JPIChE 53(2) 2025: 47-55

Journal of the Pakistan Institute of Chemical Engineers

Journal homepage: www.piche.org.pk/journals

DOI: https://doi.org/10.54693/piche.05325

AI-Driven Prediction of Sulfide Removal from Wastewater through Aeration and Iron Dosing

[1]F. Ahmad

Submitted: 16/12/2025, Accepted:18/02/2026, Published: 17/03/2026

Abstract

Wastewater containing sulfide needs to be segregated to save the environment and preserve the quality of water. Precise prediction of sulfide is important to monitor the environment, the exploration of minerals, and industrial operations. In this study, five widely used machine learning (ML) models namely, XGBoost, Artificial Neural Networks (ANN), Support Vector Regression (SVR), Random Forest (RF), and AdaBoost, using the published data of suldife removal by aeration and iron salts. XGBoost was selected as the best model (RMSE: 160.86, R^2: 0.973) and was much better than others due to hyper parameter optimization and extensive analysis (RMSE, MAE, R^2). The correlation study of features indicated that there are fundamental input relationships that dictate the behavior of sulfide. The optimized XGBoost structure achieves highest predictive accuracy. The results explain the potential application of machine learning models to optimize waste water treatment parameters and increase efficiency.

Keywords: Wastewater Treatment; Sulfide Removal; Machine Learning, Artificial Neural Network, XGBoost Model

1. Introduction:

The problem of sulfide in wastewater is widespread and has extensive implications on the sustainability of the environment and human well-being[1]. The product of industrial activities such as petrochemical refining, automotive, oil and gas, and the anaerobic degradation of organic materials, sulfides, especially hydrogen sulfide (H_2S) are multi-faceted problems[1, 2]. These are odorative wastes, corrosion of infrastructure, harmfulness to aquatic organisms, and possible risks to human health [1-3]. The successful elimination of sulfide in wastewater is thus not just a regulatory requirement but a necessary need as a measure to not only protect ecological integrity but also the welfare of a community [1].

Traditional methods of removing sulfide include physical, chemical and biological methods[1]. Physical processes such as aeration seek to remove H_2S in the liquid phase[4, 5]. Chemical processes include oxidation or precipitation using chlorine, ozone, or iron salts [1]. Biological processes involve the use of microorganisms to reduce the toxicity of sulfides into non-toxic compounds such as sulfur or sulfate [6]. All the methods have their strengths and limitations with regard to concentration of sulfide, the composition of wastewater, and cost[1]. Aeration and iron dosing are commonly employed techniques for sulfide removal from waste water **[4, 7]**. Sulfur is oxidized through aeration, and the insoluble iron sulfide is precipitated through the application of iron dosing[1]. Aeration combined with the dosing of iron may increase the removal of sulfides [1]. Nevertheless, the optimization of these processes is a complicated process, depending on such factors as the rate of aeration, pH, temperature, and iron dosage[7]. The complexity of these variables requires a thorough knowledge of the underlying processes to achieve a maximum removal of sulfides with minimum costs and side effects [7, 8]. Machine learning (ML) has shown a

[1] Department of Chemical and Materials Engineering, College of Engineering, Northern Border University, Arar, Kingdom of Saudi Arabia
Corresponding author: Farooq Ahmad **(**faroog.amin@nbu.edu.sa**)**

bright future of optimizing the process of wastewater treatment[9]. ML algorithms have the ability to process bulk data and find the correlation between input variables and output variables, which allow decision-making and control of the process through prediction **models [19-25].** ML has been proven to be successful in wastewater treatment such as predicting the effluent quality, optimal chemical dosage, and detection of anomaly [10]. Nevertheless, the predictive capability of ML applied to forecast the sulfide removal through aeration and iron dosing is not well explored [10]. In this research, this gap is filled by establishing ML models to predict the sulfide removal with the help of aeration and iron dosing [6, 7]. Through the use of published data to train the ML models, the efficiency of sulfide removal can be forecasted as a function of the aeration rate, iron dosage, pH, temperature, and starting concentration of sulfides [11]. These models have the capability of streamlining aeration and dosing of iron, enhancing removal of sulfides, a reduction in the use of chemicals and stabilizing of the process [6, 7]. There is also the ability of ML models to offer insights into the significance of the operational parameters and their interaction. Several popular machine-learning models were chosen to show how the system's nonlinear and complicated interactions work. We used XGBoost, Random Forest, and AdaBoost to test how well ensemble and boosting methods work when they include built-in regularization and are resistant to overfitting. We included Support Vector Regression with an RBF kernel since it has a solid theoretical basis and can handle nonlinear regression with little data. We used an Artificial Neural Network (MLP) to model very complicated input-output interactions by approximating nonlinear functions. These models show how ensemble, kernel, and neural network-based regression methods compare to each other [20-23].

