

Table S1 List of abbreviations.

Abbreviation	Full expression
CFU	Colony-Forming Unit
F	Fahrenheit
min	Minute
lbs/day	Pounds Per Day
KW	Kilowatt
MGD	Million Gallons per Day
MMCF	Million Cubic Feet
MPN	Most Probable Number
MW	Megawatt
MWh	Megawatt Hours
TPD	Tons per day
T	Ton

Table S2 Model performance metrics for 90:10 and 80:20 train-test splits.

Target Variable	Model	Set	(90:10)			(80:20)		
			MAE	R ²	RMSE	MAE	R ²	RMSE
BOD _i	ANN	Training	7.47	0.99	11.25	5.70	1.00	8.61
		Test	13.58	0.93	39.10	12.21	0.97	25.73
	GBM	Training	4.64	1.00	6.16	4.40	1.00	5.84
		Test	15.99	0.97	23.28	15.46	0.98	23.77
	RF	Training	13.22	0.98	20.74	14.18	0.98	21.54
		Test	32.20	0.89	48.95	38.83	0.86	57.84
	XGBoost	Training	4.84	1.00	6.37	4.24	1.00	5.47
		Test	14.43	0.98	22.52	15.40	0.98	23.01
RF-GBM	Training	6.10	1.00	7.69	2.87	1.00	3.71	
	Test	14.21	0.98	19.91	15.88	0.98	24.22	
BOD _e	ANN	Training	0.09	1.00	0.14	0.09	1.00	0.14
		Test	0.60	0.96	1.25	0.60	0.96	1.25
	GBM	Training	0.28	1.00	0.37	0.27	1.00	0.37
		Test	0.72	0.95	1.35	0.68	0.94	1.35
	RF	Training	0.45	0.99	0.77	0.48	0.99	0.79
		Test	1.27	0.86	2.36	1.18	0.86	2.07
	XGBoost	Training	0.33	1.00	0.45	0.29	1.00	0.39
		Test	0.74	0.95	1.36	0.69	0.94	1.36
RF-GBM	Training	0.21	1.00	0.28	0.30	1.00	0.38	
	Test	0.68	0.96	1.32	0.69	0.95	1.28	
(NH ₃) _i	ANN	Training	1.40	0.93	1.98	0.97	0.97	1.33
		Test	10.22	0.82	3.20	2.30	0.81	3.16
	GBM	Training	0.65	0.99	0.89	0.52	1.00	0.70

