Machine-Learning Modeling of Sulfuric-Acid Leaching of Metalliferous Oil-Shale Ash: Predictive Performance, Feature Importance, and Confounding in an Unbalanced Experimental Design

Shamshiddin K. Kurbanov

Military Institute of Information and Communication Technologies and Communications, University of Military Security and Defense of the Republic of Uzbekistan, Uzbekistan
Corresponding author: [e-mail to be provided]

ABSTRACT

A dataset of 28 sulfuric-acid leaching experiments on metalliferous oil-shale ash was analyzed using multiple linear regression, second-order polynomial regression, and random forest models to predict the recovery of uranium, thorium, vanadium, molybdenum, scandium, and rhenium. The input variables were liquid-to-solid ratio, temperature, leaching time, and H_2SO_4 concentration. Because of the small sample size, model performance was evaluated by leave-one-out cross-validation (LOO-CV). Random forest provided the highest predictive performance for all six elements, with LOO-CV R^2 values of 0.823-0.906, whereas multiple linear regression gave 0.515-0.784. Polynomial regression performed similarly to random forest for most elements, confirming a substantial nonlinear component in the response surface. Random-forest feature importance attributed approximately 89-98% of total model importance to sulfuric-acid concentration. In the full linear model, the H_2SO_4 coefficient was significant for all elements ($p < 0.001$). In contrast, the temperature coefficient was negative in the pooled regression but statistically non-significant for five of the six elements. A controlled comparison at 50 g L^{-1} H_2SO_4 reversed the apparent temperature effect; for uranium, mean recovery increased from 48.0% at 30 °C to 51.2% at 75 °C. The sign reversal is consistent with confounding caused by an unbalanced experimental design rather than a physically adverse temperature effect. The study shows that nonlinear machine-learning models can improve prediction for small hydrometallurgical datasets, but interpretable process conclusions require explicit diagnosis of experimental-design structure, confounding, and uncertainty.

Keywords: oil-shale ash; sulfuric acid leaching; machine learning; random forest; confounding; design of experiments

Highlights

- Random forest achieved LOO-CV R^2 values of 0.823-0.906 for six metals.
- Nonlinear models strongly outperformed multiple linear regression for U, V, Mo and Sc.
- Sulfuric-acid concentration accounted for approximately 89-98% of RF importance.
- An apparent negative temperature coefficient was traced to experimental-design confounding.
- Design-balance diagnostics are essential before interpreting process-regression coefficients.

1. Introduction

Machine learning is increasingly used to describe and predict hydrometallurgical leaching processes because metal dissolution is governed by interacting chemical, mineralogical, and operational variables that are often nonlinear. Reviews of mineral-leaching applications have highlighted artificial neural networks, support-vector machines, tree ensembles, and hybrid modeling as useful complements to classical empirical models [1]. Recent leaching studies also show that ensemble methods can achieve higher predictive accuracy than linear approaches when relationships between process variables and recovery are strongly nonlinear [2-5].

Model accuracy alone, however, is insufficient for process interpretation. Hydrometallurgical datasets are frequently small and generated by experimental campaigns in which factor combinations are not perfectly balanced. Under such conditions, regression coefficients may describe correlations induced by the design rather than causal physicochemical effects. This distinction is particularly important when model outputs are used to formulate operating recommendations. Comparative studies in copper leaching have emphasized that the best-performing algorithm depends on the specific dataset and that model generalization must be verified rather than assumed [6]. Design-of-experiments approaches applied to high-pressure laterite leaching similarly demonstrate the value of structured factor variation for separating process effects [7].

The present study uses 28 agitation-leaching experiments on metalliferous oil-shale ash to address two complementary questions: (i) how well linear, polynomial, and random-forest models predict the recovery of six target elements; and (ii) whether the fitted effects can be interpreted physically in the presence of an unbalanced experimental design. Particular attention is given to the apparent sign of the temperature effect and its dependence on sulfuric-acid concentration. The objective is therefore not only to identify a high-performing predictive model, but also to demonstrate a practical workflow for diagnosing confounding before model coefficients are translated into hydrometallurgical conclusions.

2. Materials and methods

2.1. Experimental dataset and variables

The dataset comprised 28 agitation-leaching experiments performed on 50 g samples of oil-shale ash with particle size below 1 mm. Four operating variables were used as model inputs: liquid-to-solid ratio (L/S; two levels, 3 and 5 mL g^{-1}), temperature (30 and 75 °C), leaching time (1 and 2 h), and sulfuric-acid concentration (10, 20, 50, and 100 g L^{-1} H_2SO_4). The response variables were the recoveries (%) of U, Th, V, Mo, Sc, and Re.