The study will come up with precise ML models to forecast the presence of sulfide removal in wastewater through aeration and iron dosing to enhance sustainable practices of wastewater treatment [10]. The models will be tested to make sure they will be applicable and reliable[12, 13]. The findings will inform the operators of wastewater treatment plants to optimize sulfide protraction and obtain regulatory compliance[13, 14]. The originality of the present work is that it uses machine learning to maximize sulfide removal and passes through the limitations of traditional techniques [10, 14]. This study provides a data-driven solution to increase the efficiency and sustainability of wastewater treatment by involving ML in the strategy to improve the efficiency of the aeration and iron dosing processes [11]. These findings will be applied to develop intelligent water systems, leading to improved wastewater management practices [15-18].

2. Methodology:

The workflow involved, compiling a dataset from published sulfide-removal experiments [4, 19-22], engineering features and pre-processing, (3) training five ML regression models with hyper parameter tuning, and (4) evaluating performance. Key steps are summarized below. The flow diagram of the methodology is shown in Fig.1.

2.1. Data Collection and Preprocessing:

A composite dataset was assembled from literature-reported experiments on sulfide removal by aeration and iron salts (ferrous/ferric). Each data point comprises features such as initial sulfide concentration, aeration time (or airflow rate), iron dosage, initial pH, temperature, and dissolved oxygen, with the target variable being either sulfide removal efficiency (%) or residual sulfide concentration. Prior to modeling, the dataset was cleaned and standardized to zero mean and unit variance. To evaluate potential multicollinearity among input features, a Pearson correlation matrix was calculated, as depicted in Fig. 2. Strong linear correlations were observed between several variables such as dissolved oxygen showed a strong positive correlation with temperature and a strong negative correlation with initial sulfide concentration. Likewise, iron dosage showed moderate to strong correlations with both initial sulfide and final sulfide levels. Despite these interdependencies, such as dissolved oxygen showed a strong positive correlation with

