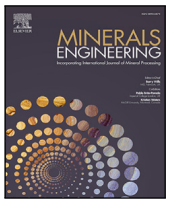
Minerals Engineering 232 (2025) 109551

Contents lists available at ScienceDirect

Minerals Engineering

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



Optimising SAG mill throughput and circulating load using machine learning models: A multi-objective approach for identifying optimal process parameters

Zahra Ghasemi [a,*], Mehdi Neshat [a], Chris Aldrich [b], Max Zanin [c,d], Lei Chen [a]

[a] School of Electrical and Mechanical Engineering, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
[b] Western Australian School of Mines, Curtin University, Perth, Western Australia, 6845, Australia
[c] School of Chemical Engineering, The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia
[d] School of Civil, Environmental and Mining Engineering, The University of South Australia, Mawson Lakes, Adelaide, SA 5095, Australia

ARTICLE INFO

Keywords:
Semi-autogenous grinding (SAG)
Mill throughput
Circulating load (CL)
Machine learning (ML)
Mult-objective optimisation
Evolutionary algorithm

ABSTRACT

Grinding is a fundamental operation in mineral processing plants. Semi-autogenous grinding (SAG) mills are widely utilised in these circuits. Enhancing SAG mill throughput is crucial due to its significant financial impact. However, controlling the mill discharge particle size is also essential. If the discharge particles are oversized, a substantial portion of the SAG mill output is recirculated to pebble crushers. Therefore, the challenge addressed in this research is how to set controllable process parameters to achieve both maximum mill throughput and minimum circulating load. This research addresses the problem by formulating it as a constrained multi-objective optimisation problem. An industrial dataset, consisting of 66,776 records is utilised. Initially, three gradient boosting methods including CatBoost, XGBoost, and HGBM are compared for modelling both objectives. The most precise prediction models are subsequently employed within four multi-objective optimisation algorithms consisting of NSGA-II, MOEA/D, RVEA, and SPEA2. The comparative results indicated that CatBoost is the best-performing prediction model with an average R^2 of 0.9304 and 0.9031 for mill throughput and circulating load, respectively. NSGA-II is also identified as the best-performing optimiser, achieving the highest average hypervolume of 291,489. A sensitivity analysis demonstrated how changes in input features affect objectives and confirmed the robustness of the optimal solutions. By analysing the top-performing solutions, recommendations for process parameter setting are provided. Additionally, the advantage of the developed framework in this research is providing a set of Pareto-optimal solutions, allowing experts to select the most suitable settings based on real-time process conditions.

1. Introduction

The separation of valuable minerals from gangue material comprises a sequence of complex processes, typically categorised into crushing, grinding, and concentration. Grinding, as the most energy-intensive step, is particularly critical, employing grinding media to achieve the necessary particle size reduction. This process is generally performed with the addition of water to produce slurry for the subsequent concentration stage (Wills and Napier-Munn, 2006). SAG mills are commonly utilised in the grinding circuits of mineral processing plants for particle size reduction, leveraging a mix of steel grinding media and coarse ore to effectively achieve this objective (Wills and Finch, 2015).

The throughput of a SAG mill is a crucial performance metric that quantifies the amount of ore processed by the mill per hour (Fuerstenau and Han, 2003). Enhancing SAG mill throughput is critically important because of its considerable financial impact. In addition to the quantity of material processed, the quality of grinding is also critical, as indicated by the particle size distribution of the SAG mill discharge. Oversized particles, which are larger than the screen's aperture size at SAG mill discharge, are recirculated back into the grinding circuit for further size reduction in pebble crushers. A high circulating load is unfavourable as it results in additional operational costs. These costs arise from increased energy consumption required to reprocess material in the pebble crushers and the SAG mill, as well as the wear and tear on this equipment. Furthermore, a high circulating load reduces the effective capacity of the mill to process new feed material. Consequently, it is crucial not only to maximise mill throughput but also to minimise the circulating load, thereby optimising both the quantity and quality of the grinding process.

* Corresponding author.
E-mail address: zahra.ghasemi@adelaide.edu.au (Z. Ghasemi).

https://doi.org/10.1016/j.mineng.2025.109551
Received 30 December 2024; Received in revised form 17 June 2025; Accepted 18 June 2025
Available online 1 July 2025

However, achieving this balance is a complex and challenging task. Grinding is an inherently intricate process influenced by numerous factors, such as ore properties and mill conditions. Impactful factors have dynamic interactions and the interdependencies between these factors make it difficult to isolate the effect of each factor. Consequently, these complexities make it challenging to determine the optimal process settings that satisfy both objectives simultaneously.

Numerous studies have investigated mill throughput. Some focus on the development of empirical models predominantly based on Bond's equation (Bond, 1961), a widely used method for calculating the specific energy of grinding mills. Examples of these models include Morell's model (Morrell, 2004), as well as models developed in Minera Escondida mining sites (Flores and Limitada, 2005) and Collahuasi grinding circuit (Alruiz et al., 2009).

Several other research works focus on methods to enhance mill throughput through practical experiments and simulations. Behnamfard et al. (2020) demonstrated that increasing the proportion of coarse material (larger than 200 mm) in the mill feed from 10% to 30% significantly improves throughput, although further increases beyond 45% negatively affect performance. Powell et al. (2009) found that mill throughput reaches its peak within a filling range of 23% to 34%, and that higher turning speeds can further boost throughput. Additionally, research conducted at the Antamina mining complex (Rybinski et al., 2011) and the Cadia Hill SAG mill circuit (Hart et al., 2001) highlighted the substantial impact of adjusting feed size, ball charge, and lifter configuration on throughput.

