raghunathpur Zone 1-2, Birnagar AIRP, Dhantala AIRP, Dubli 1 AIRP, Garibpur Zone 2, Muradpur Jalangi Zone 1, Purba Khamar Simulia PWSS, Ukhilnara WSS zone 1, Sonda PWSS, Pansila PWSS, Banpur WSS, Beharia WSS, Dhakhinpara zone 2, Fatepur WSS, Helenchi WSS zone 2-3, Itaberia WSS, Jaypur WSS, Nrishinghapur AIRP, Kakramari WSS, Mahishnengra WSS Zone 1, Monipota WSS
11	Media-HIX Oxidant- $Ca(ClO)_2$/ NaClO Coagulant- $FeCl_3$ Disinfectant- NaClO/ Cl		Ground Water Based Piped WSS for Digri & Its adj. mouzas, Ramnagar PWSS, Dubli PWSS, Taranipur WSS zone 2, Damodarkhali ground water based PWSS, Bajitpur Zone 1 WSS
1 and 3	Media-AA Coagulant- alum+$FeSO_4$/ $FeCl_3$ Oxidant-NaClO Disinfectant- $Ca(ClO)_2$		Tarapur WSS, Santoshpur WSS, Palla WSS, Bajitpur Zone2 WSS, Gazna WSS, Ichhlampur WSS Zone 2, Kalaighata PWSS, Maia WSS, Paschim Panchberia PWSS zone 2, Sabdalpur AIRP

10	Media-GFH Coagulant-$NaAlO_2$ Oxidant-Na_2S Disinfectant-NH_2Cl		Aturia PWSS, Jasaikathi PWSS Zone 2, Mollahati PWSS, Paschim Panchberia PWSS zone 1
1	Media-AA Coagulant- $FeCl_3$ Oxidant-NaClO or $Ca(ClO)_2$ Disinfectant- Cl		Dhalani WSS, Bahirgachhi WSS Zone 2, Taldah WSS, Hariharpur WSS, Kola PWSS

WSS - Water Supply Scheme, PWSS -Public Water Supply Scheme

Section S2:

Technology	Type
AA+$FeCl_3$	Tech 1
AA/alum+$FeCl_3$	Tech 2
AA/alum+$FeSO_4$	Tech 3
AA+alum	Tech 4
AA+$FeSO_4$	Tech 5
GFH/alum+$FeSO_4$	Tech 6
GFH+$FeCl_3$	Tech 7
GFH	Tech 8
GFH+Na_2S	Tech 9
GFH+$NaAlO_2$	Tech 10
HIX+$FeCl_3$	Tech 11
HIX+FeS	Tech 12
HIX+$NaAlO_2$	Tech 13

Section S3:

Dataset for box plot of efficiency of major technologies:

Type 1	Type 2	Type 3	Type 5	Type 8	Type 10	Type 11
65.455	88.889	87.143	84.463	88.75	83.333	79.31
60	87.654	76.238	54.455	64.286	73.545	100
100	89.796	73.333		92.593	71.429	98.319
48.837	85.714	71.429		88.571	97.159	96.875
69.277	67.308	80.62		80.392	58.871	86.486
76.19	50	90.141		80		84.337
81.944	87.5			92.308		
65.854	70.492			95		
85.263				64.286		
76.316				81.25		
77.1				80		
59.259						
58.182						
74.359						
75						
68.504						
43.827						
79.675						
50						
81.034						
80						

82.727
80.412
83.333
68.382
67.939
65.823
89
75.676
85.246
81.315
66.667
60.714
82.741
80.282
70
72.816
82.292
68.421
76.744
76.744
79.592
76.744
72.414
74.51
70.701
76.378
68.75
66.154
73.611
82.456
54.455
61.702
79.167
99.167
63.158
38.776
27.642
43.827
39.744

Section S4:

Error function for statistical analysis

Several error functions were used to evaluate the statistical significance of soft-computation models. The sum of the square of the error (SSE), sum of absolute error (SAE), average relative error (ARE), hybrid fractional error function (HYBRID), Marquardt's percent standard deviation (MPSD), chi-square test statistic (χ^2) are chosen as error functions and their mathematical expressions are as follows:

The sum of the square of the error (Kundu and Gupta, 2006)

$$SSE = \sum_{i=1}^{n} \left(Y_{pred} - Y_{\exp}\right)_i^2 \tag{1}$$

The sum of absolute error (Kundu and Gupta, 2006)

$$SAE = \sum_{i=1}^{n} \left|Y_{pred} - Y_{\exp}\right|_i \tag{2}$$

The mean square error (Kundu and Gupta, 2006)

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(Y_{pred} - Y_{\exp}\right)_i^2 \tag{3}$$

The mean absolute error (Tang et al., 2021)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left|Y_{pred} - Y_{\exp}\right|_i \tag{4}$$

The average relative error (Kundu and Gupta, 2006)

$$ARE = \frac{100}{n} \sum_{i=1}^{n} \left|\frac{Y_{pred} - Y_{\exp}}{Y_{\exp}}\right|_i \tag{5}$$

The hybrid fractional error function (Kundu and Gupta, 2006)

$$HYBRID = \frac{100}{n-p} \sum_{i=1}^{n} \left[\frac{\left(Y_{pred} - Y_{\exp}\right)^2}{Y_{\exp}}\right]_i \tag{6}$$

The Marquardt's percent standard deviation (Kundu and Gupta, 2006)

$$MPSD = 100\sqrt{\frac{1}{n-p}\sum_{i=1}^{n}\left[\frac{\left(Y_{pred}-Y_{\exp}\right)}{Y_{\exp}}\right]_i^2} \quad (7)$$

Where, n is the number of data points, and p is the number of parameters considered for modeling, n-p is the degree of freedom.