		Test	7.52	0.87	2.74	1.99	0.85	2.82
	RF	Training	0.74	0.98	1.03	0.75	0.98	1.03
		Test	8.17	0.85	2.86	2.09	0.84	2.91
	XGBoost	Training	0.41	0.99	0.57	0.61	0.99	0.83
		Test	7.02	0.87	2.65	1.91	0.86	2.70
	RF-GBM	Training	0.40	0.87	2.73	0.56	0.99	0.73
		Test	7.53	0.87	2.74	1.97	0.86	2.73
(NH ₃) _e	ANN	Training	0.16	0.98	0.23	0.11	0.99	0.16
		Test	0.53	0.33	1.06	0.46	0.35	0.89
	GBM	Training	0.16	0.98	0.25	0.24	0.95	0.36
		Test	0.48	0.60	0.82	0.47	0.45	0.82
	RF	Training	0.20	0.94	0.40	0.21	0.94	0.40
		Test	0.51	0.52	0.90	0.48	0.48	0.79
	XGBoost	Training	0.05	1.00	0.07	0.03	1.00	0.04
		Test	0.47	0.57	0.85	0.43	0.55	0.74
	RF-GBM	Training	0.18	0.97	0.26	0.22	0.96	0.31
		Test	0.52	0.50	0.91	0.46	0.42	0.84
P _i	ANN	Training	0.45	0.94	0.62	0.45	0.95	0.62
		Test	0.64	0.84	0.50	0.62	0.86	0.89
	GBM	Training	0.18	1.00	0.24	0.30	0.98	0.42
		Test	0.57	0.86	0.89	0.57	0.87	0.85
	RF	Training	0.25	0.98	0.40	0.23	0.98	0.33
		Test	0.56	0.86	0.89	0.57	0.87	0.85
	XGBoost	Training	0.29	0.98	0.38	0.26	0.98	0.34
		Test	0.55	0.85	0.90	0.58	0.87	0.87
	RF-GBM	Training	0.24	0.98	0.38	0.31	0.97	0.42
		Test	0.56	0.85	0.89	0.57	0.87	0.85
P _e	ANN	Training	0.11	0.73	0.17	0.10	0.74	0.17
		Test	0.14	0.42	0.18	0.13	0.44	0.19
	GBM	Training	0.01	1.00	0.01	0.03	0.99	0.04
		Test	0.10	0.65	0.14	0.11	0.55	0.17
	RF	Training	0.04	0.93	0.08	0.05	0.92	0.10
		Test	0.11	0.61	0.15	0.11	0.58	0.16
	XGBoost	Training	0.02	0.99	0.03	0.01	0.99	0.02
		Test	0.10	0.64	0.14	0.10	0.58	0.16
	RF-GBM	Training	0.01	1.00	0.01	0.01	1.00	0.01
		Test	0.10	0.65	0.14	0.10	0.60	0.16
TSS _i	ANN	Training	5.84	0.96	21.22	6.15	0.97	17.80
		Test	5.75	0.99	9.65	7.55	0.98	14.56
	GBM	Training	3.54	1.00	4.66	2.99	1.00	3.88
		Test	12.20	0.96	20.43	11.97	0.95	20.17
	RF	Training	9.46	0.97	17.13	11.22	0.96	22.12
		Test	26.71	0.82	41.55	23.06	0.85	36.11
	XGBoost	Training	3.17	1.00	4.06	3.19	1.00	4.12

		Test	11.18	0.97	17.69	10.82	0.96	18.31
	RF-GBM	Training	5.60	1.00	7.17	5.46	1.00	7.02
		Test	12.99	0.96	19.83	11.36	0.96	18.05
TSS _e	ANN	Training	0.21	1.00	0.28	0.15	1.00	0.20
		Test	0.52	0.95	0.91	0.56	0.94	0.96
	GBM	Training	0.17	1.00	0.23	0.03	1.00	0.04
		Test	0.38	0.97	0.70	0.33	0.97	0.71
	RF	Training	0.26	0.98	0.79	0.27	0.98	0.86
		Test	0.60	0.90	1.27	0.54	0.91	1.24
	XGBoost	Training	0.15	1.00	0.20	0.19	1.00	0.25
		Test	0.45	0.94	0.97	0.42	0.96	0.82
	RF-GBM	Training	0.16	1.00	0.21	0.08	1.00	0.10
		Test	0.40	0.97	0.71	0.41	0.96	0.83

