Journal of Hazardous Materials 496 (2025) 139479

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

Explainable artificial intelligence for predicting rare earth elements leaching from secondary resources

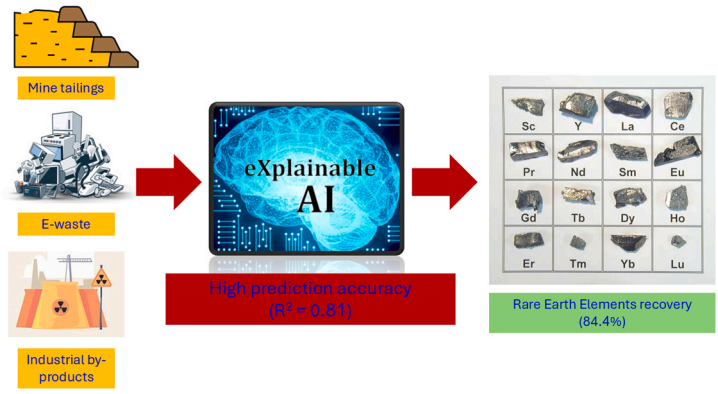
Quang Loc Nguyen [a], Huy Nguyen Lai [b], Hong T.M. Nguyen [a], Le Long Pham [a], Andrei Herdean [a], Nature Poddar [a], Unnikrishnan Kuzhiumparambil [a], Ai Nguyen [c], Mathieu Pernice [a], Peter J. Ralph [a], Phong H.N. Vo [a,*]

[a] Climate Change Cluster, University of Technology Sydney, Ultimo, NSW 2007, Australia
[b] Environmental Engineering and Management, Asian Institute of Technology (AIT), Klong Luang, Pathumthani, Thailand
[c] School of the Environment, The University of Queensland, Brisbane, Australia

HIGHLIGHTS

- The AI model was trained using 572 experimental datasets.
- The AI model show high level of prediciton accuracy (R^2=0.81).
- Silica concentration is the most critical factor in REEs leaching.
- REEs classification and pH possess lower impact on REEs recovery efficiency.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Explainable artificial intelligence
Rare earth elements
Hydrometallurgy
Bioleaching
Mine tailing
Electronic waste

ABSTRACT

The increasing global demand for rare earth elements (REEs) presents significant mining and extraction challenges. Extraction of REEs from secondary resources such as mine tailings, electronic waste, and industrial by-products becomes an alternative solution. This study introduces an explainable artificial intelligence (AI) system designed to (i) predict the leaching efficiency rates and (ii) provide real-time explanations of the key extraction factors during the leaching process, while also recommending condition adjustments to optimize the leaching efficiency of REEs from secondary resources. Trained on 572 experimental datasets sourced from the Web of Science database, the system enhances leaching performance by offering explainable recovery rates and identifying influential process parameters. The results showed that silica concentration is the most critical factor, followed by REEs classification (light vs. heavy). In contrast, acid strength (pH), aluminum content, and temperature exhibited moderate but comparatively lower contributions to the overall leaching performance. By identifying suboptimal parameters and suggesting adjustments that lead to improvements in the predicted recovery (R^2=0.81), the approach exemplifies how explainable AI can bridge the gap between empirical data and

* Corresponding author.
E-mail address: phong.vo@uts.edu.au (P.H.N. Vo).

https://doi.org/10.1016/j.jhazmat.2025.139479
Received 3 May 2025; Received in revised form 6 July 2025; Accepted 6 August 2025
Available online 6 August 2025

process innovation. This methodology offers a broadly applicable framework for enhancing decision-making and process efficiency in the complex extraction systems, with the potential to extend beyond REEs into other resource- and energy-intensive industries.

1. Introduction

Rare earth elements (REEs) are 17 elements with similar chemical and physical properties. REEs are essential in advanced manufacturing due to their unique magnetic, luminescent, and highly reactive properties. REEs are critical in various advanced applications such as electronics, renewable energy technologies, and defence systems [10]. The demand for REEs has grown significantly in the last couple of years, expecting up to $USD 10.9 billion by 2029 [20]. This indicates the necessity for a sustainable REEs supply to fulfill the market demand.

Hydrometallurgy techniques, including biohydrometallurgy, have been studied widely for REEs extraction from ores and tailings [13,34]. Hydrometallurgy extraction offers several benefits, such as significant efficiency and selective recovery from low-grade resources. Although hydrometallurgy provides substantial benefits, optimising leaching conditions such as pH, pulp ratio, stirring speed, and temperature vary on a case-by-case basis. An online platform to predict the REEs leaching efficiency and advise the optimum leaching conditions is critically needed to save resources and reduce labor.

Machine learning, a subfield of artificial intelligence (AI), performs statistical analysis of a large dataset and unveils the correlation between the leaching input, such as mineral types, operating conditions, and the output, specifically to the leaching efficiency [21]. However, the application of machine learning prediction in extended environments such as different acid tanks, varying outdoor temperature conditions, and specialized types of equipment is scarce. Several current applications favor simpler, more interpretable models (e.g., Bayesian regression models) at the expense of accuracy [19]. This is due to the lack of interpretability in the complex practical conditions of the current machine learning approach. To understand the main driving factors for the optimization of the leaching process, an ensemble model of machine learning methods is employed to predict the leaching efficiency and describe the particular conditions contributing to that risk.

Gradient Boosting Machines (GBM) leverage the principles of boosting and gradient descent in function space to iteratively refine predictive performance [4,8,9]. Boosting is an ensemble learning technique that sequentially trains weak learners, typically regression trees, where each new model corrects the residual errors of its predecessors by assigning higher weights to mispredicted instances. Unlike traditional gradient descent, which optimizes model parameters, GBM employs gradient descent in function space, adjusting the predictions of the model by iteratively minimizing the gradient of a specified loss function. This approach enables GBM to effectively reduce both bias and variance, leading to enhanced predictive accuracy across a wide range of applications. In this study, we employed Extreme Gradient Boosting, which is an advanced subclass of Gradient Boosting Machines [31].