These studies provide insightful findings while also emphasising the intricate relationship between effective factors and mill throughput. However, real-world experiments are often costly and time-consuming. Such approaches do not offer a generalised framework and are constrained by the specific conditions under which they were conducted. That is while the grinding process itself is dynamic and subject to significant variability in critical factors, including material properties (such as input ore hardness and particle size distribution), equipment conditions (such as the state of liners and grates), and operational parameters (such as rotational speed and inlet water flow). Consequently, even if a study provides an empirical model for mill throughput or recommendations for enhancing it, these findings may not be applicable with the same level of accuracy in future conditions for the same mining plant or to other mining plants due to the inherent variability of these factors. This highlights the importance of accounting for such variability and developing dynamic models that can be continuously updated based on the most recent data. Unlike static models, which remain fixed, dynamic models adapt to current conditions, reflecting changes in the influential factors. This is where machine learning (ML) can play a pivotal role in modelling SAG mill throughput. On one hand, recent advancements in sensor technology have generated a wealth of data in mining operations, which can be leveraged for ML modelling. On the other hand, these models can be continually updated with new data inputs, enabling a dynamic and adaptive approach to process modelling.

In some research works, ML models have been used for predicting mill performance. In most of them, they are focused on predicting power draw (Curilem et al., 2011; Hoseinian et al., 2017; Avalos et al., 2020; Katoch et al., 2021; Ruiz et al., 2024). Both and Dimitrakopoulos (2021) used neural networks (NN) and multiple linear regression (MLR) for predicting ball mill throughput in Tropicana Gold mining complex. Their results demonstrated the superior accuracy of NN models. Their results showed a significant correlation between the mill feed size, power consumption, and mill throughput. Ou et al. (2023) integrated Discrete Element Modelling (DEM) with the XGBoost algorithm, to create a robust and accurate mill throughput prediction model. This method leverages DEM simulations, incorporating various configurations of worn shell liners to analyse the impact of liner wear on collisional energy. The study found that total impact energy increased and then plateaued as shell liners wore, while shear collisional energy peaked at 4 to 6 months of liner service life before declining. Further analysis of feature importance within the XGBoost model revealed that water feed and liner wear had the most significant impact on mill throughput, followed by ore properties. Ghasemi et al. (2024b) compared six ML models and found that recurrent neural networks (RNNs) were the most accurate for predicting SAG mill throughput. Their results demonstrated the significant impact of SAG mill turning speed and inlet water on SAG mill throughput.

Although machine learning approaches have been utilised to model mill throughput in previous studies, there remains a considerable gap in the literature regarding the identification of optimal control parameters that optimise SAG mill performance. That is where artificial intelligence (AI)-based optimisation algorithms can be utilised to find optimal settings. Ghasemi et al. (2024a) addressed this problem by developing an integrated intelligent framework that first models mill throughput using machine learning techniques and then determines the optimal parameter settings to maximise throughput using evolutionary algorithms. In their study, they compared 17 machine learning models, including ensemble models, neural networks, tree-based models, and traditional ML approaches. For the optimisation purpose, they employed particle swarm optimisation (PSO), differential evolution (DE), and genetic algorithms (GA). Their analysis revealed that Categorical Boosting (CatBoost), an ensemble model, was the most accurate predictor, while DE emerged as the most effective optimisation algorithm, achieving the highest mill throughput predictions with superior robustness compared to other methods. However, the study focused solely on optimising mill throughput as a single objective, without addressing circulating load, which serves as a key indicator of grinding quality.

This research is aimed at addressing this gap by developing a comprehensive framework that considers both mill throughput and circulating load as objectives. By integrating these objectives, the study seeks to optimise SAG mill performance, ensuring improvements in both the quantity and quality of the grinding process. An industrial dataset from an Australian mining complex is used for this purpose. Initially, 17 features have been selected according to recommendations from professionals in the industry, including mill weight, mill turning speed, inlet water, ore feed fresh ratio, three conveyor readings, and input particle size distribution. Three gradient boosting models including CatBoost, Histogram-Based Gradient Boosting Machine (HGBM), and Extreme Gradient Boosting (XGBoost) are compared to determine the most accurate predictive model. Once the most accurate model is identified, four multi-objective evolutionary algorithms (EAs), including Non-dominated Sorting Genetic Algorithm II (NSGA-II), Strength Pareto Evolutionary Algorithm 2 (SPEA2), Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D), and Reference Vector Guided Evolutionary Algorithm (RVEA), are evaluated using the best predictive model as a surrogate to assess their optimisation performance. Finally, the best prediction model, integrated with the top-performing multi-objective optimisation algorithm, is used to find optimal set points that maximise mill throughput and minimise circulating load. A sensitivity analysis is performed to assess the robustness of the optimal solutions and to understand the impact of changes in any features on these solutions. The proposed optimal settings are presented as a parallel plot and suggestions for adjusting the set points are provided.