The available dataset is not a fully balanced four-factor factorial design. Consequently, factor frequencies and combinations differ across the experimental matrix. This structure was retained because it reflects the original experimental campaign; however, it was explicitly considered during model interpretation and confounding diagnostics.

2.2. Predictive models and validation

Three regression approaches were evaluated separately for each element: multiple linear regression, second-order polynomial regression, and random forest. The random forest used 300 trees, maximum tree depth of 4, and a minimum of two observations per terminal leaf to reduce overfitting risk in the small dataset. Model performance was evaluated using leave-one-out cross-validation (LOO-CV), which uses 27 observations for training and one for validation in each iteration and therefore avoids an arbitrary single train/test split.

For the multiple linear regression model, coefficient estimates, standard errors, t statistics, p values, and 95% confidence intervals were evaluated. Variance inflation factors (VIFs) were used to diagnose multicollinearity. Random-forest feature importance was used as a predictive importance measure; it is interpreted here as a measure of model dependence on each variable and not as direct evidence of causal process importance.

2.3. Confounding diagnostic

The temperature effect was examined in two ways. First, the coefficient of temperature was inspected in the pooled multivariable linear model. Second, a controlled comparison was performed within the subset of experiments conducted at 50 g L^{-1} H_2SO_4. If pooled and concentration-controlled comparisons give opposite temperature trends, the discrepancy is interpreted as evidence that the pooled association is confounded by the distribution of acid concentration across the experimental design. This pattern is structurally analogous to a regression form of Simpson-type sign reversal.

3. Results and discussion

3.1. Predictive performance of the three model classes

Random forest provided the highest LOO-CV R^2 for every target element: 0.871 for U, 0.823 for Th, 0.860 for V, 0.886 for Mo, 0.879 for Sc, and 0.906 for Re. Multiple linear regression was substantially less accurate for U, V, Mo, and Sc, with R^2 values near 0.52-0.54, whereas polynomial regression increased R^2 to 0.858-0.877 for these elements. The smaller improvement for Th reflects the relatively strong performance of the linear model for that response. Overall, the comparison indicates that nonlinear response structure is important for most elements in the investigated operating domain.

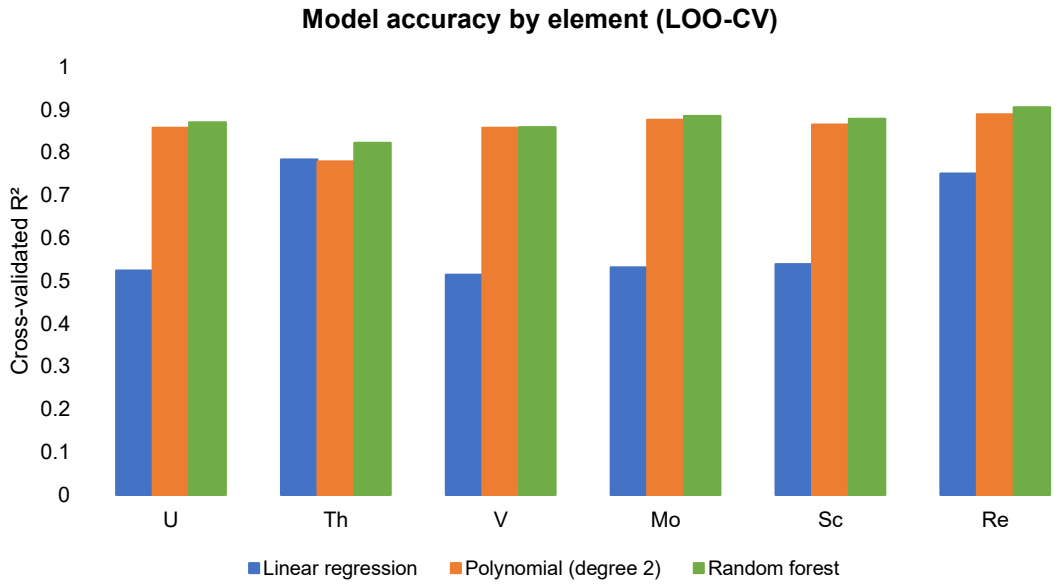

Figure 1. Leave-one-out cross-validated R^2 for multiple linear regression, second-order polynomial regression, and random forest. The chart and its embedded data are editable in Word.