Significant data regarding the hydrometallurgical extraction of REEs from ores and mine tailings has been published in the literature [1,24]. However, the data remains heterogeneous and under-exploited by data scientists to support the hydrometallurgical extraction of REEs. Given the tremendous potential of the available database, the objective of this study is to develop an explainable AI module, based on the GBM, to predict the leaching efficiency of REEs from the secondary materials such as mine tailings, industrial by-products, and electronic waste in complex conditions. Unlike traditional approaches that enhance interpretability by simplifying models, this method maintains high transparency while leveraging the predictive power and decision-support of advanced, non-parametric ensemble techniques.

2. Materials and methods

2.1. Data collection, preprocessing, and model development

The dataset in this study was extracted from articles indexed in the Web of Science database (appendix). The dataset integrated both chemical composition and experimental parameters. The chemical composition includes the type of REEs, materials, leaching solution; and the experimental parameters are the recovery rate (%), leaching time (min), hydrogen ion concentration (pH), solid-liquid ratio (g/mL), stirring speed (rpm), pretreatment, silica concentration (Si (%)), aluminum (Al (%)), iron (Fe (%)), and temperature (°C). The input variables of the module include categorical and numerical variables. The output variable is the recovery rate (%), which measures the leaching efficiency of experiments.

The categorical variables include Materials (material types for leaching), Solution (physical chemical properties of the leachate), and REEs (type of REEs). Categorical variables were preprocessed via one-hot encoding, given that each unique category was transformed into a distinct binary feature. The Materials variable consisted of seven sources (coal gangue, used magnet, iron residue, deep sea mud, zircon tailings, coal refuse clay, coal ash). The Solution variable comprises six leaching agents (HCl, HNO_3, H_2SO_4, H_3PO_4, Methanesulphonic acid (MSA), p-toluenesulphonic acid (PTSA)). The REEs variables were classified as Light REEs (LREEs), including La, Ce, Pr, Nd, Pm, Sm, and Eu, Heavy REEs (HREEs), including Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, and Mixed REEs including both LREEs and HREEs (Y and Sc excluded from the dataset). To clearly distinguish the temperature for leaching (<180 °C) and for thermal pre-treatment (180 °C – 900 °C), we introduced an additional binary categorical variable named 'Pretreatment', which encodes whether the temperature recorded in the dataset corresponds to a thermal pre-treatment (code 1) or to the actual leaching process (code 0).

Numerical variables consist of Si (%), Al (%), Fe (%), temperature (°C), pH, solid-liquid ratio (g/mL), stirring speed (rpm), and leaching time (min). After encoding, each categorical level was represented as a distinct feature, resulting in a total feature space of 26 dimensions for model development.

The Gradient Boosting Machines (GBM) is selected as a non-parametric method in this study to accurately capture a range of complex, non-linear interactions of all the features extracted from the database. Given a training dataset $\mathscr{D} = \{x_i y_i\}_1^N$, the goal of gradient boosting is to approximate the target function $F(x)$ which maps input instances x to their corresponding output values y by minimizing the expected value of a specified loss function $L(y, F(x))$. Gradient boosting constructs an additive approximation of $F^*(x)$ by iteratively building a weighted sum of simpler functions. The model is expressed mathematically as below [4]:

$$F_m(x) = F_{m-1}(x) + \rho_m h_m(x) \tag{1}$$

The approximation of $F^*(x)$ is constructed iteratively. Initially, a constant approximation of $F^*(x)$ is obtained as:

$$F_0(x) = arg\ min \sum_{i=1}^{N} L(y_i, \alpha) \tag{2}$$

Each successive model is expected to minimize the remaining errors from the previous models, progressively refining the overall prediction accuracy:

$$(p_m, h_m(x)) = \underset{p,h}{\operatorname{argmin}} \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + ph(x_i)) \quad (3)$$

Instead of solving the optimization problem directly, each h_m represents a greedy step in a gradient descent process for optimizing F^*. In this approach, each model h_m is trained on a modified dataset $\mathscr{D} = \{x_i r_{mi}\}_{m=1}^{N}$ where pseudo-residuals r_{mi} are computed by:

$$r_{mi} = \left[\frac{\partial L(y_i, F(x))}{\partial F(x)}\right]_{F(x)=F_{m-1}(x)} \quad (4)$$

Whereas:

p_m: the weight of the m^{th} function,
$h_m(x)$: individual models within the ensemble (e.g., decision trees).
$F_m(x)$: step m of function F^*
r_{mi}: pseudo-residuals

The value of p_m is then determined by solving a line search optimization problem. Approximately 80 % of the dataset was used for performance evaluation (training). The model was validated via three approaches: (i) internal validation using the other 20 % of the whole dataset, (ii) k-fold cross-validation to assess model robustness and minimize over-fitting, (iii) external validation using an independent dataset to evaluate the generalizability of the model. To prevent any potential bias toward the test set, the tested data remained compressed and inaccessible until the completion of method development.

To compare the performance of the GBM, two other models were developed, including the Support Vector Machine (SVM) autoregressive model [30], and the Bayesian Gaussian Process Regression (BGPR) model [5]. Both models were previously used to predict chemical leaching and chemical reaction, and synthesis [11,33].

2.2. Feature explanations

The GBM produces results that are difficult to interpret in depth. To address the issue, the model-agnostic approach was adopted to assess the impact of the features in the prediction via the Shapley values (SHAP) (Eq. 5) [18]. The Shapley values are unique in their ability to assign the feature importance while upholding two key properties: local accuracy and consistency. The estimation methods developed using the GBM and SHAP can be computed in real time and provide interpretable insights into each prediction.

For this study, f represents a GBM, and x is the completed set of input features at a given time point. The SHAP, $\phi_i(f, x)$, assigned to each feature in the input vector x (e.g., temperature, solution, leaching time), represent the allocation of credit for each feature's contribution to the prediction $F(x)$. Here, $\phi_i(f, x)$ is a numerical value indicating the specific impact of a feature i on the model's prediction given the input x.