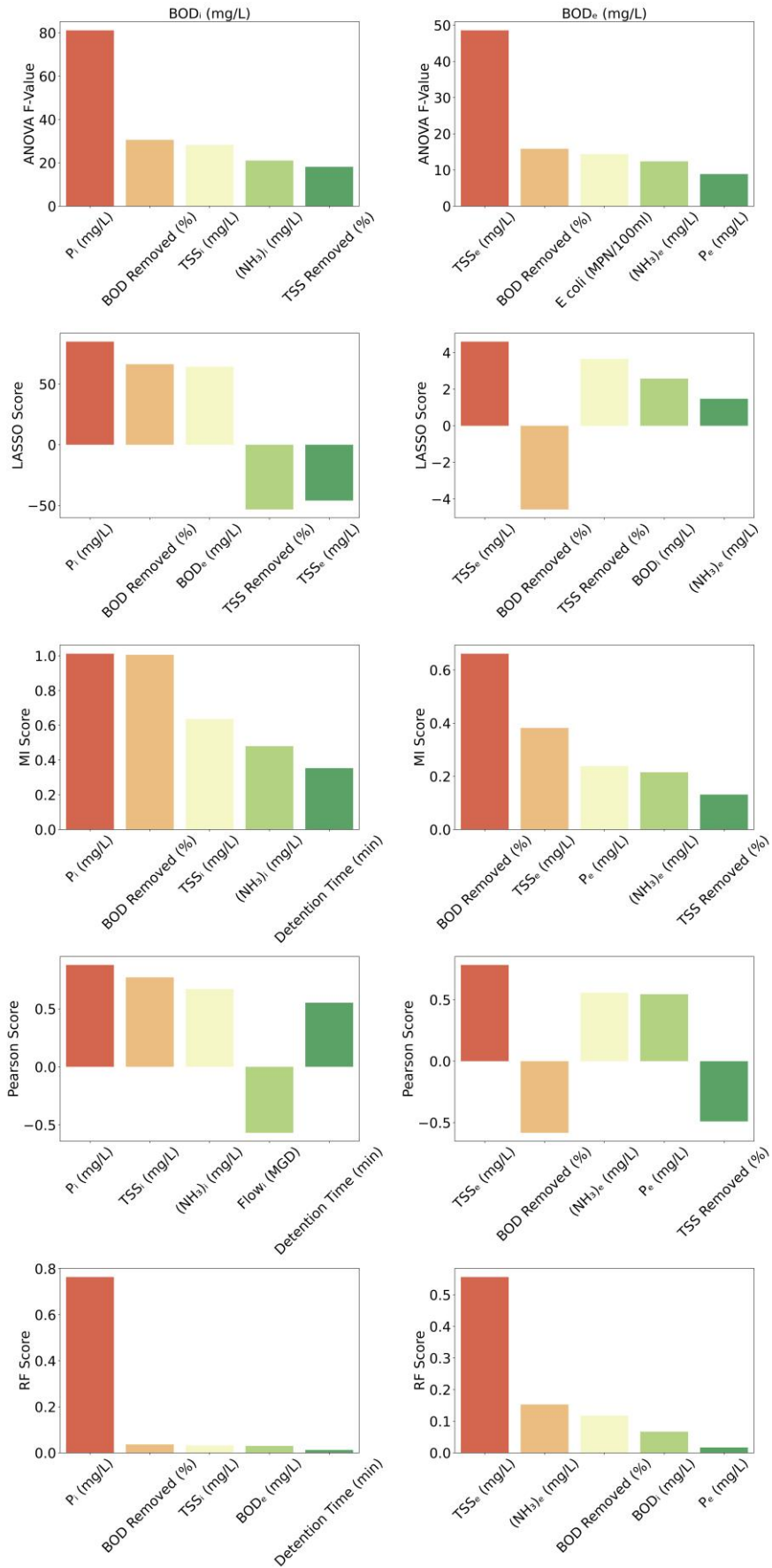


Figure S1 (i) Top five strongly correlated variables related to target variables.