2. Process description and data analysis

The dataset utilised in this study consists of 66,776 operational records obtained from a Gold mining complex in Australia. These data represent per-minute recordings, spanning approximately 1.5 months of operation. The mineral composition of the ore includes albite, sericite, chlorite, biotite, pyrrhotite, pyrite, and arsenopyrite. Key parameters from the comminution test work indicate an average Crushing Work Index (CWi) of 6.0 kWh/t, Bond Rod Work Index (RWi) of 20.8 kWh/t,

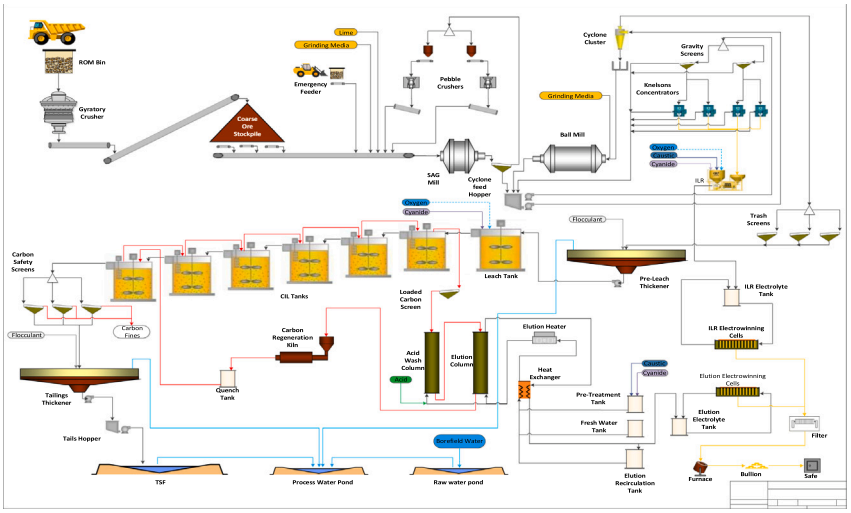


Fig. 1. Grinding circuit flowsheet.

Bond Ball Work Index (BWi) of 17.6 kWh/t, and an Abrasion Index (Ai) of 0.49 g.

Fig. 1 presents a schematic of the grinding process at this mining complex. The ore is initially crushed by gyratory crushers and then transported to a coarse ore stockpile, which supplies the SAG mill via three feeders. The SAG mill has a diameter of 10.97 metres and an effective grinding length of 5.79 m, driven by a 15 MW dual-pinion system. The mill discharge is screened by a vibrating screen, with oversized material recirculated to two pebble crushers for further size reduction before being returned to the SAG mill feed conveyor. The undersized material from the screen is transferred to the mill discharge hopper and subsequently pumped to a cyclone cluster for classification.

In this mining complex, operational data is recorded in real time through a network of circuit sensors. Laser diffraction analysers are employed to monitor the particle size of the ore fed into the SAG mill. As the crushed ore moves along the conveyor belt, the analysers emit a laser beam onto the material, and the resulting light scattering patterns are analysed to provide detailed information on particle size distribution. The dataset employed in this research includes 10 variables related to particle size, with the percent passing values ranging from 13.2 to 300 mm. Other input variables include mill weight, mill rotation speed, inlet water flow, fresh feed ratio (indicating the proportion of fresh ore relative to the total feed of both fresh and oxide ores), output from the pebble crushers (CV006 and CV007), and the mass flow of uncrushed pebbles recirculated to the mill (CV008).

Initially, the dataset contained some erroneous measurements, such as negative values, percentage values exceeding 100%, and data recorded during non-operational periods of the mill. Following the removal of these invalid entries, 6764 records were discarded, yielding a final dataset of 60,012 valid observations. Table 1 provides the descriptive statistics of the cleaned dataset, and Fig. 2 presents histograms for all variables. Bivariate distribution plots for some features are also represented in Fig. 3. The contours in this Figure represent areas of equal data density, with red areas indicating regions where the data points are most concentrated. As shown in this Figure, there is a dense cluster of operating points where throughput ranges between 1350 to 1400(t/h) and circulating load is between 300 to 350(t/h). The maximum mill throughput typically remains under 1500(t/h) for a wide range of circulating load values. However, there are some data points with higher throughputs, approaching 1700(t/h). In particular, there is a data point with a throughput of around 1650(t/h) at a circulating load of approximately 250 (t/h). Further investigation is needed to determine whether these rare observations are related to normal operating conditions or are the result of measurement noise or transient operational factors that are not captured in the available dataset. While these instances suggest the potential for achieving higher throughput at lower circulating loads, more detailed process data is needed to confirm the consistency of such performance.

Table 1
Basic descriptive statistics of the data (updated)

Variable	Unit	Min	Max	Mean	Std. Deviation
Mill throughput	t/h	900.08	1672.30	1292.68	120.74
Circulating load	t/h	10.98	456.22	309.71	47.65
Mill weight	t	436.02	701.48	601.87	25.00
Mill turning speed	rpm	7.74	10.30	10.09	0.29
Inlet water	m^3/h	140.56	656.70	431.34	49.12
Fresh ratio	%	70.00	100.00	88.40	6.39
%PL 13.2 mm	%	19.37	43.18	30.89	2.96
%PL 13.2–19 mm	%	0.51	11.93	5.28	0.91
%PL 19–26.5 mm	%	3.00	12.00	7.81	1.00
%PL 26.5–37.5 mm	%	5.00	13.00	9.37	0.88
%PL 37.5–53 mm	%	6.00	11.00	9.12	0.58
%PL 53–75 mm	%	8.00	15.00	12.51	0.73
%PL 75–106 mm	%	6.00	23.00	14.23	2.11
%PL 106–150 mm	%	2.00	27.00	8.31	2.56
%PL 150–212 mm	%	0.00	10.00	1.63	0.88
%PL 212–300 mm	%	0.00	2.00	0.79	0.40
CV006	t/h	0.00	245.18	127.69	70.31
CV007	t/h	0.00	248.96	131.44	67.36
CV008	t/h	0.00	478.13	41.73	72.78