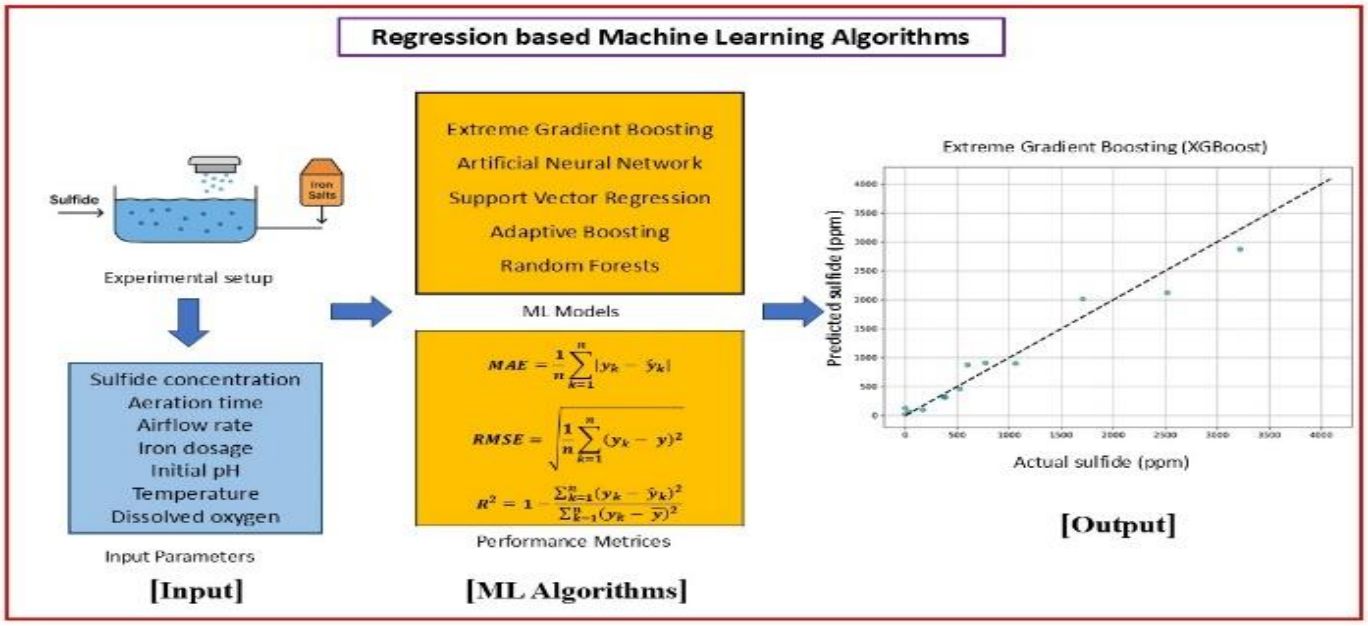

Figure 1: Flow diagram of the methodology

temperature and a strong negative all features were retained for modeling since tree-based algorithms are generally robust to multicollinearity. The data set are shown in Table 01.

Table 1: Data set used in the methodology

Process	Initial Sulfide (mg)	Fe2_Fe3	Volume_ml	Time_min	PH	Temperature (C^0)	DO_ppm	Air_Flow_Lmin	Sulfate_ppm	Sulfide_ppm (FeS)
Precipitation	2400	0.5	500	25	5.1	27	0	0	0	1825
Precipitation	2400	0.5	500	50	4.6	27.9	0	0	0	1447
Precipitation	2400	0.5	500	75	4.32	28	0	0	0	1080
Precipitation	2400	0.5	500	100	4.22	28.4	0	0	0	771
Precipitation	2400	0.5	500	125	3.99	27.9	0	0	0	600
Precipitation	2400	0.3	1000	30	5.13	27.1	0	0	0	2210
Precipitation	2400	0.3	1000	60	4.91	27.9	0	0	0	1870
Precipitation	2400	0.3	1000	90	4.43	28.2	0	0	0	1490
Precipitation	2400	0.3	1000	120	4.2	28.4	0	0	0	981
Precipitation	2400	0.3	1000	150	3.88	27.9	0	0	0	750
Precipitation	2400	0.8	1000	20	5.2	27	0	0	0	1705
Precipitation	2400	0.8	1000	40	4.97	26.8	0	0	0	1557
Precipitation	2400	0.8	1000	60	4.61	26.4	0	0	0	1378
Precipitation	2400	0.8	1000	80	4.3	25.9	0	0	0	1061
Precipitation	2400	0.8	1000	100	4.01	25.4	0	0	0	697
Precipitation	2400	0.8	1000	120	3.93	25.4	0	0	0	520
Aeration	600	0.5	1000	20	4.1	26.9	7.98	4	3625	561
Aeration	600	0.5	1000	40	4.17	26.4	8.23	4	3690	492
Aeration	600	0.5	1000	60	4.23	25.8	8.2	4	3701	369
Aeration	600	0.5	1000	80	4.32	25.2	8.01	4	3759	258