Table 1. Leave-one-out cross-validated predictive performance (R^2) of the compared models.

Element	Multiple linear regression	Quadratic polynomial regression	Random forest
U	0.525	0.858	0.871
Th	0.784	0.780	0.823
V	0.515	0.858	0.860
Mo	0.532	0.877	0.886
Sc	0.540	0.866	0.879
Re	0.751	0.890	0.906

The predicted-versus-observed diagrams for U, V, and Re further illustrate the model behavior. Most high-recovery observations cluster near the 1:1 line, whereas deviations are larger in the low-to-intermediate recovery range. This is consistent with the limited number of experiments and the strong concentration dependence of the response. For journal-quality transparency, the editable scatter plots reproduce coordinates digitized from the source figure because the manuscript did not contain the complete 28-row prediction table; the reported R^2 values are taken directly from the source analysis.

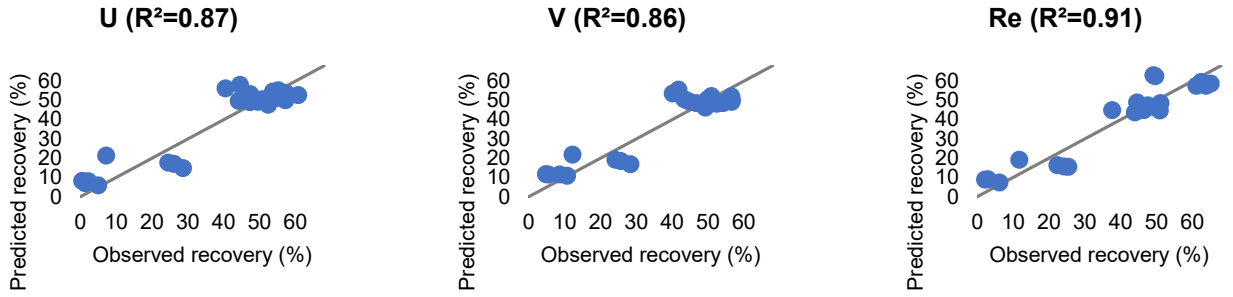

Figure 2. Observed versus LOO-CV random-forest predictions for (a) U, (b) V, and (c) Re. The 1:1 line represents perfect prediction. Point coordinates were reconstructed from the source figure solely to make the plots editable; source-reported cross-validated R^2 values are preserved.

3.2. Random-forest feature importance

Sulfuric-acid concentration dominated random-forest feature importance for all six elements. Its contribution ranged from approximately 0.89 for Th to 0.982 for Re, leaving no more than about 11% of the total importance for the combined effects of L/S ratio, temperature, and leaching time. The result indicates that, within the sampled operating domain, changes in acid concentration are the principal source of predictive variation. This conclusion is also supported by the parametric regression analysis, in which the acid-concentration coefficient is statistically significant for all six elements.

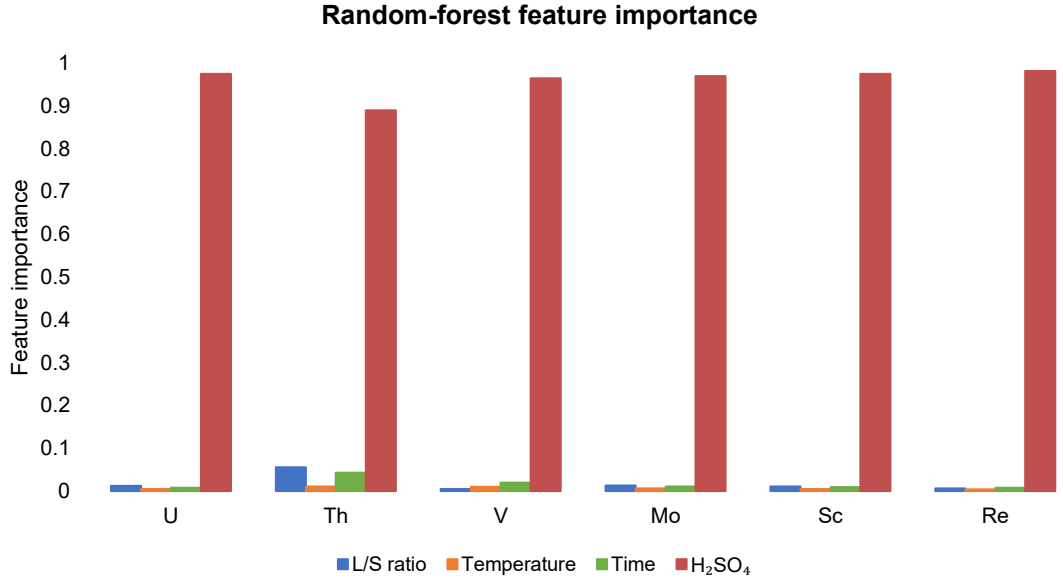

Figure 3. Random-forest feature importance for the six target elements. Minor feature-importance values were reconstructed from the source chart; the reported 89-98% dominance of H_2SO_4 concentration is preserved. All chart data are editable in Word.