By investigating the relationship between mill weight and throughput, it can be seen that the majority of data points fall within the 600 to 650(t) range for mill weight and around 1400 to 1450 (t/h) for mill throughput. Another cluster can be seen slightly below 600(t) in mill weight, with throughput from 1200 to 1300(t/h). It seems to be a tendency for higher mill weight to correspond to higher throughput values.

For mill speed, it is mostly set at the maximum (10.3 rpm), with a high concentration of data points at this speed and throughput values between 1200 and 1350(t/h). However, a wide range of throughput values is observed at the maximum turning speed, suggesting that other input variables significantly influence mill throughput even when speed remains constant.

Fresh feed ratio is a discrete variable with distinct values (e.g., 70%, 75%, 77%, etc.). The bivariate plot of fresh feed ratio versus circulating load shows that most observations fall within a fresh feed ratio of approximately 87% to 91% and a circulating load between 300 and 350(t/h). Another noticeable cluster is observed around a 100% fresh feed ratio with circulating loads around 300 (t/h), representing multiple operating conditions.

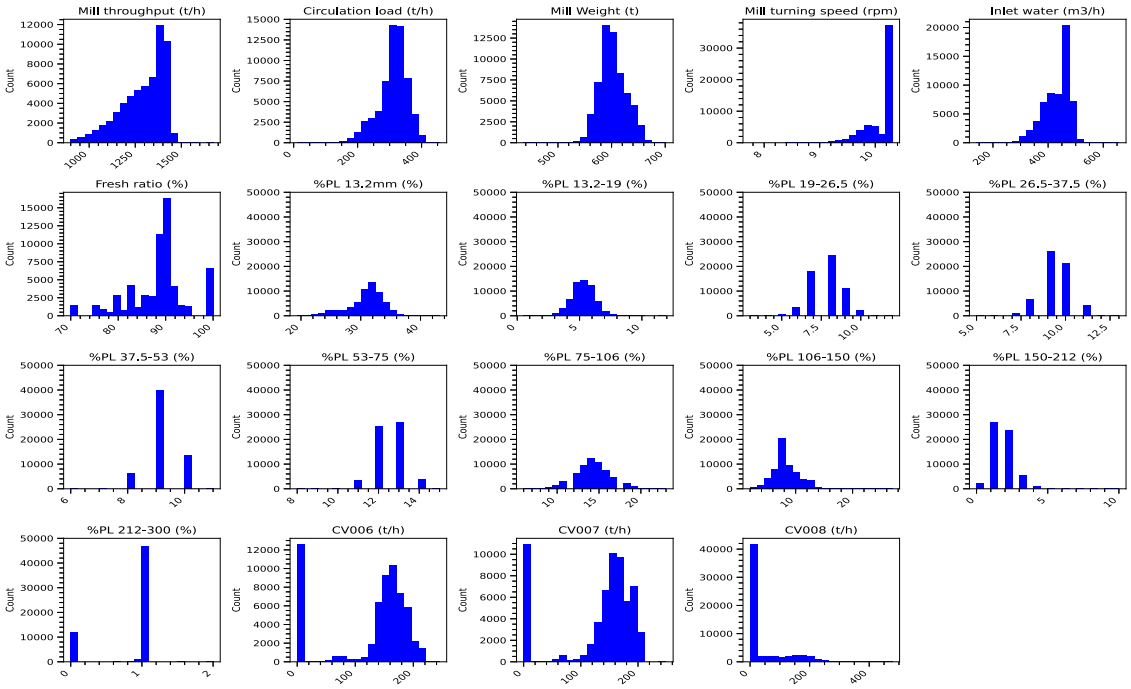


Fig. 2. Histogram plot of all features.

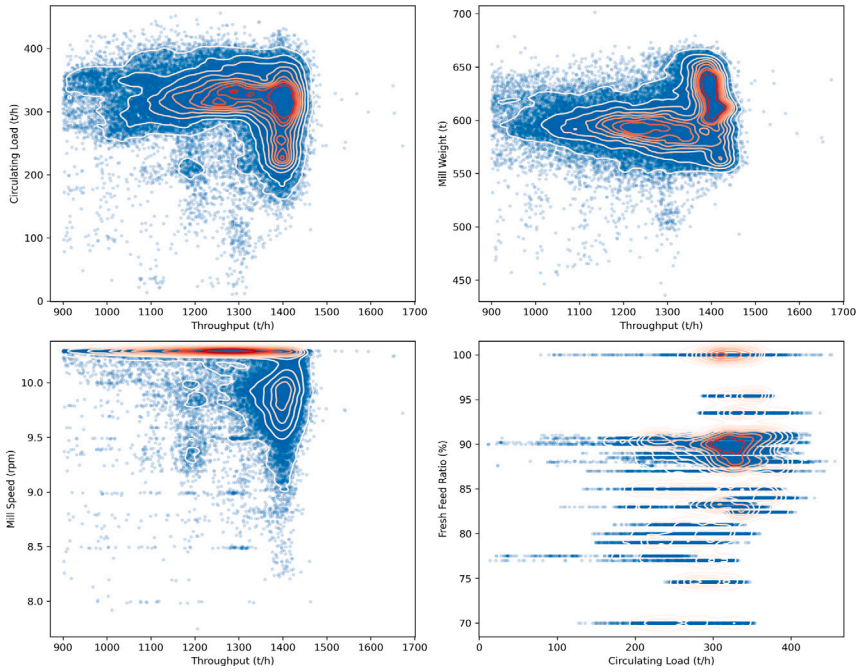


Fig. 3. Bivariate distribution plots for selected features.