Aeration	600	0.5	1000	100	4.35	25	8.71	4	3850	102
Aeration	600	0.5	1000	120	4.39	25	8.8	4	3990	0
Aeration	750	0.3	1000	20	4.05	25.4	7.71	4	2753	671
Aeration	750	0.3	1000	40	4.15	25	7.81	4	2890	522
Aeration	750	0.3	1000	60	4.23	24.7	8.2	4	2950	382
Aeration	750	0.3	1000	80	4.32	24.7	8.1	4	3110	180
Aeration	750	0.3	1000	100	4.37	24.5	7.9	4	3220	23
Aeration	750	0.3	1000	120	4.41	24.3	8.31	4	3330	0
Aeration	520	0.8	1000	18	4.04	25.4	7.59	4	4500	452
Aeration	520	0.8	1000	36	4.11	24.9	7.85	4	4890	371
Aeration	520	0.8	1000	54	4.26	24.4	8.32	4	5120	262
Aeration	520	0.8	1000	72	4.36	24.1	8.1	4	5490	169
Aeration	520	0.8	1000	90	4.39	24	7.9	4	5890	37
Aeration	520	0.8	1000	100	4.46	24	8.23	4	6010	0
Aeration	600	0.5	1000	30	4.07	26.8	5.8	2	3605	560
Aeration	600	0.5	1000	60	4.13	26.2	6.01	2	3630	491
Aeration	600	0.5	1000	90	4.19	25.7	6.15	2	3675	432
Aeration	600	0.5	1000	120	4.23	25.4	6.29	2	3720	370
Aeration	600	0.5	1000	150	4.29	25.4	5.91	2	3785	290
Aeration	600	0.5	1000	180	4.31	25.2	5.72	2	3830	201
Aeration	600	0.5	1000	210	4.34	25.1	5.98	2	3892	117
Aeration	600	0.5	1000	240	4.37	24.8	6.08	2	3922	27
Aeration	600	0.5	1000	255	4.41	24.6	6.23	2	4080	0
Aeration	600	0.5	1000	15	4.1	26.8	8.3	8	3790	520
Aeration	600	0.5	1000	30	4.14	26.2	8.41	8	4233	389
Aeration	600	0.5	1000	45	4.18	25.9	8.15	8	4590	241
Aeration	600	0.5	1000	60	4.26	25.7	7.9	8	5198	31
Aeration	600	0.5	1000	75	4.31	25.4	8.1	8	5691	0
Aeration	1328	0.5	1000	30	4.25	25.4	6.7	4	7467	1197
Aeration	1328	0.5	1000	60	4.29	25.1	7.01	4	9771	971
Aeration	1328	0.5	1000	90	4.36	24.9	7.92	4	8548	812
Aeration	1328	0.5	1000	120	4.39	24.7	7.2	4	9200	603
Aeration	1328	0.5	1000	150	4.41	24.4	6.91	4	9711	491
Aeration	1328	0.5	1000	180	4.43	24.2	6.63	4	10216	356
Aeration	1328	0.5	1000	210	4.47	24.1	8.07	4	10971	231
Aeration	1328	0.5	1000	240	4.49	24	8.21	4	11211	120
Aeration	1328	0.5	1000	270	4.51	24	8.35	4	11830	23
Aeration	1328	0.5	1000	290	4.53	24	8.49	4	12451	0
Precipitation	4800	0.5	1000	35	5.31	27.2	0	0	0	4089
Precipitation	4800	0.5	1000	70	5.1	28	0	0	0	3217
Precipitation	4800	0.5	1000	105	4.8	28.4	0	0	0	2517
Precipitation	4800	0.5	1000	140	4.57	27.9	0	0	0	1967
Precipitation	4800	0.5	1000	175	4.39	27.2	0	0	0	1635
Precipitation	4800	0.5	1000	210	4.18	26.8	0	0	0	1328