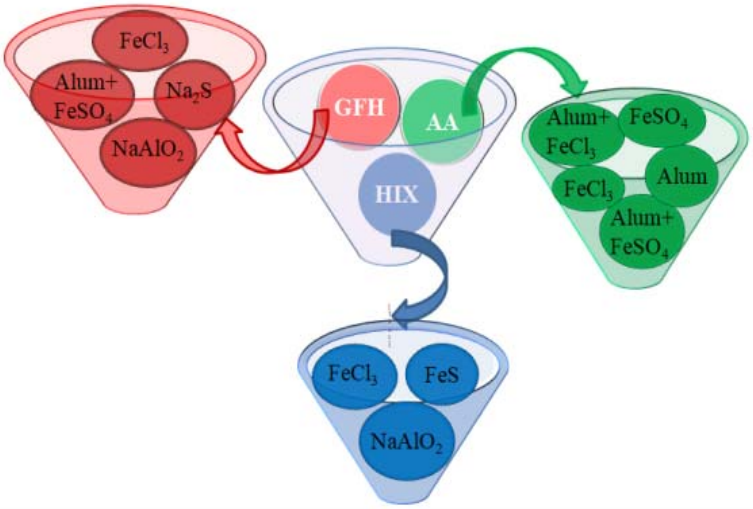


Figure S1: Combination of technologies concerning media and chemicals

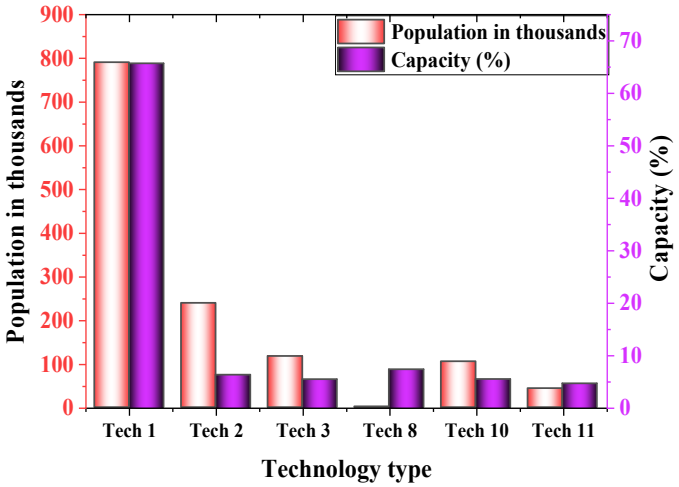


Figure S2: Contribution of major technologies to population and capacity (>90%)

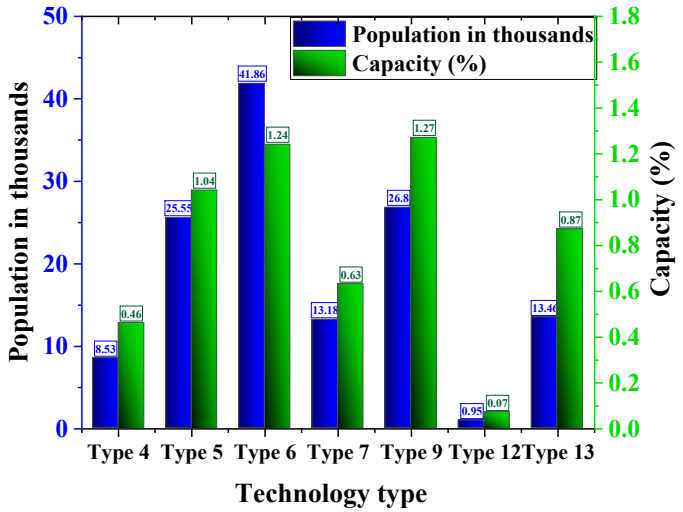


Figure S3: Contribution of minor technologies to population and capacity

(<10%)