3. Methodology

3.1. Prediction methods

In contrast to traditional data-driven modelling techniques that rely on the development of a single predictive model, ensemble-based approaches have emerged as a robust alternative, combining the predictions from multiple base estimators to enhance predictive performance and model robustness. These methods have proven highly effective in addressing the limitations of individual models, such as overfitting and sensitivity to noise, by leveraging the collective strength of multiple learners (Dietterich, 2000). The principal ensemble learning techniques include bagging (Breiman, 1996), boosting (Freund and Schapire, 1997), and stacking (Wolpert, 1992), each of which employs distinct mechanisms for improving predictive accuracy, generalisation, and robustness.

Bagging, or Bootstrap Aggregating, entails generating several variations of the training dataset by random sampling with replacement. These datasets are used to independently train the same type of base model, with the final prediction derived from an aggregation of the predictions from all models in the ensemble, typically through averaging or majority voting. The independence of model training in bagging not

only enables parallelisation but also reduces variance and mitigates the risk of overfitting.

In contrast, boosting operates by training models sequentially, with each subsequent model focused on correcting the errors of its predecessors. The fundamental idea in boosting is to assign higher weights to data points that are misclassified or have higher prediction errors in successive iterations, thereby directing the learning process towards the most challenging instances. This adaptive mechanism iteratively improves model performance, transforming a collection of weak learners into a strong learner. Boosting is regarded as one of the most successful learning techniques in machine learning (Hastie et al., 2009).

Stacking, or stacked generalisation, presents a more sophisticated ensemble methodology by introducing a two-layer structure. In stacking, multiple base models are first trained on the original dataset, and their predictions are then used as inputs to a second-level model, often referred to as a meta-learner. This meta-model is trained to optimally combine the predictions of the base models, capturing potential interactions and patterns among them that may not be discernible to the individual learners.

In this research, due to the recognition of boosting method as one of the most effective learning techniques (Hastie et al., 2009), along with its superior performance demonstrated in the previous study (Ghasemi et al., 2024a), three boosting methods including XGBoost (Chen and Guestrin, 2016), CatBoost (Prokhorenkova et al., 2018), and Histogram-based Gradient Boosting Machine (HGBM) (Ke et al., 2017) are employed as prediction models.

XGBoost is an advanced implementation of the Gradient Boosting Machine (GBM) (Friedman, 2001) algorithm that enhances predictive accuracy while controlling model complexity. Like GBM, XGBoost employs a forward stage-wise strategy, where models are sequentially added to the ensemble to refine predictions. Typically, regression trees (RTs) serve as the weak learners in this process. XGBoost improves upon standard GBM by introducing regularisation techniques to avoid overfitting, making it especially effective for high-dimensional data and complex problems. The basic structure of XGBoost begins with the specification of a differentiable loss function. In many cases, the squared error loss is used for regression problems, which is expressed as:

$$L(y, F(x)) = \frac{1}{2}(y - F(x))^2, \tag{1}$$

where y represents the true value, and $F(x)$ represents the prediction from the current model. The training process begins by initialising the model with a constant prediction, $F_0(x)$, which minimises the loss function across all data points. Mathematically, this is computed as:

$$F_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i, \gamma), \tag{2}$$

where γ is the constant prediction value, and the result of this initialisation, when using squared error loss, is the mean of the target variable across the training dataset. Subsequent steps involve computing the residuals, which represent the difference between the true values and the model's current predictions. These pseudo-residuals, r_{im}, are calculated for each data point as the negative gradient of the loss function:

$$r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}, \tag{3}$$

where $F_{m-1}(x)$ is the prediction from the previous iteration. These residuals are used as the target variable for fitting the next weak learner (often a regression tree) to the dataset. Once the weak learner is fitted, XGBoost solves an optimisation problem to compute the optimal multiplier ρ_m, which scales the contribution of the new learner. This is done by minimising the loss function:

$$\rho_m = \arg\min_{\rho} \sum_{i=1}^{n} L(y_i, F_{m-1}(x_i) + \rho h_m(x_i)), \tag{4}$$

where $h_m(x)$ represents the new weak learner. The model is then updated as follows:

$$F_m(x) = F_{m-1}(x) + \gamma_m, \tag{5}$$

where $\gamma_m = \rho_m h_m(x)$, incorporating the scaled contribution of the new learner.

XGBoost extends GBM by incorporating additional regularisation terms into the objective function to prevent overfitting and reduce the risk of overly complex models. The overall objective function for XGBoost is defined as:

$$\text{Objective}(F) = \sum_{i=1}^{n} \left[l(y_i, F(x_i)) + \Omega(f_i)\right], \tag{6}$$

where $l(y_i, F(x_i))$ is the loss function, measuring the discrepancy between the actual and predicted values, and $\Omega(f_i)$ is the regularisation term that penalises model complexity. This regularisation helps prevent overfitting by encouraging simpler models.