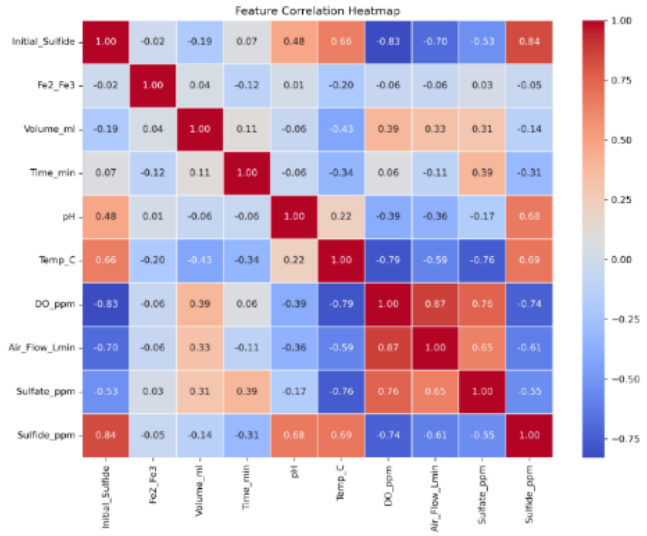

Figure 2: Feature Correlation Heat map

2.2. Feature Engineering:

In addition to raw inputs, we created composite features, such as the Fe:S molar ratio and dissolved oxygen contact time, to better represent known chemical interactions. While categorical and ordinal features were prepared for encoding, none were utilized in this analysis. The final set of features underwent exploratory analysis, confirming that higher Fe:S ratios are strongly correlated with improved removal rates, consistent with the ~1.3:1 molar guideline. Feature importance analysis from an initial XGBoost run further highlighted that iron dose and aeration duration were the most significant predictors.

2.3. Model selection and training: We implemented five supervised regression models known for nonlinear modeling:

- **XGBoost (eXtreme Gradient Boosting):** A regularized tree boosting algorithm
- **Random Forest (RF):** An ensemble of decision trees with bootstrap aggregation.
- **AdaBoost:** Adaptive boosting with decision-tree base learners.
- **Support Vector Regression (SVR):** Kernel-based regression (we used RBF kernel).
- **Artificial Neural Network (ANN):** A feedforward multilayer perceptron (MLP) with backpropagation.

For each model, hyper parameters were tuned via grid search with 5-fold cross-validation. Example tuned parameters included number of trees and max depth (RF, XGBoost), learning rate (AdaBoost, XGBoost), and SVR's C and ε. The best hyper parameters are summarized in Table 2 below. We reserved an unseen test subset (20% of data) for final evaluation to assess generalization.

2.4. Evaluation Metrics:

Root-mean-square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) between predicted and actual removal of sulfide were used to measure the model accuracy. The greater the R^2 and the lesser the RMSE/MAE the better. These are regression standardized measures. The general goal is to reduce error-based measure (MAE and RMSE) and maximize R^2. A smaller value of error measures is associated with improved consistency of the forecast and the real values hence, indicating a stronger and generalized model. The mathematical formulas for computing the values of MSE, RMSE, and R^2 are provided in equations (1), (2), and (3) [23].

$$MAE = \frac{1}{n}\sum_{k=1}^{n}|y_k - \hat{y}_k| \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - y)^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(y_k - \hat{y}_k)^2}{\sum_{k=1}^{n}(y_k - \overline{y})^2} \quad (3)$$

Table 2: Optimized hyperparameter using trained ML models

Model	Optimized Hyperparameters
XGBoost	n_estimators=185 max_depth=4, learning_rate=0.163, subsample=0.724 colsample_bytree=0.849 gamma=2.458 reg_alpha=2.777 reg_lambda=0.679
ANN	Number of hidden layers=3 n_units = [64, 112, 64] dropout = [0.069, 0.159, 0.300] lr=0.0068 batch_size=8
SVR	C=965.26

	epsilon=0.923 gamma='scale'
RF	n_estimators=178 max_depth=12 min_samples_split=4 min_samples_leaf=3 max_features='auto'
AdaBoost	n_estimators=83, max_depth=9, learning_rate=0.517

3. Results and Discussion:

3.1. Model Performance:

Table 3 shows how well five machine-learning models (XGBoost, ANN, SVR, RF, and AdaBoost) can predict how well sulfide can be removed. It does this by utilizing RMSE, MAE, and R^2. XGBoost was the best model overall, with the lowest RMSE (160.86), MAE (111.24), and R^2 (0.973). This framework, which includes built-in regularization, makes it possible to learn complicated nonlinear connections while yet being able to generalize well. XGBoost has also been shown to be the best choice for modeling wastewater, including predicting the performance of sludge and treatment plants. The ANN model came in second, with a high R^2 of 0.944 but more errors (RMSE: 230.33, MAE: 159.75). Even while ANNs can handle nonlinear dynamics, their performance depends on how they are built and how their hyperparameters are configured, which might make them less accurate with noisy wastewater datasets. This trend aligns with prior environmental modeling research indicating that ANN was competitive however outperformed by ensemble tree-based methodologies.