In these properties, $f_x(S) = E[f(x)|x_S]$ represents the expected model output with only features from subset S present. For an individual prediction, $F(x)$, SHAP can be computed as below to provide a comprehensive measure of each feature's contribution:

$$\boldsymbol{\phi_i}(\boldsymbol{f}, \boldsymbol{x}) = \sum_{\boldsymbol{S} \subseteq \boldsymbol{S_{all}}/\{i\}} \frac{|\boldsymbol{S}|!(\boldsymbol{M} - |\boldsymbol{S}| - \boldsymbol{1})!}{\boldsymbol{M}!} [\boldsymbol{f_x}(\boldsymbol{S} \cup \{\boldsymbol{i}\}) - \boldsymbol{f_x}(\boldsymbol{S})] \quad (5)$$

$$= \sum_{\boldsymbol{S} \subseteq \boldsymbol{S_{all}}/\{i\}} \frac{\boldsymbol{1}}{(\boldsymbol{M choose} |\boldsymbol{S}|)(\boldsymbol{M} - |\boldsymbol{S}|)} [\boldsymbol{f_x}(\boldsymbol{S} \cup \{\boldsymbol{i}\}) - \boldsymbol{f_x}(\boldsymbol{S})] \quad (6)$$

In addition, to compute the SHAP for each prediction, missing values in input features (those not included in the set S) are estimated using the k-nearest neighbors' algorithm (k-NN). This method estimates missing values by identifying similar instances based on available features, effectively preserving the data's natural patterns and relationships.

3. Results and discussion

3.1. Models validation and comparison

The results of the model comparison (GBM, SVM, and BGPR) are shown in Table S1 and Fig. 1. It shows that the GBM has substantial improvement in the prediction accuracy over the other two models. It demonstrates the lower mean absolute error (MAE) and mean squared error (MSE). The reason is attributable to its design and functioning, such as decision trees, which are good at capturing the complex and nonlinear relationships in the data, handling large data sets effectively, and adapting to noisy data [35]. The iterative boosting process also contributes to minimizing the residual errors from the previous iterations, resulting in a better performance on the predicting values. The robustness of all models is also validated with k-fold cross validation, and the results indicate the consistent superior performance of GBM over the two models (Appendix). Additionally, GBM offered a favorable balance between accuracy and computational efficiency for rapid training and tuning within the resource constraints. Overall, these factors provide a compelling justification for selecting GBM over other advanced models in this study.

The predicted vs. actual plots illustrate the performance of three models - GBM, SVM, and BGPR - by comparing their predicted values to the true target values. The GBM achieves the highest R^2 score of 0.81, demonstrating strong predictive capability and a close fit to the actual values. In contrast, the BGPR model attains an R^2 of 0.62, indicating moderate performance, while the SVM model yields the lowest R^2 of 0.31, suggesting weaker predictive accuracy. These results highlight the superiority of the GBM in capturing the variance within the dataset, whereas the other models exhibit varying degrees of underperformance.

The GBM was further evaluated through external validation using an independent dataset [32]. As expected, the model exhibited a noticeable decline in performance, with a Mean Absolute Error (MAE) of 19.39 and a Root Mean Squared Error (RMSE) of 23.90 (Fig. S1). This performance drop is understandable, likely due to domain shift and the model's limited exposure to experimental conditions not represented in the original training data (Fig. S5).

3.2. Global SHAP values

The relative importance of each feature in predicting REE bioleaching efficiency is illustrated in Fig. 2, based on global SHAP values, which represent the mean absolute contribution of each variable to model outputs across all samples. Higher SHAP values indicate greater influence on the model's predictions and provide insight into the key physicochemical parameters governing leaching outcomes.

The global SHAP value analysis provides insights into the key factors influencing the predicted recovery rate in the leaching of REEs collected from the literature. Among the 26 features analyzed, Si (%) exhibits the highest impact on model predictions, with a SHAP value of 11.64. High silica gel was known to hinder REE recovery due to the formation of amorphous silica gel $SiO_2 \bullet xH_2O$ under acidic condition. This gel can physically trap REEs, obstruct the pore network in mineral matrices, and increase slurry viscosity, therefore reducing the mobility and diffusivity of the leaching agents [16,29]. The polycondensation of dissolved silicic acid can happen at low pH, resulting in the gelation that impairs the mass transfer. The Si present in the mineral can also compete with REEs via complexation with the leaching acid, or interact with REEs to form less-soluble REEs silicate [15]. HREE content is the second most influential factor (8.59), reinforcing its well-established role in bioleaching due to their larger ionic size, stronger bond formation in minerals, and lower solubility in typical leaching solutions [3,6].

The pH is the third most influential feature with the SHAP value is 6.16, reflecting its critical role in leaching efficiency of REEs. The lower the pH of the leachate, the higher the extraction efficiency, which is