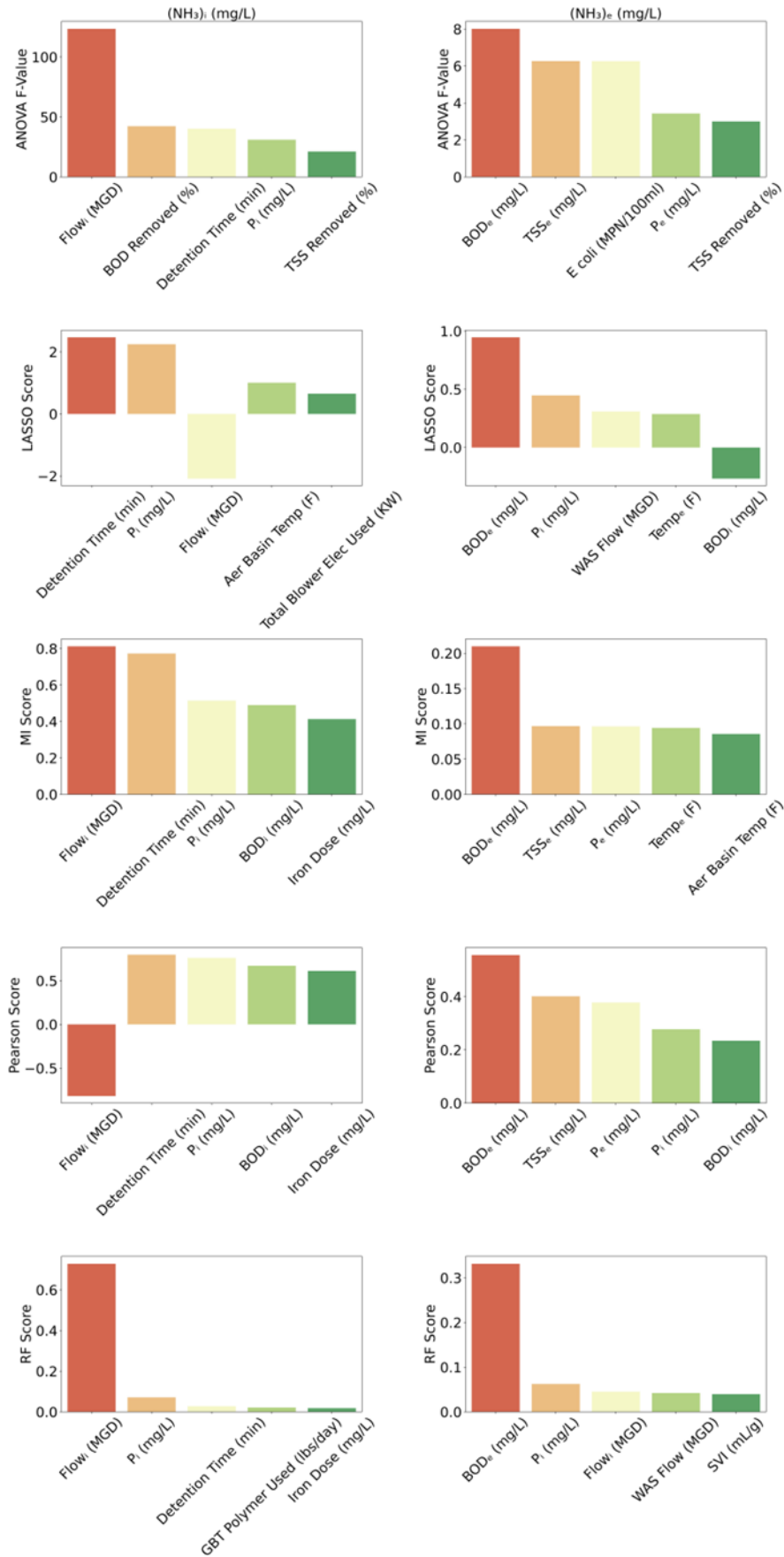


Figure S1 (ii) Top five strongly correlated variables related to target variables.