The regularisation term $\Omega(f_t)$ typically includes both L1 (Least Absolute Shrinkage and Selection Operator) and L2 (ridge) penalties, which are represented as:

$$\Omega(f_t) = \gamma T + \frac{1}{2}\alpha \sum_{j=1}^{T} |w_j| + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2, \tag{7}$$

where T is the number of leaf nodes in the regression tree, w_j are the weights associated with the leaf nodes, and γ, α, and λ are regularisation parameters. γ controls the complexity of the model by penalising the number of leaf nodes, thus encouraging simpler trees. α and λ are L1 and L2 regularisation parameters, respectively, used to penalise the magnitude and sum of the leaf weights, further constraining model complexity.

HGBM extends the foundational principles of XGBoost by introducing a more memory-efficient approach. In traditional gradient boosting methods like XGBoost, decision trees are built by selecting optimal split points for continuous features. These split points are used to determine how a tree branches at each node, thereby separating the data into groups that maximise the predictive accuracy of the model.

XGBoost evaluates all possible split points within a feature's range to find the threshold that minimises the loss function, a process that can become computationally expensive and memory-intensive, especially for high-dimensional datasets. HGBM addresses this limitation by constructing histograms for each continuous feature, where the feature values are divided into a predetermined number of bins. Instead of evaluating every possible value as a potential split point, HGBM only considers the boundaries of these bins. This discretisation significantly reduces the complexity of determining optimal split points, as the algorithm narrows down the number of candidate splits.

By restricting the search space to a smaller set of potential splits, HGBM reduces both the computational time and memory required for training, making it particularly advantageous for large-scale datasets. The memory efficiency of HGBM is further enhanced by storing bin indices instead of the raw continuous data. This compression reduces the memory usage, allowing the algorithm to handle larger datasets more efficiently, while maintaining high predictive accuracy.

CatBoost is a gradient-boosting algorithm that is capable of handling both categorical and numerical features without the need for manual feature encoding. To address the issue of target leakage, CatBoost employs an ordered boosting technique, as depicted in Fig. 4.

In ordered boosting, the training set is randomly sampled to generate n subsets, and prediction models $M_1, M_2, \ldots, M_{n-1}$ are sequentially trained on these subsets. For each model M_i, the first i samples from the selected subset are used, ensuring that the model only learns from prior data and does not incorporate information from the current or future observations. This approach fundamentally differs from traditional gradient boosting, where all residuals are computed using the same dataset, potentially leading to overfitting and target leakage.

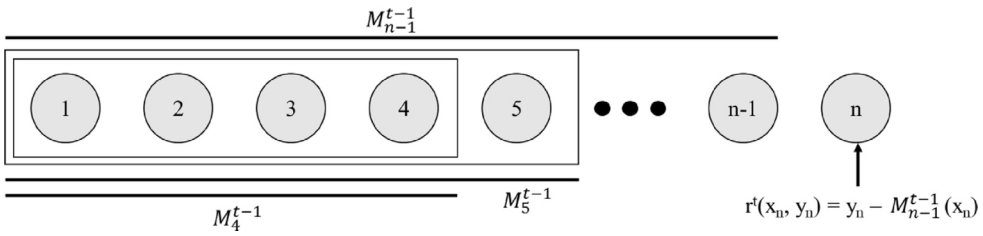


Fig. 4. Ordered boosting in CatBoost.

Table 2
Evaluation metrics for assessing the Performance of prediction models.

Metrics	Definition	Equation
RMSE	Root Mean Square Error	$\mathrm{RMSE} = \sqrt{\frac{1}{N_s}\sum_{k=1}^{N_s}(f_e(k) - f_t(k))^2}$
MAE	Mean Absolute Error	$\mathrm{MAE} = \frac{1}{N_s}\sum_{k=1}^{N_s}\lvert f_e(k) - f_t(k)\rvert$
MAPE	Mean Absolute Percentage Error	$\mathrm{MAPE} = \frac{1}{N_s}\sum_{t=1}^{N_s}\left\lvert\frac{f_e(k)-f_t(k)}{f_t(k)}\right\rvert$
R^2-value	Coefficient of Determination	$R^2 = 1 - \frac{\sum_{k=1}^{N_s}(f_e(k) - f_t(k))^2}{\sum_{k=1}^{N_s}(f_t(k) - \overline{f_t})^2}$

To prevent such issues, CatBoost calculates the residual $r_t(x_n, y_n)$ for the nth sample in the tth iteration using predictions from the previous model M_{n-1}. This ensures that the prediction for a given sample is not influenced by its actual target value during training, thus providing a robust method for avoiding overfitting and leakage.

A variety of evaluation criteria, as outlined in Table 2, are used to assess the accuracy of prediction models. In these measures, N_s indicates sample count. The index of each individual sample is indicated by k. $f_e(k)$ and $f_t(k)$ stands for the predicted mill throughput of the kth sample, and its actual value, respectively. Each utilised evaluation metric provides unique insights into model performance. RMSE represents the magnitude of prediction errors, giving more weight to larger errors, while MAE offers the average absolute difference between predicted and actual values, providing a straightforward measure of accuracy. MAPE expresses errors as a percentage of the actual values, which allows for easier comparison. R^2 indicates the proportion of variance in the data explained by the model, giving an indication of goodness of fit for models. Together, these metrics offer a comprehensive evaluation of the predictive accuracy and overall performance of models. These metrics enable an objective comparison of different models, allowing for a precise quantification of their predictive accuracy and performance. The application of these metrics provides a structured and quantitative framework for evaluating model performance, ensuring an accurate and unbiased assessment.