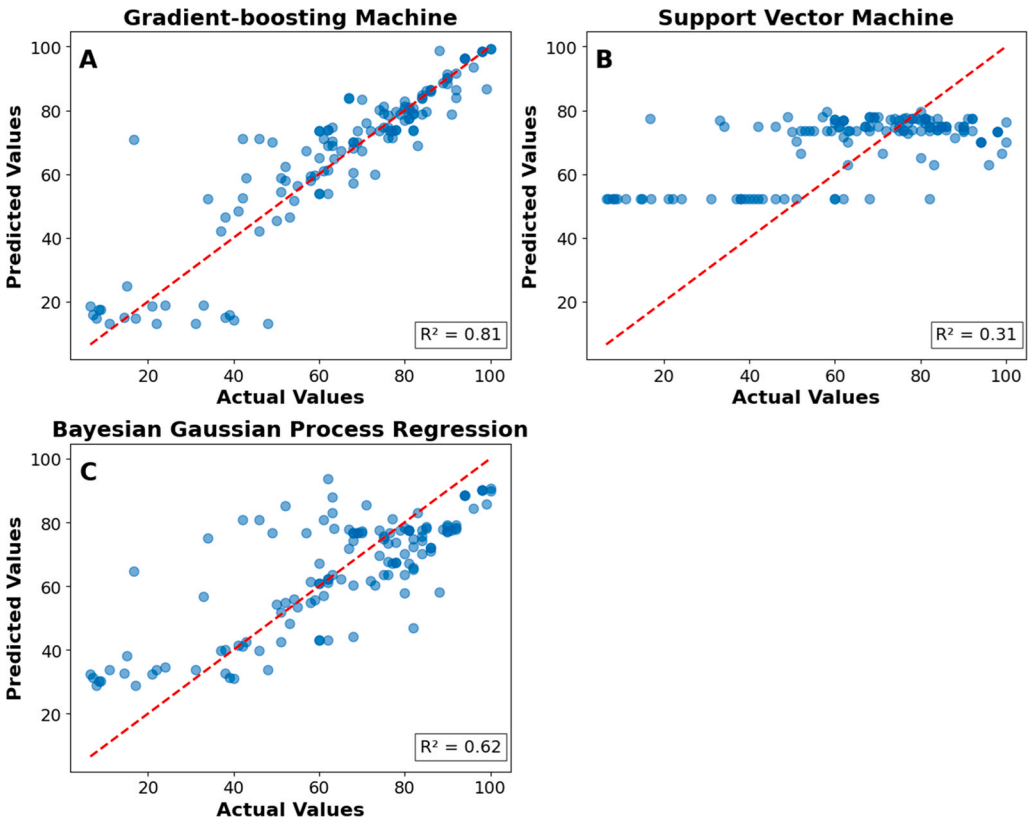


Fig. 1. Comparison of the prediction accuracy amongst three models. Scatter plots show predicted versus actual values for the Gradient-boosting Machines model (panel A), Support Vector Machine model (panel B), and Bayesian Gaussian Process Regression model (panel C). The x-axis represents actual values (y_{test}), and the y-axis represents predicted values. The red dashed line indicates the ideal prediction line ($y_{prediction}=y_{actual}$). Points closer to this line indicate better model performance. The R^2 values demonstrate predictive accuracy: GBM (R^2 =0.81) shows the best fit, followed by BGPR (R^2 =0.62), while SVM (R^2 =0.31) exhibits weaker predictive capability.

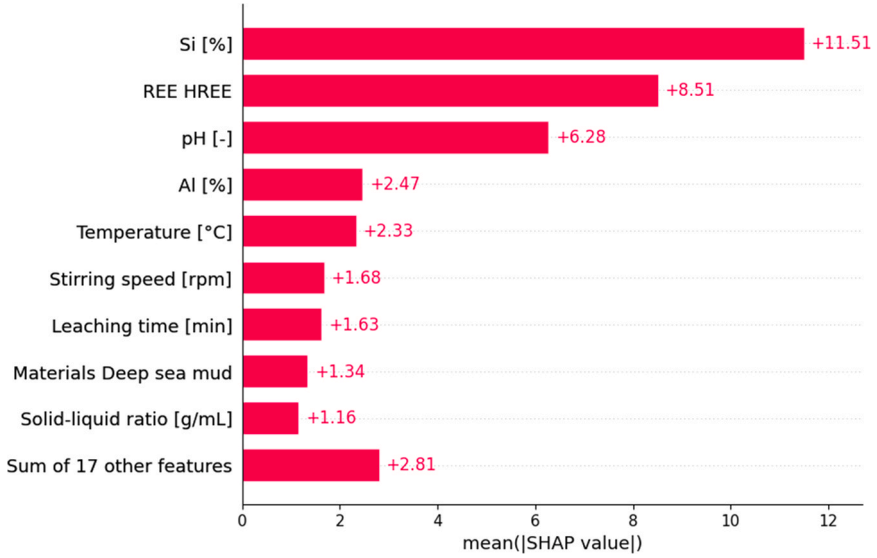


Fig. 2. Global SHAP values of extraction factors. SHAP (SHapley Additive exPlanations) values are used to distribute feature contributions, ensuring interpretability in model predictions.

attributable to excessive ion H^+ participating in the exchange reaction with REEs (Fig. 2) [23]. In addition, the high concentration of H^+ ions promotes the dissolution of REEs-bearing phases, such as monazite, basnasite, and apatite, via surface protonation and cleavage of metal-oxygen bonds [26,27]. The acidic environment also provide a favorable condition for acidophile species in bioleaching, contributing to maintaining the acid production and mineral breakdown [14]. The presence of Al (%) (2.40) also contributes notably, likely due to its involvement in redox reactions that facilitate bioleaching processes [36]. Temperature, stirring speed, and leaching time with SHAP values of 2.12, 1.97 and 1.49, respectively, further emphasize the importance of operating conditions in optimizing metal recovery and reaction

duration [25,28]. For example, increased temperatures improve the leaching kinetics by enhancing the rate constants of surface reactions and the diffusion of reactants and products across the solid-liquid interface [37]. Particularly in the bioleaching system, temperature poses a significant effect on the microbial growth and acid production [22].

The remaining 19 features (such as type of solutions and materials) contribute 4.25 SHAP value to the overall predictive model. This highlights the presence of additional but less influential variables in the bioleaching process.

The correlation of Si (%), pH, REE HREE, and temperature with SHAP values is shown in Fig. 3. The SHAP dependence plot for Si (%) demonstrates a threshold behavior in its impact on REE recovery. When silica content is below approximately 20 %, its contribution to the predicted recovery rate is largely positive, with SHAP values ranging from + 5 to + 15. This suggests that moderate levels of silica may facilitate leaching, possibly through enhanced acid-mineral interactions or by not impeding reagent penetration. However, as Si (%) increases beyond around 25 %, the SHAP values drop sharply, indicating a strong negative effect on the leaching efficiency. In addition, SHAP value decreases with the increase of solution pH and HREE presence, indicating that lower pH and non-HREE content are favorable for leaching. On the other hand, an increase in temperature enhances the leaching efficiency.