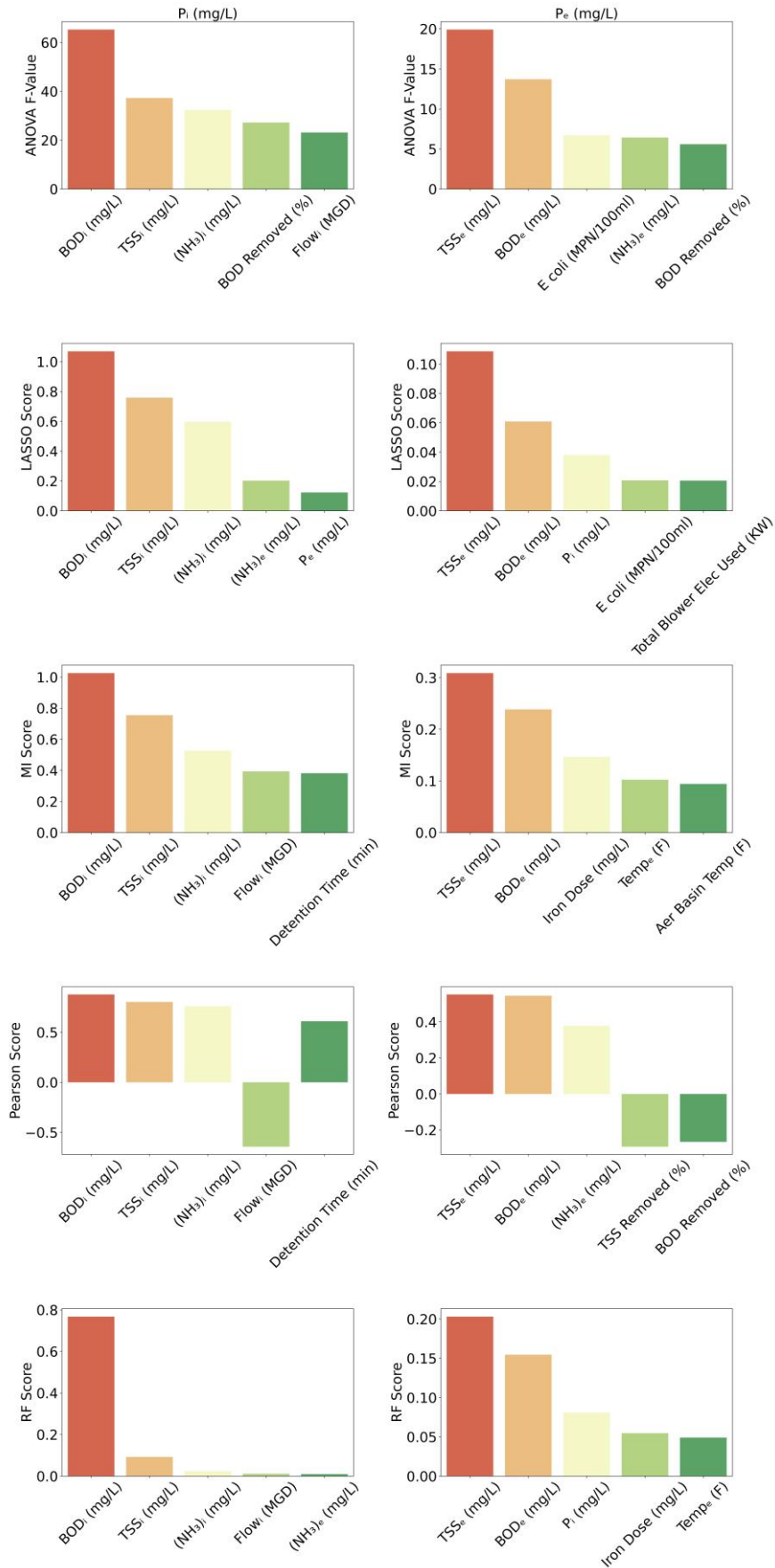


Figure S1 (iii) Top five strongly correlated variables related to target variables.