SVR showed moderate performance, reflecting its dependence on kernel selection and parameter optimization. While effective for smaller or smoother datasets, SVR may struggle to fully capture the operational variability and nonlinearities inherent in sulfide removal processes, as also observed in related wastewater applications. The Random Forest model provided stable but lower accuracy than boosting-based methods. Its independent tree construction improves robustness but limits its ability to iteratively correct prediction errors, explaining its inferior performance relative to XGBoost. AdaBoost performed the weakest among all models, likely due to its sensitivity to noise and outliers, which are common in wastewater datasets.

Figure 3 (top) and bottom show the comparison of actual and predicted sulfide values for Training and Testing of XGBoost respectively. Figure 4 shows that XGBoost is quite good at making predictions. The projected values are very near to the observed values along the 1:1 line, which backs up its high R^2 and low error metrics.

Table 3: Performance of ML models on predicting sulfide removal (test set). Lower RMSE/MAE and higher R^2 are better.

Model	RMSE (ppm)	MAE (ppm)	R^2
XGBoost	160.86	111.24	0.973
ANN	230.33	159.75	0.944
SVR	266.94	165.29	0.925
RF	383.13	253.94	0.845
AdaBoost	411.72	225.69	0.822

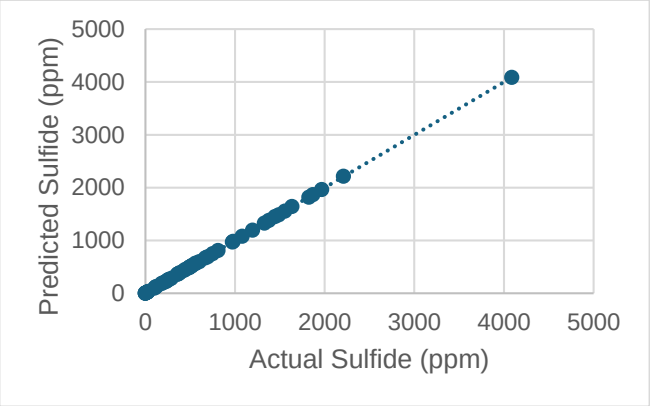

Figure 3: Regression curve for XgBoost model both train and test set

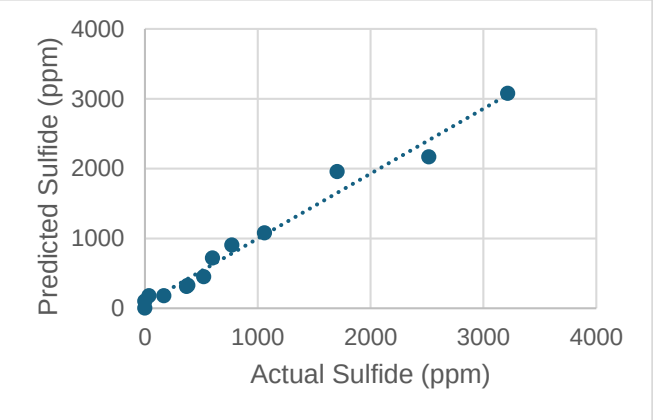

Figure 4: Scatter plot comparing the predicted versus actual values of sulfide removal efficiency using XGBoost model

3.2. Interpretation of Results:

The iron dosage and aeration dominate in feature analysis. Our correlation analysis showed that the Fe dose and the aeration time had a strong positive relationship with the removal efficiency, which is consistent with chemical law (the more the Fe, the more the oxidation/precipitation). To give an example, raising the Fe:S molar ratio above its optimal level of about 1:1 gave diminishing returns, which are a reflection of the 1.3:1 suggestion of 0.1mgS/L targets. These factors were validated by feature-importance scores obtained with the XGBoost as most predictive. On the contrary, initial sulfide concentration had a less (nonlinear) pronounced impact: at very high initial S, the removal efficiency reduced slightly because of the incomplete oxidation under fixed time. pH and temperature had minor correlations, presumably because standard wastewater pH (6.5-8.5) creates limited variations in the chemistry of iron in that range.

The high ML performance implies some practical implications. Design can be done in silico with the help of a predictive model: e.g., estimates of required aeration time and Fe dose to reach a desired level of H_2S might be made at a certain level under a certain influent condition. This is more lenient than simple ratio rules, whereby the model takes into account interactions and kinetic bounds. Notably, operators will be able to optimise the use of chemicals and energy (aeration) to reduce the minimum costs and environmental impact by correctly predicting the outcome. An example is when XGBoost forecasts that doubling the dose of Fe will not result in any significant addition to the removal (plateau behavior), the operator may select a smaller dose.