The correlation of SHAP values and stirring speed appears non-monotonic (Fig. S2). At low stirring speeds (~50–100 rpm), SHAP values are generally positive, suggesting that gentle stirring enhances the predicted leaching efficiency. However, as stirring speed increases towards 300 rpm, the SHAP values decrease, with many values becoming slightly negative, implying a neutral or mildly detrimental impact on the leaching efficiency at moderate speeds. Further increases in stirring speed, particularly around 400–500 rpm, lead to a rise in SHAP values, indicating that higher agitation rates positively affect leaching efficiency once again. Beyond 500 rpm up to 600 rpm, the SHAP values cluster around zero, suggesting a diminishing influence at high stirring speeds. This trend suggests that an optimal range of stirring speed exists: both very low and moderately high stirring speeds favor improved leaching efficiency, whereas intermediate speeds (~300 rpm) may be less effective. Regarding leaching time, it exhibits a positive correlation with REEs recovery rate, indicating that prolonged contact time enhances extraction efficiency under the studied conditions (Fig. S1B). In contrast, the solid–liquid ratio (g/mL) shows a negative SHAP trend (Fig. S2), it suggests that higher solid concentrations (or lower dilution) reduce REE recovery efficiency in the model. This aligns with hydrometallurgical principles: higher pulp density can hinder mass transfer, acid penetration, or microbial activity in bioleaching [12].

3.3. Individual prediction and interpretation

To illustrate the predictive capacity and interpretability of the model, an individual sample was randomly selected and analyzed using SHAP values (Fig. 4). For this specific sample, the predicted bioleaching recovery rate was 79.017 %, indicating a substantial increase (14.469 %) over the baseline recovery (64.548 %). This baseline represents the model's predicted output when all feature contributions are absent, hence enabling the attribution of incremental gains to individual experimental parameters.

Among the input features, Si (%) emerged as the most influential positive contributor, with a SHAP value of + 7.88. The silica concentration of 2.91 % appears to enhance leaching performance, likely due to the association of REEs with silicate phases that promote the mobilization under acidic conditions. This aligns with prior studies reporting the structural role of silica in REE-hosting minerals. The presence of light REEs (LREEs) as opposed to heavy REEs (HREEs) yielded a significant positive impact (+6.85), suggesting favorable interactions between LREEs and the leaching environment, possibly due to their distinct geochemical behavior and mineral associations [3]. Stirring speed at

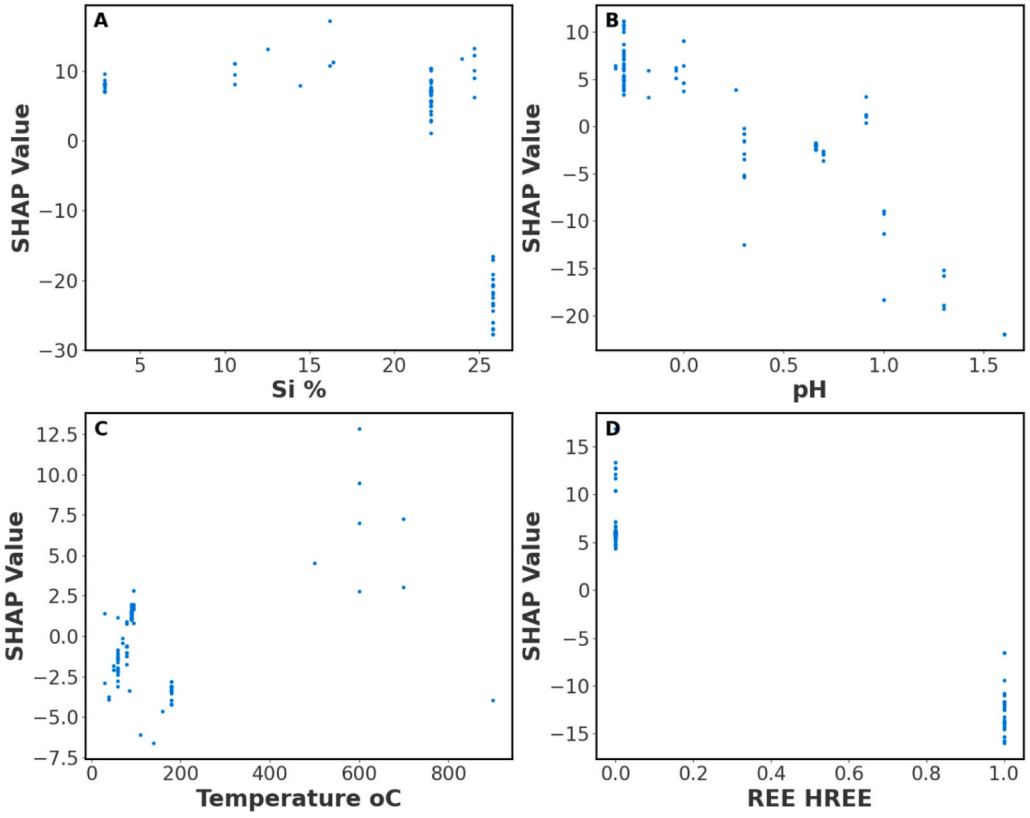


Fig. 3. The correlation of the %Si (Panel A), pH (Panel B), temperature (Panel C), and REE HREE (Panel D) and their corresponding SHAP values to predict leaching efficiency rates.