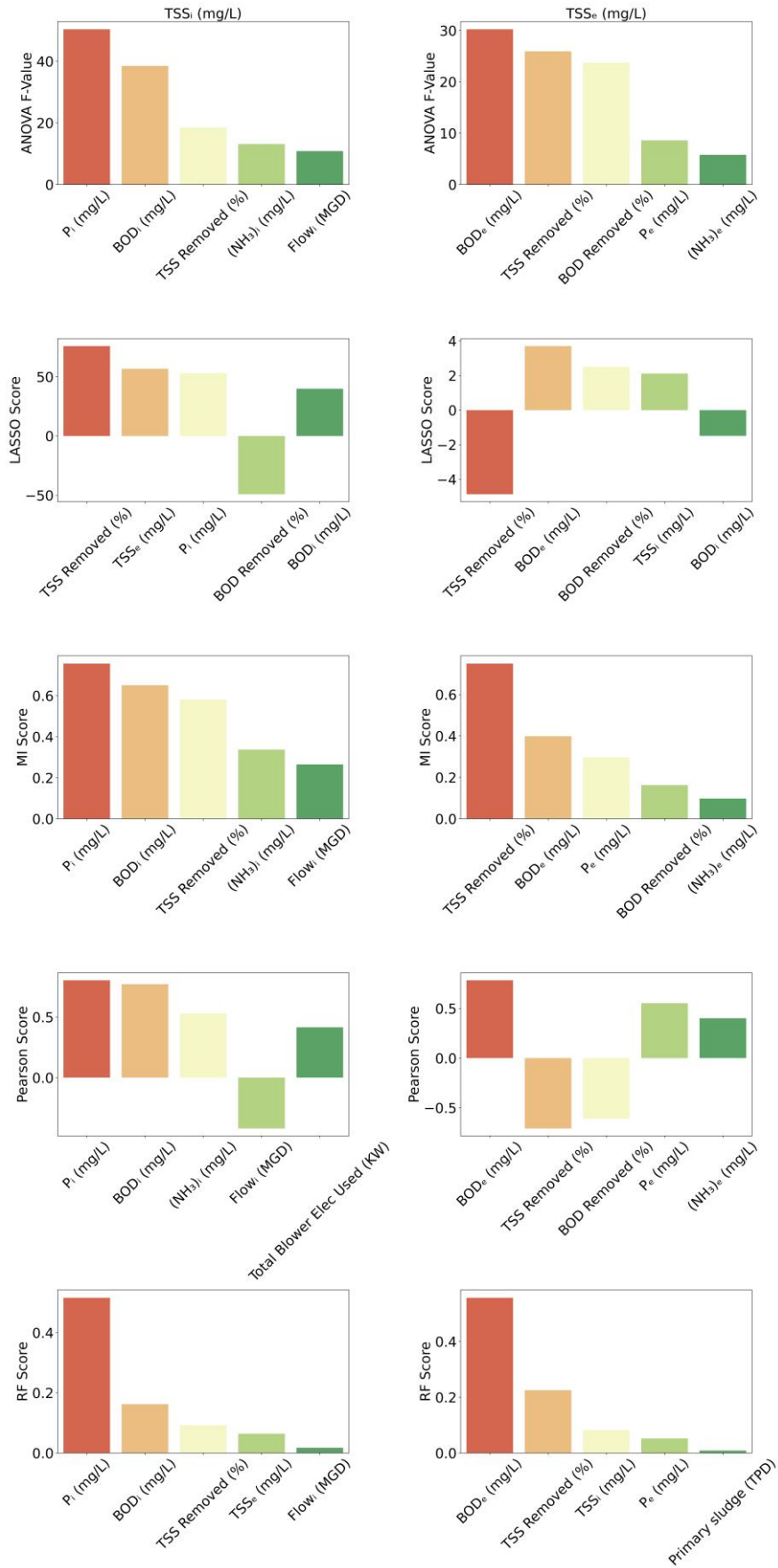


Figure S1 (iv) Top five strongly correlated variables related to target variables.

Table S3 LIME and SHAP selected top five influential variables in descending order of influence.