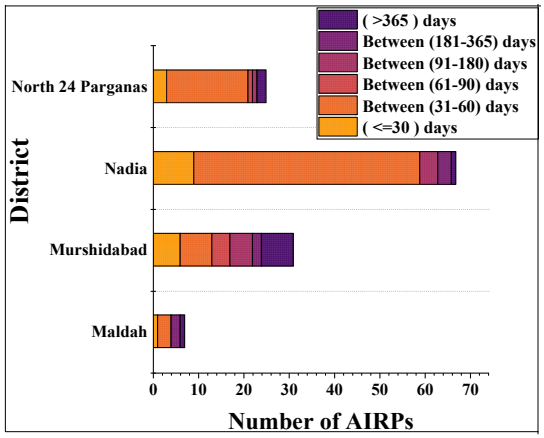


Figure S4: Frequency of sampling of effluent water in AIRPs

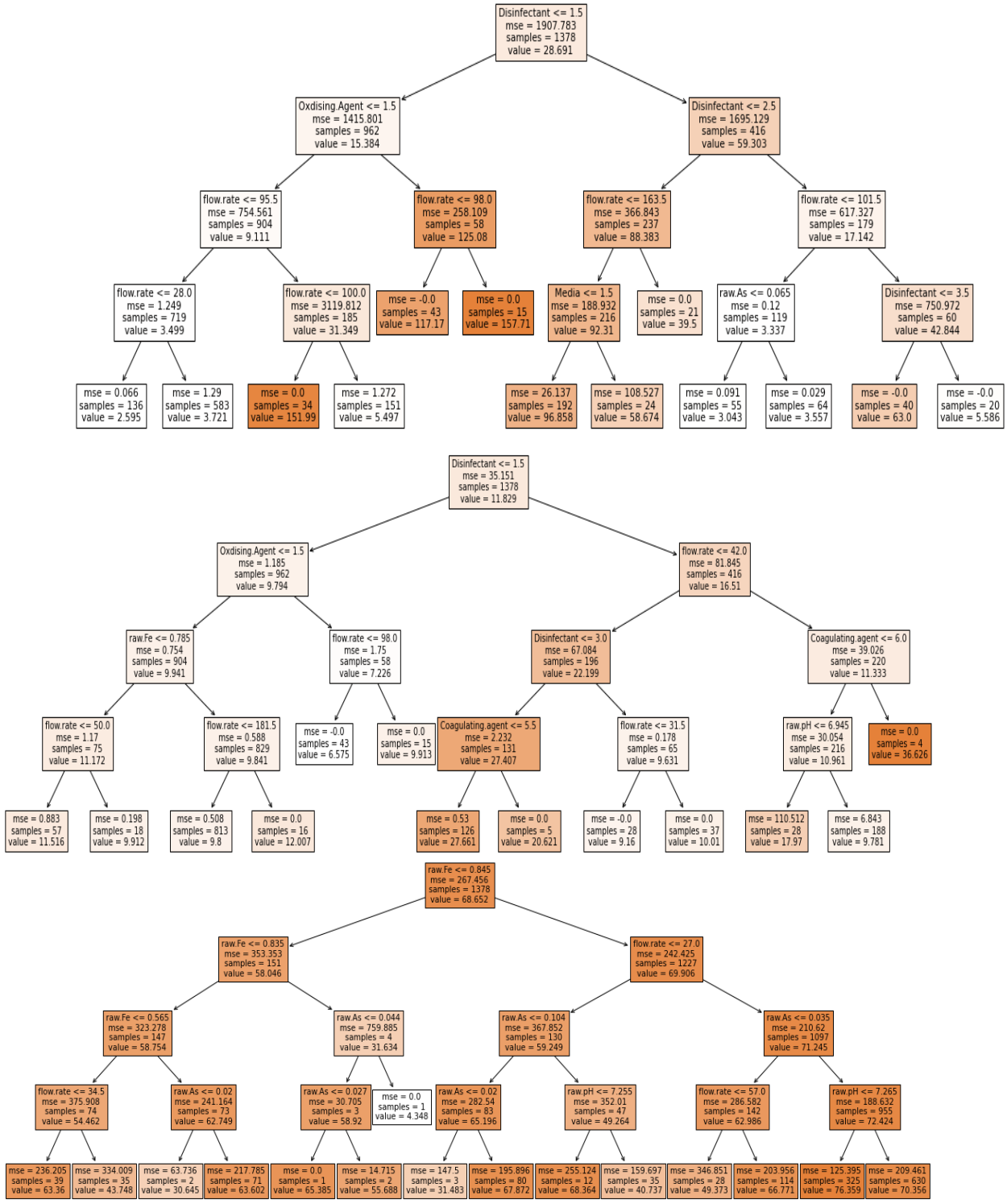


Figure S5: Random forest model trees for (a) capital cost, (b) operation and maintenance cost, (c) removal efficiency

Reference

Ayoob, S., Gupta, A.K., 2008. Insights into isotherm making in the sorptive removal of fluoride from drinking water. J. Hazard. Mater. 152, 976–985. https://doi.org/10.1016/j.jhazmat.2007.07.072

Kundu, S., Gupta, A.K., 2006. Arsenic adsorption onto iron oxide-coated cement (IOCC): Regression analysis of equilibrium data with several isotherm models and their optimization. Chem. Eng. J. 122, 93–106. https://doi.org/10.1016/j.cej.2006.06.002

Tang, M., Pérez-Fernández, R., De Baets, B., 2021. A comparative study of machine learning methods for ordinal classification with absolute and relative information. Knowledge-Based Syst. 230, 107358. https://doi.org/10.1016/J.KNOSYS.2021.107358