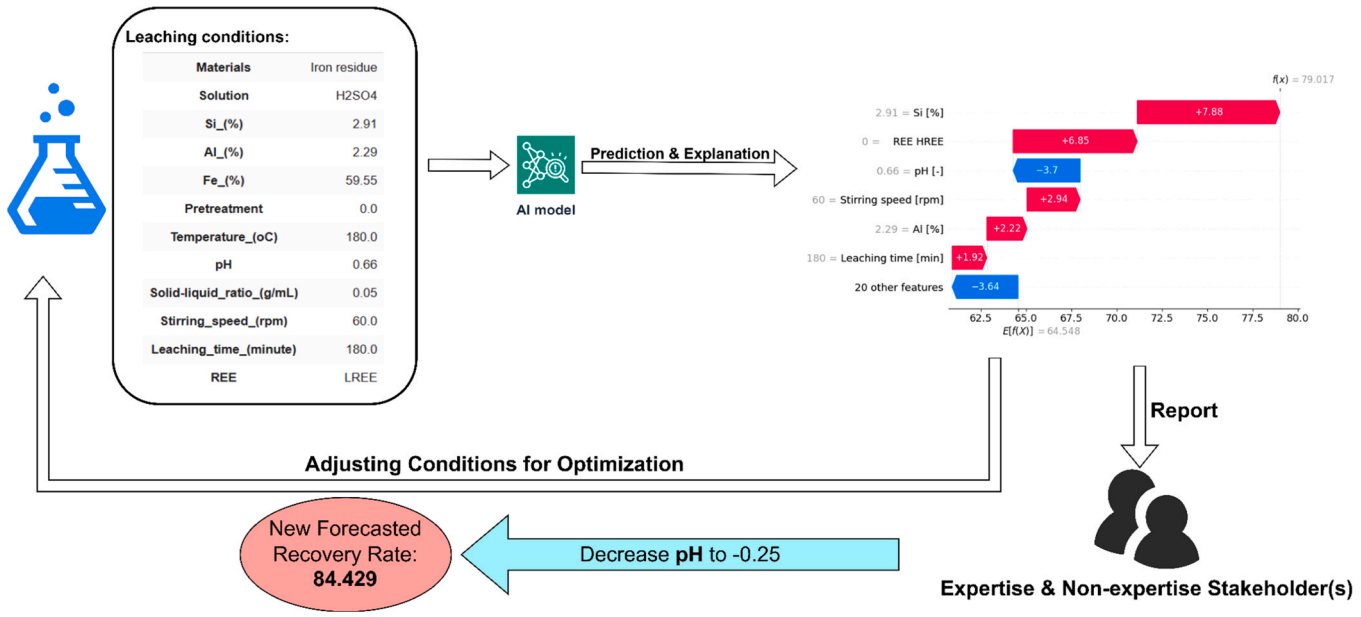


Fig. 4. Workflow of the proposed GBM illustrated with a *r*andom *e*xample. The SHAP value analysis reveals the most influential features that drive this prediction, highlighting both positive and negative contributions. Red features indicate increased value associated with an increased leaching rate on the final prediction, whereas blue features indicate a decreased leaching rate.

60 rpm also contributed positively (+2.94), reinforcing the role of mechanical agitation in enhancing mass transfer and solid-liquid contact during bioleaching.

In contrast, pH (0.66) was associated with negative contributions, at −3.70. These values imply that, under the experimental conditions studied, the acidity was suboptimal for maximizing REE recovery. This insight prompted a hypothesis: further tuning of this parameter—specifically, lowering the pH—could yield substantial performance gains. To test this, a guided optimization was performed by modifying pH alone-decreasing it from 0.66 to −0.25, while all other parameters remained constant. The updated model prediction yielded 84.429 %, reflecting an increase of 5.412 % over the initial prediction. The guidance of optimization can be incorporated with the overall trend of the impact of individual factors on the recovery rate (Fig. 3). By identifying and quantifying the influence of individual features, the SHAP analysis together improves model transparency and enables rational experimental design (Fig. S4, Fig. S5). These results strengthen the importance of integrating explainable AI into the experimental workflows.

3.4. Implications and future work

The application of AI in REEs extraction is still scarce. In the previous study, Bashiri et al. [2] used AI to predict the recovery efficiency of REEs from the sorbent. It was concluded that the molecular weight of the first functional group is the most impactful factor in the recovery efficiency, while pH is less critical in this context. Similarly, the cation exchange capacity of a sorbent is the most important feature of heavy metals recovery [38]. It suggests that the chemical bonding of REEs in the mining materials is more important than the engineering factors, such as stirring speed, and leaching time. Those factors should be considered for future studies of REEs leaching and recovery.

The insights drawn from both individual and global SHAP values of this study can inform process optimization strategies in the bioleaching industry. The consistently high importance of Si (%) across both individual predictions and the global analysis suggests that optimizing silicon levels with the bioleaching agents may significantly enhance recovery rates. Similarly, the positive impact of leaching time further supports the need for extended leaching periods to maximize extraction efficiency.

On the other hand, the negative contributions of pH indicate that adjustments to the parameter may be necessary to mitigate their inhibitory effects, given that lower pH values support the leaching process [17,7]. The global and local SHAP values, focusing on this key parameter, along with controlling other variables, could lead to more efficient and cost-effective bioleaching processes. Importantly, guided adjustments based on these insights resulted in improved predicted outcomes, demonstrating the model's capacity to inform actionable decision-making. This capability shifts the role of machine learning from passive prediction to active decision support, enabling the rational design of extraction processes grounded in both empirical evidence and scientific reasoning.

While the model demonstrated strong predictive performance during internal cross-validation, a noticeable drop was observed during external validation. This is likely attributable to domain shift, wherein the external data differ in mineral composition, pretreatment history, experimental protocols, or measurement precision, which were not fully captured in the training set. Such shifts are common in bioleaching datasets compiled from heterogeneous literature sources and highlight the need for more standardized and comprehensive datasets to improve model generalizability. Even though the performance of the model is dropped in external validation, the predicted recovery trends on the independent samples still captured the change direction, indicating that the model encodes meaningful mechanistic insights into the bioleaching process. As such, this work lays the foundation for scalable AI-driven approaches in sustainable rare earth extraction.

Moreover, although SHAP analysis greatly enhances interpretability by assigning feature-wise contributions, it is important to recognize its limitations. SHAP assumes feature additivity and conditional independence, which may not hold in systems with strong nonlinear feature interactions. Additionally, its attributions can be influenced by imbalanced training data or hidden biases, especially when key experimental conditions are underrepresented. As such, SHAP insights should be interpreted in conjunction with chemical knowledge and experimental context. Future work will aim to incorporate interaction-aware interpretation methods and expand the dataset to better capture the diversity of REE leaching systems.