Target variable	Model	Influential variable (LIME)	Influential variable (SHAP)
BOD _i	ANN	BOD _e (-), P _i (-), BOD Removed (%) (+), GBT Polymer Used(+), (NH ₃) _e (+)	P _i , DO Set Pt, WAS Flow, pH _e , TRC
	GBM	BOD _e (+), BOD Removed (%) (-), P _i (-), TSS _i (-), WAS(-)	BOD _e , TSS _i , P _i , BOD Removed (%), Detention Time
	RF	P _i (-), BOD _e (+), TSS _i (-), BOD Removed (%) (-), Temp _e (-)	TSS _i , BOD _e , BOD Removed (%), P _i , DO Set Pt
	RF-GBM	BOD _e (+), BOD Removed (%) (-), P _i (-), TSS _i (-), MCRT(-)	BOD _e , TSS _i , P _i , BOD Removed (%), Detention Time
	XGBoost	BOD _e (+), BOD Removed (%) (-), P _i (-), TSS _i (-), Temp _e (-)	BOD _e , TSS _i , P _i , BOD Removed (%), Detention Time
BOD _e	ANN	TSS _e (+), TSS Removed (%) (-), Aer Basin Temp(-), BOD _i (-), BOD Removed (%) (-)	(NH ₃) _e , DO Set Pt, WAS Flow, pH _e , TRC
	GBM	TSS _e (+), BOD _i (-), (NH ₃) _e (-), Aer Basin Temp(-), BOD Removed (%) (+)	(NH ₃) _e , BOD _i , TSS _e , BOD Removed (%), Aer Basin Temp
	RF	TSS _e (+), BOD Removed (%) (+), (NH ₃) _e (-), BOD _i (-), P _e (+)	(NH ₃) _e , P _e , TSS _e , BOD Removed (%), BOD _i
	RF-GBM	TSS _e (+), (NH ₃) _e (-), BOD _i (-), BOD Removed (%) (+), Aer Basin Temp(-)	(NH ₃) _e , BOD _i , TSS _e , BOD Removed (%), E coli
	XGBoost	TSS _e (+), BOD _i (-), (NH ₃) _e (-), WAS Flow(+), Aer Basin Temp(-)	NH ₃) _e , BOD _i , TSS _e , BOD Removed (%), E coli
(NH ₃) _i	ANN	BOD _e (+), TSS Removed (%) (-), Flow _i (+), Aer Basin Temp(+), Total Blower Elec Used(+)	P _e , DO Set Pt, WAS Flow, pH _e , TRC
	GBM	Flow _i (+), P _i (-), Aer Basin Temp(+), Temp _e (+), GBT Polymer Used(-)	Flow _i , P _i , Iron Dose, Aer Basin Temp, GBT Polymer Used
	RF	Flow _i (+), P _i (-), Temp _e (+), Aer Basin Temp(+), GBT Polymer Used(-)	Flow _i , P _i , Iron Dose, Detention Time, GBT Polymer Used
	RF-GBM	Flow _i (+), P _i (-), GBT Polymer Used(-), Aer Basin Temp(+), Temp _e (+)	Flow _i , P _i , Iron Dose, Detention Time, GBT Polymer Used
	XGBoost	Flow _i (+), Aer Basin Temp(+), Temp _e (+), P _i (-), GBT Polymer Used(-)	Flow _i , P _i , Iron Dose, Detention Time, GBT Polymer Used
(NH ₃) _e	ANN	BOD _e (+), P _e (+), Temp _e (+), TSS _e (-), DO Set Pt(-)	Total Elec Generated, WAS Flow, pH _e , DO Set Pt, TRC
	GBM	BOD _e (+), TSS _e (+), E Coli(-), P _e (+), Temp _e (+)	BOD _e , SVI, E Coli, WAS Flow, Temp _e