Our results are consistent with the literature: ensemble models are especially well-suited to nonlinear wastewater data, and the advantage of XGBoost has been previously reported in a number of treatment prediction tasks. In our work, XGBoost was able to learn nonlinear sulfide-iron and sulfide-oxygen interactions. Here, the ANN, though flexible, cannot have performed excellently because there were only limited data and that it was hard to tune deep layers with very few hundred samples.

4. Conclusions:

This paper shows that the current ML methods are useful in predicting the removal of sulfide in wastewater through aeration and iron addition. Using literature experimental data, we demonstrated that XGBoost specifically is more accurate (maximum R^2, minimum errors) than RF, AdaBoost, SVR and ANN. Tables 1-2 that have the performance of the models and the parameters of each model tuned give tangible benchmarks to future work. Notably, the ML analysis validates established chemical intuitions (e.g. Fe:S stoichiometry) and measures the effect of operations parameters on removal.

The implications of these findings are practical and predictive models, such as the XGBoost can be incorporated into the control system of the treatment plant or used in its design, because the models do not require a lot of pilot testing to work. They are a data-based complement to the conventional mechanistic models. Future research ought to increase the dataset (some different types of wastewater, larger datasets), investigate time-varying models (e.g., recurrent networks as dynamic behavior), and use explainable AI (e.g., SHAP values) to further understand feature interactions.The impressive predictive accuracy of the machine learning models demonstrates their potential for practical applications, including optimization, proactive management, and real-time monitoring of the sulfide separation process. By integrating machine learning into the separation system, adjustments can be made to operational parameters to ensure compliance with environmental standards while reducing resource consumption. The predictive models created in this study can be integrated into the control systems of wastewater treatment. Real-time monitoring of sulfide concentrations through sensors provides data to the trained model, which forecasts the removal efficiency under various operational conditions. This approach to predictive analytics empowers operators to make timely adjustments in aeration rates, oxygen levels, thereby enhancing treatment performance while minimizing energy and reagent usage.

Abbreviations and Nomenclature

ML: Machine Learning

SVR: Support Vector Regression

XGBoost: Extreme Gradient Boosting
LSTM: Long Short-Term Memory
R^2: Coefficient of Determination
RMSE: Root Mean Squared Error
MSE: Mean Squared Error

References

[1] A. Elkhazraji *et al.*, "Oxidation of hydrogen sulfide and CO2 mixtures: Laser-based multi-speciation and kinetic modeling," *Chemical Engineering Journal,* vol. 486, p. 150421, 2024.

[2] N. Ahmad, F. Ahmad, I. Khan, and A. D. Khan, "Studies on the oxidative removal of sodium thiosulfate from aqueous solution," *Arabian Journal for Science and Engineering,* vol. 40, no. 2, pp. 289-293, 2015.

[3] E. Aghdam, S. R. Mohandes, P. Manu, C. Cheung, A. Yunusa-Kaltungo, and T. Zayed, "Predicting quality parameters of wastewater treatment plants using artificial intelligence techniques," *Journal of Cleaner Production,* vol. 405, p. 137019, 2023.

[4] B. Adibimanesh, S. Polesek-Karczewska, F. Bagherzadeh, P. Szczuko, and T. Shafighfard, "Energy consumption optimization in wastewater treatment plants: Machine learning for monitoring incineration of sewage sludge," *Sustainable energy technologies and assessments,* vol. 56, p. 103040, 2023.

[5] M. Ksibi, "Chemical oxidation with hydrogen peroxide for domestic wastewater treatment," *Chemical Engineering Journal,* vol. 119, no. 2-3, pp. 161-165, 2006.

[6] F. Ahmad, S. Maitra, and N. Ahmad, "Treatment of sulfidic wastewater using iron salts," *Arabian Journal for Science and Engineering,* vol. 42, no. 4, pp. 1455-1462, 2017.