Future work will also need to consider experimental validation of the predicted optimal conditions, focusing on verifying the impact of key

variables such as silica concentration, REE type, and pH on the leaching efficiency. These experiments will confirm the reliability of the model and provide new data to further improve its performance, strengthening its potential integration into real-world REE recovery applications.

4. Conclusion

The AI model identifies silica concentration, REE classification, and pH as the most influential factors governing the recovery of REEs from secondary resources. The developed AI model also provides a high level of accuracy in the prediction of REEs recovery (R^2=0.81). More broadly, this AI model provides a transferable template for optimizing complex secondary resources where performance, resource efficiency, and environmental impact must be simultaneously balanced. As global supply chains of REEs face increasing pressure, data-driven and transparent tools such as this will be essential for guiding low-impact, high-yield resource strategies.

Environmental implications

This research facilitates the recovery of hazardous pollutants from waste streams such as mine tailings, electronic waste, and industrial by-products. By enabling the recycling of toxic metals for use in advanced manufacturing, it simultaneously prevents their uncontrolled leaching into the environment—serving as both a recovery and preventive pollution control strategy.

CRediT authorship contribution statement

Andrei Herdean: Validation, Resources. **Le Long Pham:** Writing – review & editing, Validation. **Unnikrishnan Kuzhiumparambil:** Writing – review & editing. **Nature Poddar:** Formal analysis, Data curation. **Quang Loc Nguyen:** Writing – original draft, Conceptualization. **Vo Phong:** Writing – review & editing, Supervision, Investigation, Funding acquisition. **Nguyen Hong T. M.:** Writing – review & editing. **Lai Nguyen Huy:** Validation, Data curation. **Peter J. Ralph:** Writing – review & editing, Resources, Methodology. **Mathieu Pernice:** Writing – review & editing. **Ai Nguyen:** Writing – review & editing.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Phong Vo reports financial support was provided by University of Technology Sydney. I would like to disclose that I am currently serving as an Editorial Board Member of the Journal of Hazardous Materials; however, I have had no involvement in the editorial processing or peer review of this manuscript. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Phong H.N. Vo is funded by the University of Technology Sydney (UTS) Chancellor's Research Fellowship. The authors acknowledge the critical comments and suggestions from Dr. Danh Tai Hoang (National Institutes of Health, USA). The authors would like to acknowledge the assistance of ChatGPT (OpenAI) in refining and polishing the language of this manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jhazmat.2025.139479.

Data availability

Data will be made available on request.

References

[1] Abbadi, A., Mucsi, G., 2024. A review on complex utilization of mine tailings: recovery of rare earth elements and residue valorization. J Environ Chem Eng 12 (3), 113118. https://doi.org/10.1016/j.jece.2024.113118.
[2] Bashiri, A., Habibi, M., Sufali, A., Shekarsokhan, S., Maleki, R., Razmjou, A., 2024. Artificial intelligence models for efficiency estimation of adsorbents in Rare-Earth element recovery based on feature engineering. Ind Eng Chem Res 63 (42), 17930–17948. https://doi.org/10.1021/acs.iecr.4c01935.
[3] Battsengel, A., Batnasan, A., Narankhuu, A., Haga, K., Watanabe, Y., Shibayama, A., 2018. Recovery of light and heavy rare earth elements from apatite ore using sulphuric acid leaching, solvent extraction and precipitation. Hydrometallurgy 179, 100–109.
[4] Bentéjac, C., Csörgő, A., Martínez-Muñoz, G., 2021. A comparative analysis of gradient boosting algorithms. Artif Intell Rev 54 (3), 1937–1967. https://doi.org/10.1007/s10462-020-09896-5.
[5] Cardoso, A.C., Dias, C.S., Moura, C.H.Rd, Ferreira, J.L., Rodrigues, E.C., Macêdo, E. N., Estumano, D.C., Viegas, B.M., 2023. Use of Bayesian methods in the process of uranium bioleaching by acidithiobacillus ferrooxidans. Appl Sci 14 (1), 109.
[6] Chen, S.-Y., Lin, J.-G., 2001. Bioleaching of heavy metals from sediment: significance of ph. Chemosphere 44 (5), 1093–1102.
[7] Cunningham, S., Etherington-Rivas, M., Azimi, G. 2025. Effect of Agitation, pH, and Particle Size on Rare Earth Element Extraction from an Ionic Clay. *Proceedings of the 63rd Conference of Metallurgists, COM 2024*, 2025//, Cham. Springer Nature Switzerland. pp. 1355-1360.
[8] Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann Stat 1189–1232.
[9] Friedman, J.H., 2002. Stochastic gradient boosting. Comput Stat data Anal 38 (4), 367–378.
[10] Ghorbani, Y., Ilankoon, I.M.S.K., Dushyantha, N., Nwaila, G.T., 2025. Rare earth permanent magnets for the Green energy transition: bottlenecks, current developments and cleaner production solutions. Resour Conserv Recycl 212, 107966. https://doi.org/10.1016/j.resconrec.2024.107966.
[11] Guo, J., Ranković, B., Schwaller, P., 2023. Bayesian optimization for chemical reactions. CHIMIA 77 (1/2), 31–38. https://doi.org/10.2533/chimia.2023.31.
[12] Josso, P., Roberts, S., Teagle, D.A., Pourret, O., Herrington, R., de Leon Albarran, C. P., 2018. Extraction and separation of rare earth elements from hydrothermal metalliferous sediments. Miner Eng 118, 106–121.
[13] Li, J., Xiao, Y., Feng, X., Wang, J., Ma, Z., Yao, R., Zhai, Y., Tian, L., 2024. Leaching of ion adsorption rare earths and the role of bioleaching in the process: a review. J Clean Prod 468, 143067. https://doi.org/10.1016/j.jclepro.2024.143067.
[14] Liang, Y., Huang, Z., Xiao, Q., Guan, X., Xu, X., Jiang, H., Zhong, Y., Tu, Z., 2025. Dissolution and mechanism of bastnaesite mediated by acidithiobacillus ferrooxidans. J Rare Earths. https://doi.org/10.1016/j.jre.2025.03.025.
[15] Lin, Y., Li, S., Li, X., Lin, H., Lei, N., Wu, D., Tong, J., Xie, H., 2023. Bioleaching of silicon from Fly ash by Co-culture of silicate bacteria and fungi. Water Air Soil Pollut 234 (12), 755. https://doi.org/10.1007/s11270-023-06775-x.
[16] Liu, W., Xiao, Z., Das, S., Zhang, W., 2024. Mechanism and kinetic study of rare earth extraction from allanite by direct acid leaching. Miner Eng 205, 108489. https://doi.org/10.1016/j.mineng.2023.108489.
[17] Lokshin, E.P., Tareeva, O.A., Elizarov, I.R., 2016. Agitation leaching of rare earth elements from phosphogypsum by weak sulfuric solutions. Theor Found Chem Eng 50 (5), 857–862. https://doi.org/10.1134/S0040579516050134.
[18] Lundberg, S., 2017. Arxiv preprint. A Unifi Approach Interpret Model Predict. *arXiv:1705.07874*.
[19] Luo, Y., Tseng, H.H., Cui, S., Wei, L., Ten Haken, R.K., El Naqa, I., 2019. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. BJR Open 1 (1), 20190021. https://doi.org/10.1259/bjro.20190021.
[20] Matkets, Ma. 2024. Rare Earth Metals Market.
[21] Mehrotra, D., 2019. Basics of artificial intelligence & machine learning. Notion Press.
[22] Modak, J.M., Natarajan, K.A., Mukhopadhyay, S., 1996. Development of temperature-tolerant strains of thiobacillus ferrooxidans to improve bioleaching kinetics. Hydrometallurgy 42 (1), 51–61. https://doi.org/10.1016/0304-386X(95)00072-O.
[23] Mokoena, B.K., Mokhahlane, L.S., Clarke, S., 2022. Effects of acid concentration on the recovery of rare earth elements from coal Fly ash. Int J Coal Geol 259, 104037. https://doi.org/10.1016/j.coal.2022.104037.
[24] Owusu-Fordjour, E.Y., Yang, X., 2023. Bioleaching of rare earth elements challenges and opportunities: a critical review. J Environ Chem Eng 11 (5), 110413. https://doi.org/10.1016/j.jece.2023.110413.
[25] Pan, J., Nie, T., Zhou, C., Yang, F., Jia, R., Zhang, L., Liu, H., 2022. The effect of calcination on the occurrence and leaching of rare earth elements in coal refuse. J Environ Chem Eng 10 (5), 108355. https://doi.org/10.1016/j.jece.2022.108355.
[26] Pan J., Zhao X., Zhou C., Yang F., Ji W.Effect of Agitation, pH, and Particle Size on Rare Earth Element Extraction from an Ionic Clay. Proceedings of the 63rd