	RF	BOD _e (+), TSS _e (+), E Coli(-), (NH ₃) _i (-), SVI(-)	SVI, E Coli, (NH ₃) _i , WAS Flow, BOD _e
	RF-GBM	BOD _e (+), TSS _e (+), P _e (+), E Coli(-), Temp _e (+)	SVI, E Coli, (NH ₃) _i , WAS Flow, BOD _e
	XGBoost	BOD _e (+), TSS _e (+), Temp _e (+), P _e (+), E Coli(-)	E Coli, Primary sludge, (NH ₃) _i , BOD _e , WAS Flow
P _i	ANN	BOD _i (-), P _e (+), TSS _i (-), BOD _e (-), (NH ₃) _i (+)	BOD Removed (%), P _e , DO Set Pt, TRC, pH _e
	GBM	BOD _i (-), (NH ₃) _i (+), TSS Removed (%)(-), TSS _i (-), P _e (+)	(NH ₃) _i , TSS Removed (%), TSS _i , Flow _i , BOD _i
	RF	BOD _i (-), TSS _i (-), (NH ₃) _i (+), P _e (+), GBT Polymer Used(+)	(NH ₃) _i , Flow _i , TSS _i , BOD _i , Detention Time
	RF-GBM	BOD _i (-), TSS _i (-), (NH ₃) _i (+), TSS Removed (%)(-), WAS Flow(+)	(NH ₃) _i , Flow _i , TSS _i , BOD _i , Iron Dose
	XGBoost	BOD _i (-), TSS _i (-), Temp _e (+), P _e (+), (NH ₃) _i (+)	(NH ₃) _i , Total Blower Elec Used, TSS _i , Flow _i , BOD _i
P _e	ANN	TSS _e (+), BOD _e (+), TSS Removed (%) (+), Temp _e (-), Total Blower Elec Used(+)	(NH ₃) _e , DO Set Pt, WAS Flow, pH _e , TRC
	GBM	TSS _e (+), Temp _e (-), BOD _e (+), TSS _i (-), Aer Basin Temp(+)	Temp _e , Iron Dose, TSS _e , Aer Basin Temp, BOD _e
	RF	TSS _e (+), BOD _e (+), Temp _e (-), Aer Basin Temp(+), TSS _i (-)	BOD _e , TSS _e , Aer Basin Temp, Temp _e , Iron Dose
	RF-GBM	TSS _e (+), Temp _e (-), BOD _e (+), Aer Basin Temp(+), TSS _i (-)	BOD _e , TSS _e , Iron Dose, Aer Basin Temp, Temp _e
	XGBoost	TSS _e (+), Temp _e (-), BOD _e (+), Aer Basin Temp(+), DO Set Pt (-)	BOD _e , TSS _e , Iron Dose, Aer Basin Temp, Temp _e
TSS _i	ANN	TSS Removed (%) (-), TSS _e (+), BOD _e (+), P _i (-), Total Blower Elec Used(+)	TSS Removed (%), DO Set Pt, WAS Flow, pH _e , TRC
	GBM	TSS _e (+), TSS Removed (%) (-), P _i (-), BOD _i (-), pH _e (-)	BOD _i , TSS _e , P _i , TSS Removed (%), Detention Time
	RF	TSS Removed (%) (-), TSS _e (+), P _i (-), BOD _i (-), Flow _i (-)	TSS _e , Flow _i , TSS Removed (%), BOD _i , P _i
	RF-GBM	TSS Removed (%) (-), TSS _e (+), P _i (-), BOD _i (-), GBT Polymer Used(-)	Flow _i , BOD _i , TSS _e , P _i , TSS Removed (%)
	XGBoost	TSS Removed (%) (+), TSS _e (-), P _i (-), BOD _i (-), (NH ₃) _e (+)	Flow _i , P _i , BOD _i , TSS _e , TSS Removed (%)
TSS _e	ANN	P _e (+), TSS Removed (%) (+), BOD _e (+), (NH ₃) _e (-), Temp _e (-)	TSS Removed (%), DO Set Pt, TRC, P _e , pH _e
	GBM	TSS Removed (%) (+), BOD _e (+), TSS _i (+), P _e (-), (NH ₃) _i (-)	P _e , TSS _i , TSS Removed (%), BOD _e , Aer Basin Temp
	RF	TSS Removed (%) (+), BOD _e (+), TSS _i (-), P _e (+), TRC(+)	P _e , TSS _i , TSS Removed (%), BOD _e , Aer Basin Temp
	RF-GBM	TSS Removed (%) (+), BOD _e (+), TSS _i (-), P _e (+), TRC(+)	P _e , TSS _i , TSS Removed (%), BOD _e , Aer Basin Temp
	XGBoost	TSS Removed (%) (+), BOD _e (+), TSS _i (-), P _e (+), Fecal Coliforms(+)	P _e , TSS _i , TSS Removed (%), BOD _e , BOD Removed (%)