Table 2. Sulfuric-acid concentration coefficient in the full multiple linear regression model.

Element	β(H_2SO_4)	SE	t	p
U	0.417	0.082	5.06	<0.001
Th	0.333	0.036	9.38	<0.001
V	0.377	0.071	5.33	<0.001

Element	β(H_2SO_4)	SE	t	p
Mo	0.381	0.073	5.19	<0.001
Sc	0.372	0.070	5.29	<0.001
Re	0.508	0.058	8.70	<0.001

3.3. Linear-model coefficients and uncertainty

The full multiple linear regression provides a complementary view of factor effects. The H_2SO_4 coefficient is positive and highly significant for every element ($p < 0.001$). By contrast, most coefficients for L/S ratio, temperature, and time have wide confidence intervals that overlap zero. The negative temperature coefficient in the pooled model is therefore not sufficient evidence that higher temperature suppresses dissolution. For five of the six elements, the source analysis reports p values between 0.053 and 0.137 for temperature, i.e., no conventional 5% statistical significance.

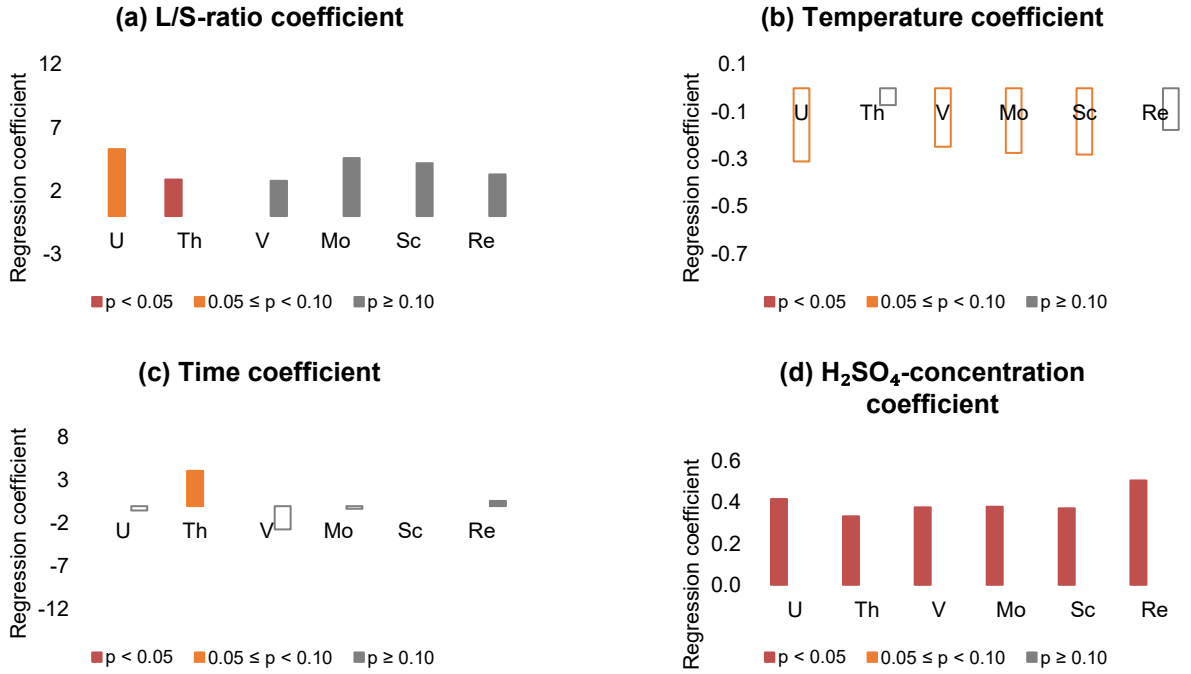

Figure 4. Multiple-linear-regression coefficient estimates by factor: (a) L/S ratio, (b) temperature, (c) time, and (d) H_2SO_4 concentration. The editable charts show point estimates and significance classes. Approximate 95% confidence intervals reconstructed from the source coefficient plot are listed in Table 3.