Conference of Metallurgists, COM 2024, 2025//, Cham. Springer Nature Switzerland. pp. 1355-1360..
[27] Rahmati, S., Adavodi, R., Romano, P., Vegliò, F., 2025. Evaluation of the mechanisms of rare earth elements extraction from citrate solutions in the recycling of NdFeB magnets. Process Saf Environ Prot 195, 106788. https://doi.org/10.1016/j.psep.2025.106788.
[28] Rasoulnia, P., Barthen, R., Lakaniemi, A.-M., 2021. A critical review of bioleaching of rare earth elements: the mechanisms and effect of process parameters. Crit Rev Environ Sci Technol 51 (4), 378–427.
[29] Rogers, J.R., Bennett, P.C., 2004. Mineral stimulation of subsurface microorganisms: release of limiting nutrients from silicates. Chem Geol 203 (1-2), 91–108.
[30] Ruhatiya, C., Gandra, R., Kondaiah, P., Manivas, K., Samhith, A., Gao, L., Lam, J.S. L., Garg, A., 2021. Intelligent optimization of bioleaching process for waste lithium-ion batteries: an application of support vector regression approach. Int J Energy Res 45 (4), 6152–6162.
[31] Sahin, E.K., 2020. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Appl Sci 2 (7), 1308.
[32] Sarker, S.K., Pownceby, M., Yadav, S., Bruckard, W., Haque, N., Singh, N., Pramanik, B.K., 2024. Selective removal of fe impurities in the recovery of rare earth elements from carbonatite tailings using chemical routes. Hydrometallurgy 224, 106249. https://doi.org/10.1016/j.hydromet.2023.106249.
[33] Shields, B.J., Stevens, J., Li, J., Parasram, M., Damani, F., Alvarado, J.I.M., Janey, J.M., Adams, R.P., Doyle, A.G., 2021. Bayesian reaction optimization as a tool for chemical synthesis. Nature 590 (7844), 89–96. https://doi.org/10.1038/s41586-021-03213-y.
[34] Tuncay, G., Yuksekdag, A., Mutlu, B.K., Koyuncu, I., 2024. A review of greener approaches for rare earth elements recovery from mineral wastes. Environ Pollut 357, 124379. https://doi.org/10.1016/j.envpol.2024.124379.
[35] Utkarsh, Jain, P.K., 2024. Predicting bentonite swelling pressure: optimized XGBoost versus neural networks. Sci Rep 14 (1), 17533.
[36] Wang, S., Zheng, Y., Yan, W., Chen, L., Mahadevan, G.D., Zhao, F., 2016. Enhanced bioleaching efficiency of metals from E-wastes driven by biochar. J Hazard Mater 320, 393–400.
[37] Wu, X., Wang, Z., Xia, C., Shi, X., Luo, T., Bao, X., Liu, R., Xie, S., 2020. Kinetics study on leaching of rare earth and aluminum from polishing powder waste using hydrochloric acid. J Rare Earths 38 (9), 1009–1018. https://doi.org/10.1016/j.jre.2020.04.004.
[38] Zhu, X., Wang, X., Ok, Y.S., 2019. The application of machine learning methods for prediction of metal sorption onto biochars. J Hazard Mater 378, 120727. https://doi.org/10.1016/j.jhazmat.2019.06.004.