3.2. Multi-objective optimisation methods

Multi-objective optimisation involves the simultaneous optimisation of several objectives, which often exhibit conflicting characteristics. The complexity of such problems arises when the optimal solution for one objective is not aligned with the optimal solution of another, leading to inherent trade-offs between the objectives. In these cases, a single global solution does not exist; rather, the problem yields a set of trade-off solutions known as Pareto-optimal solutions. A solution is considered Pareto-optimal if no other solution can improve one objective without degrading the performance of at least one other objective.

The set of Pareto-optimal solutions forms what is referred to as the Pareto front, which provides decision-makers with a spectrum of possible trade-offs. Given the complexity and nature of these problems, evolutionary algorithms have proven to be highly effective for multi-objective optimisation. These algorithms employ a population-based search strategy, which allows the identification of a diverse set of solutions (Deb, 2011).

In this study, four well-established multi-objective evolutionary algorithms (MOEAs) are employed including Non-dominated Sorting Genetic Algorithm II (NSGA-II) (Deb et al., 2002), Multi-Objective Evolutionary Algorithm based on Decomposition (MOEA/D) (Zhang and Li, 2007), Strength Pareto Evolutionary Algorithm 2 (SPEA2) (Zitzler et al., 2001), and Reference Vector Guided Evolutionary Algorithm (RVEA) (Cheng et al., 2016).

NSGA-II (Deb et al., 2002) is a widely applied method for addressing multi-objective optimisation problems, particularly when objectives conflict. NSGA-II utilises a population-based evolutionary approach, where a set of candidate solutions evolves over successive generations, aiming to approximate the Pareto optimal front which represents the best possible trade-offs among the objectives.

The process starts by creating an initial set of candidate solutions, which is subsequently ranked based on dominance relationships. Solutions are divided into distinct Pareto fronts, with the first front comprising non-dominated solutions, meaning no other solution in the population outperforms them in all objectives. Subsequent fronts contain solutions that are dominated by those in the earlier fronts. To maintain diversity among the solutions and ensure a broad exploration of the solution space, NSGA-II incorporates a crowding distance mechanism. This metric favours solutions in less densely populated areas of the Pareto front during selection, helping to prevent premature convergence to a narrow set of solutions.

After sorting the population and computing crowding distances, parent solutions are chosen for reproduction based on their rank and crowding distance. These parents undergo genetic operations such as crossover and mutation to generate offspring, which are then combined with the parent population. A selection process follows, retaining the best-performing solutions to form the population for the next generation. This cycle of selection, variation, and ranking continues until the algorithm converges, yielding a diverse set of Pareto-optimal solutions that represent balanced trade-offs among the conflicting objectives.

MOEA/D (Zhang and Li, 2007) is another powerful multi-objective optimisation algorithm employed in this study. This method decomposes a multi-objective optimisation problem into several single-objective subproblems, which are then optimised simultaneously using an evolutionary algorithm. Each solution in MOEA/D corresponds to a specific subproblem, and these subproblems are organised based on a neighbourhood structure, determined by the similarity of their weight vectors. The population at each generation consists of the best solution for each subproblem. Given that neighbouring subproblems often have similar optimal solutions, MOEA/D optimises each subproblem by leveraging information from its neighbouring subproblems.

Through an iterative process involving genetic operators such as crossover, mutation, and selection, new candidate solutions are generated and evaluated for each subproblem. The best-performing solutions are retained, while inferior solutions are discarded. This process is repeated until the termination criterion is met. To ensure a diverse set of solutions along the Pareto front, MOEA/D incorporates diversity-preserving mechanisms, including crowding and niching strategies. These strategies prevent solutions from clustering in specific regions of

the Pareto front, promoting a more thorough exploration of the solution space.

SPEA2 (Zitzler et al., 2001) is another multi-objective optimisation algorithm employed in this study. The algorithm begins by creating a population of random solutions P_0, alongside an empty archive $\overline{P_0}$, with the archive used to store non-dominated solutions as the algorithm progresses. At each iteration t, SPEA2 evaluates the fitness of individuals in both the current population P_t and the archive $\overline{P_t}$ using a combination of three key criteria: strength, raw fitness, and density.

The strength of each individual is calculated based on the number of other solutions it dominates. This value quantifies how many other solutions in both the population and the archive are outperformed by that individual in terms of their objective values. Following this, the *raw fitness* of an individual is computed by summing the strengths of all the solutions that dominate it.

To ensure diversity across the Pareto front, SPEA2 incorporates a density metric, which is calculated by measuring the distance of each individual to its nearest neighbours. This density calculation ensures that solutions are well-distributed. In the next step, the fitness of each solution is calculated by summing its density and raw fitness.

Once the fitness values are computed, all non-dominated solutions from both the population and archive are copied to form the next generation's archive. If the archive exceeds its predefined size, a truncation operator is used to remove solutions. This operator prioritises the removal of individuals that are located closest to others in the objective space, maintaining diversity among the retained solutions. If the archive is smaller than the required size, dominated solutions are added based on their fitness, ensuring that the archive remains at the desired size.