Table 3. Regression coefficient estimates and approximate 95% confidence intervals reconstructed from the source coefficient plot.

Element	L/S ratio β [95% CI]	Temperature β [95% CI]	Time β [95% CI]	H_2SO_4 β [95% CI]
U	5.3 [-1.0; 11.7]	-0.309 [-0.64; 0.03]	-0.5 [-10.5; 9.7]	0.417 [0.256; 0.578]
Th	2.9 [0.1; 5.6]	-0.071 [-0.22; 0.07]	4.1 [-0.4; 8.5]	0.333 [0.262; 0.404]
V	2.8 [-2.6; 8.4]	-0.247 [-0.53; 0.04]	-2.7 [-11.4; 6.0]	0.377 [0.238; 0.516]
Mo	4.6 [-1.1; 10.3]	-0.273 [-0.57; 0.02]	-0.3 [-9.4; 8.7]	0.381 [0.238; 0.524]
Sc	4.2 [-1.3; 9.6]	-0.280 [-0.57; 0.01]	0.0 [-9.9; 8.7]	0.372 [0.235; 0.509]

Element	L/S ratio β [95% CI]	Temperature β [95% CI]	Time β [95% CI]	H_2SO_4 β [95% CI]
Re	3.3 [-1.2; 7.9]	-0.176 [-0.41; 0.06]	0.6 [-6.6; 7.8]	0.508 [0.394; 0.622]

3.4. Confounding of the temperature effect

When all experiments are pooled, mean recovery appears lower at 75 °C than at 30 °C for U, V, and Re. This aggregate comparison is confounded because the distribution of sulfuric-acid concentrations differs between temperature groups. Restricting the comparison to experiments conducted at 50 g L^{-1} H_2SO_4 reverses the temperature trend. Uranium recovery increases from 48.0% at 30 °C to 51.2% at 75 °C, consistent with the expected acceleration of heterogeneous dissolution at higher temperature. The corresponding controlled comparisons for V and Re show the same positive direction in the source figure.

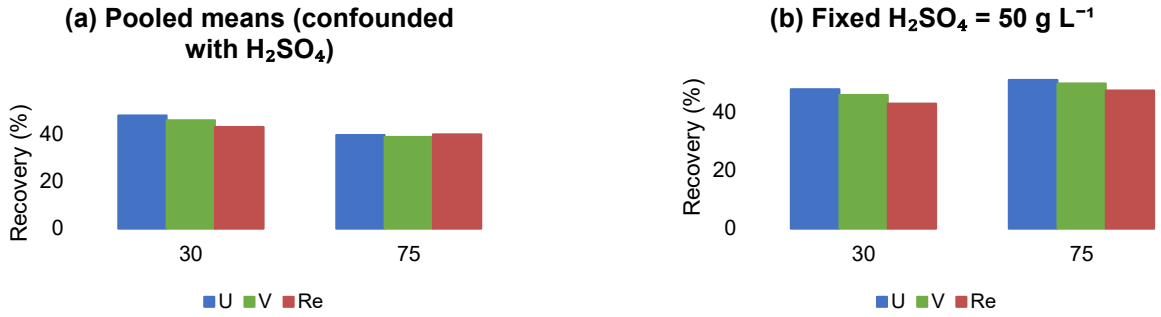

Figure 5. Illustration of confounding in the temperature effect: (a) pooled means across the full dataset and (b) means at fixed H_2SO_4 concentration of 50 g L^{-1}. Values not explicitly tabulated in the source manuscript were reconstructed from the original chart to preserve editability; the uranium values 48.0% and 51.2% are source-reported.

The VIF for temperature was approximately 1.07, which argues against conventional multicollinearity as the explanation for the sign reversal. The problem is instead the structure of factor combinations in the experimental matrix: a regression coefficient estimated from an unbalanced design may represent a conditional association that is sensitive to the distribution of other factors. This result demonstrates why process-physics interpretation should not rely solely on coefficient sign, even when a multivariable model is statistically well specified.

3.5. Implications, limitations, and recommended next steps

The study has two practical implications. First, random forest and polynomial regression can provide useful predictive performance even for a small leaching dataset when validation is performed observation by observation. Second, interpretability requires stronger safeguards than prediction: feature importance is not causal importance, and regression coefficients can be distorted by unbalanced factor combinations. For the next experimental stage, a full or fractional factorial design should be used so that temperature, acid concentration, time, and L/S ratio are varied independently enough to estimate interactions and main effects with controlled uncertainty.