[7] N. Ahmad, S. Maitra, B. K. Dutta, and F. Ahmad, "Remediation of sulfidic wastewater by catalytic oxidation with hydrogen peroxide," *Journal of Environmental Sciences,* vol. 21, no. 12, pp. 1735-1740, 2009.

[8] F. Ahmad, A. Z. Al-Khazaal, N. Ahmad, and I. Alenezi, "Remediation of sulfidic wastewater by aeration in the presence of ultrasonic vibration," *Engineering, Technology & Applied Science Research,* vol. 8, no. 3, pp. 2919-2922, 2018.

[9] F. Van der Zee, S. Villaverde, P. Garcia, and F. F.-Polanco, "Sulfide removal by moderate oxygenation of anaerobic sludge environments," *Bioresource technology,* vol. 98, no. 3, pp. 518-524, 2007.

[10] K. El Bouchefry and R. S. de Souza, "Learning in big data: Introduction to machine learning," in *Knowledge discovery in big data from astronomy and earth observation*: Elsevier, 2020, pp. 225-249.

[11] M. Javaid, A. Haleem, I. H. Khan, and R. Suman, "Understanding the potential applications of Artificial Intelligence in Agriculture Sector," *Advanced agrochem,* vol. 2, no. 1, pp. 15-30, 2023.

[12] M. Pirdashti, S. Curteanu, M. H. Kamangar, M. H. Hassim, and M. A. Khatami, "Artificial neural networks: applications in chemical engineering," *Reviews in Chemical Engineering,* vol. 29, no. 4, pp. 205-239, 2013.

[13] H. Samuel, E. E. Etim, U. Nweke-Maraizu, and S. Yakubu, "Machine Learning in Chemical Kinetics: Predictions, Mechanistic Analysis, and Reaction Optimization," *Applied Journal of Environmental Engineering Science,* vol. 10, no. 1, pp. & al./Appl. J. Envir. Eng. Sci. 10 N° 1 (2024) 36-61, 2024.

[14] D. A. Narciso and F. G. Martins, "Application of machine learning tools for energy efficiency in industry: A review," *Energy Reports,* vol. 6, pp. 1181-1199, 2020.

[15] B. Xiong, K. Chen, C. Ke, S. Zhao, Z. Dang, and C. Guo, "Prediction of heavy metal removal performance of sulfate-reducing bacteria using machine learning," *Bioresource Technology,* vol. 397, p. 130501, 2024.

[16] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning: with applications in R*. Springer, 2013.

[17] M. M. Askar, M. A. E. Ela, A. H. El-Banbi, and M. H. Sayyouh, "Development of machine learning models for predicting the deposition of sulfide scales in oil production wells," *Journal of King Saud University–Engineering Sciences,* vol. 37, no. 7, pp. 1-34, 2025.

[18] M. Bahramian, R. K. Dereli, W. Zhao, M. Giberti, and E. Casey, "Data to intelligence: The role of data-driven models in wastewater treatment," *Expert systems with applications,* vol. 217, p. 119453, 2023.

[19] H. A. Saddiqi *et al.*, "Predictive modeling of emulsion stability and drop characteristics using machine learning: A study on surfactant influence and time dynamics," *Food and Bioproducts Processing,* vol. 149, pp. 190-198, 2025.

[20] M. Akoulih *et al.*, "Electrocoagulation-based AZO DYE (P4R) removal rate prediction model using deep learning," *Procedia Computer Science,* vol. 236, pp. 51-58, 2024.

[21] H. Mahanna, N. El-Rashidy, M. R. Kaloop, S. El-Sapakh, A. Alluqmani, and R. Hassan, "Prediction of wastewater treatment plant performance through machine learning techniques," *Desalination and Water Treatment,* vol. 319, p. 100524, 2024.

[22] H. Tao *et al.*, "Coupled extreme gradient boosting algorithm with artificial intelligence models for predicting compressive strength of fiber reinforced polymer-confined concrete," *Engineering Applications of Artificial Intelligence,* vol. 134, p. 108674, 2024.

[23] M. Ahmadi and Z. Chen, "Machine learning-based models for predicting permeability impairment due to scale deposition," *Journal of Petroleum Exploration and production technology,* vol. 10, no. 7, pp. 2873-2884, 2020.