The principal limitations are the small sample size (n = 28), absence of an independent external validation dataset, and incomplete reporting of the row-level cross-validated predictions in the source manuscript. These constraints preclude strong claims about model generalization outside the investigated range. Future work should expand the experimental matrix, preserve replicate measurements, report prediction intervals, compare additional regularized/boosted models, and quantify random-forest prediction uncertainty using resampling or jackknife-type approaches [8].

4. Conclusions

1. Random forest achieved the best LOO-CV predictive performance for all six metals, with R^2 = 0.823-0.906, while multiple linear regression yielded R^2 = 0.515-0.784.
2. The large gain of polynomial and random-forest models for U, V, Mo, and Sc demonstrates

substantial nonlinear dependence of recovery on the investigated operating variables.

3. Sulfuric-acid concentration is the dominant predictive variable in the sampled dataset, accounting for approximately 89-98% of random-forest importance; its linear-model coefficient is significant for every element at $p < 0.001$.

4. The apparent negative temperature coefficient in the pooled regression is not a reliable physicochemical effect. At fixed H_2SO_4 concentration, the temperature trend becomes positive, indicating confounding caused by the unbalanced design.

5. Subsequent work should use balanced factorial experimentation and uncertainty-aware validation before predictive models are converted into engineering operating rules.

Declarations

Funding: [Please provide funding information, or confirm that no specific funding was received.]

Conflict of interest: The author declares no competing financial or personal interests.

Data availability: The underlying experimental dataset may be made available by the author upon reasonable request. Before submission, the repository or access conditions should be specified if required by the target journal.

Author contributions: S.K. Kurbanov - conceptualization, investigation, methodology, statistical analysis, data interpretation, visualization, and manuscript preparation. [Please confirm or revise the CRediT roles before submission.]

References

[1] Saldaña, M., Neira, P., Gallegos, S., Salinas-Rodríguez, E., Pérez-Rey, I., Toro, N., 2022. Mineral Leaching Modeling Through Machine Learning Algorithms - A Review. Frontiers in Earth Science 10, 816751. https://doi.org/10.3389/feart.2022.816751.

[2] Zhang, Z., Zhang, X., Zhang, D., Zhang, X., Qiu, F., Li, W., Liu, Z., Shu, J., Tang, C., 2022. Application of Machine Learning in a Mineral Leaching Process - Taking Pyrolusite Leaching as an Example. ACS Omega 7, 48130-48138. https://doi.org/10.1021/acsomega.2c06129.

[3] Weng, W., Lan, Y., Tang, D., Chen, H., Xia, J., Chen, J., Zhong, S., 2025. Acid leaching of copper from copper slag based on machine learning analysis. Separation and Purification Technology 379, 135025. https://doi.org/10.1016/j.seppur.2025.135025.

[4] Feng, Y., Jiang, M., Yang, L., Ren, W., Wang, H., Yang, G., Yang, C., Zhou, L., Shi, H., Shao, P., Luo, X., 2025. Machine learning for efficient metal leaching from spent LiFePO4: Predictive modeling and sustainability assessment. Separation and Purification Technology 370, 133334. https://doi.org/10.1016/j.seppur.2025.133334.

[5] Kuzu, E., Top, S., Kursunoglu, S., Altiner, M., 2026. Predictive Modeling and SHAP-Based Interpretability of Manganese and Iron Dissolution in Multi-Acid Leaching Systems Using Hybrid Machine Learning. Processes 14, 1716. https://doi.org/10.3390/pr14111716.

[6] Flores, V., Leiva, C., 2021. A Comparative Study on Supervised Machine Learning Algorithms for Copper Recovery Quality Prediction in a Leaching Process. Sensors 21, 2119. https://doi.org/10.3390/s21062119.

[7] Milivojević, M., Stopić, S., Friedrich, B., Stojanović, B., Drndarević, D., 2012. Computer modeling of high-pressure leaching process of nickel laterite by design of experiments and neural networks. International Journal of Minerals, Metallurgy, and Materials 19, 584-594. https://doi.org/10.1007/s12613-012-0599-x.

[8] Wager, S., Hastie, T., Efron, B., 2014. Confidence intervals for random forests: the jackknife and the infinitesimal jackknife. Journal of Machine Learning Research 15, 1625-1651.