After updating the archive, a new population is formed through the selection of parents from the archive and applying genetic operators such as crossover and mutation. These operators create new candidate solutions, which are evaluated in the next iteration. This process of fitness evaluation, archive updating, and genetic operations continues until a termination criterion, such as a maximum number of generations or adequate convergence, is met.

Similar to MOEA/D, RVEA utilises the decomposition approach to distribute the search across multiple subproblems. In RVEA, a set of uniformly distributed reference vectors is defined to partition the objective space into smaller subspaces. This decomposition enables the simultaneous optimisation of subproblems, similar to the approach employed by MOEA/D. However, RVEA distinguishes itself through the use of reference vector-guided selection.

RVEA also incorporates an elitism strategy, where the parent population is combined with the offspring population at each generation to undergo an elitist selection process. The combined population is partitioned into N subpopulations, with each solution associated with its nearest reference vector. Here, N represents the number of reference vectors, which determines the number of subspaces in the objective space. From each subpopulation, a single elitist solution is selected to form the next generation.

A new scalarisation method, entitled Angle Penalized Distance (APD), is introduced in RVEA, which dynamically balances convergence and diversity. The distance between candidate solutions and the ideal point is used as the convergence criterion in APD, while the acute angle between candidate solutions and the reference vectors serves as the diversity measure. The elite solution that best balances convergence and diversity during the optimisation process is selected from each subpopulation based on its lowest APD value (Trivedi et al., 2016).

4. Experimental results

In this section, the prediction results achieved by utilising three gradient boosting methods including XGBoost, CatBoost, and HGBM for the two objectives of mill throughput and circulating load are compared. The most accurate model for each objective is identified and subsequently employed as the surrogate model within four well-known multi-objective optimisation algorithms including NSGA-II, MOEA/D, RVEA, and SPEA2. The performance of these multi-objective EAs is then compared, and the best-performing algorithm is selected. This selected algorithm, incorporating the most accurate surrogate models, is able to suggest optimal parameter settings to maximise mill throughput while minimising circulating load. Subsequently, a sensitivity analysis is conducted to assess the robustness of the identified optimal solutions and to evaluate the impact of key factors on these solutions. Finally, recommendations for adjusting the set points are proposed. It is worth noting that for surrogate modelling and multi-objective optimisation, the scikit-learn (Pedregosa et al., 2011) and Pymoo (Blank and Deb, 2020) frameworks were used, respectively.

Table 3
Summary of evaluation metrics results for mill throughput using CatBoost, HGBM, and XGBoost models.

Model	Metrics	Min	Max	Mean	Median	STD
CatBoost	RMSE	30.5913	32.8839	31.8498	31.8672	0.5276
	MAE	22.8038	24.0535	23.2992	23.2587	0.3173
	MAPE	0.0181	0.0191	0.0185	0.0184	0.0003
	R^2	0.9262	0.9364	0.9304	0.9304	0.0025
HGBM	RMSE	34.8859	37.1086	35.9361	35.8770	0.5224
	MAE	25.6605	26.9914	26.2794	26.2325	0.3284
	MAPE	0.0204	0.0215	0.0209	0.0208	0.0003
	R^2	0.9060	0.9173	0.9114	0.9116	0.0028
XGBoost	RMSE	31.7769	34.0242	32.9295	32.7693	0.5721
	MAE	23.6079	24.5821	24.0088	23.8959	0.3012
	MAPE	0.0188	0.0195	0.0191	0.0190	0.0002
	R^2	0.9210	0.9314	0.9256	0.9263	0.0028

4.1. Prediction results

In this study, three machine learning models were utilised to predict mill throughput and circulating load expressed as:

$f_1(x_1, x_2, \ldots, x_{17})$ (Predicted Mill Throughput)

$f_2(x_1, x_2, \ldots, x_{17})$ (Predicted Circulating Load)

where the predictors x_i for $i = 1, \ldots, 17$ correspond to the input variables which are listed ones in Table 1, excluding target variables.

A 10-fold cross-validation was employed in this study to compare the predictive performance of the models. The results for the predictive performance of the employed models on the test data for mill throughput and circulating load, as measured by the evaluation metrics, are presented in Tables 3 and 4, respectively. Additionally, the box plots of the evaluation metrics for mill throughput and circulating load are depicted in Figs. 5 and 6. These box plots illustrate key summary statistics such as the median, quartiles, and potential outliers, enabling a quick understanding of the central tendency and variability of the prediction performance results.

Based on the obtained results, CatBoost demonstrated the best performance for mill throughput prediction, achieving the lowest average RMSE of 31.8498 and the highest average R^2 of 0.9304. Similarly, for circulating load prediction, CatBoost outperformed other models with the lowest average RMSE of 14.8256 and the highest average R^2 of 0.9031. Catboost results are closely followed by XGBoost results. For instance, the average R^2 values for XGBoost are 0.9256 for mill throughput and 0.8941 for circulating load. To establish a more rigorous foundation for identifying the best-performing model, the Friedman Test was performed, followed by paired T-tests. The key assumptions of the Friedman test were satisfied, as the same dataset was used to evaluate all three prediction models. The R^2 results for the prediction models are independent, and the R^2 values are continuous and can be ranked, with higher values indicating better predictive performance. Finally, as the Friedman test is non-parametric, it does not require the assumption of normality. The resulting average ranks and p-values are

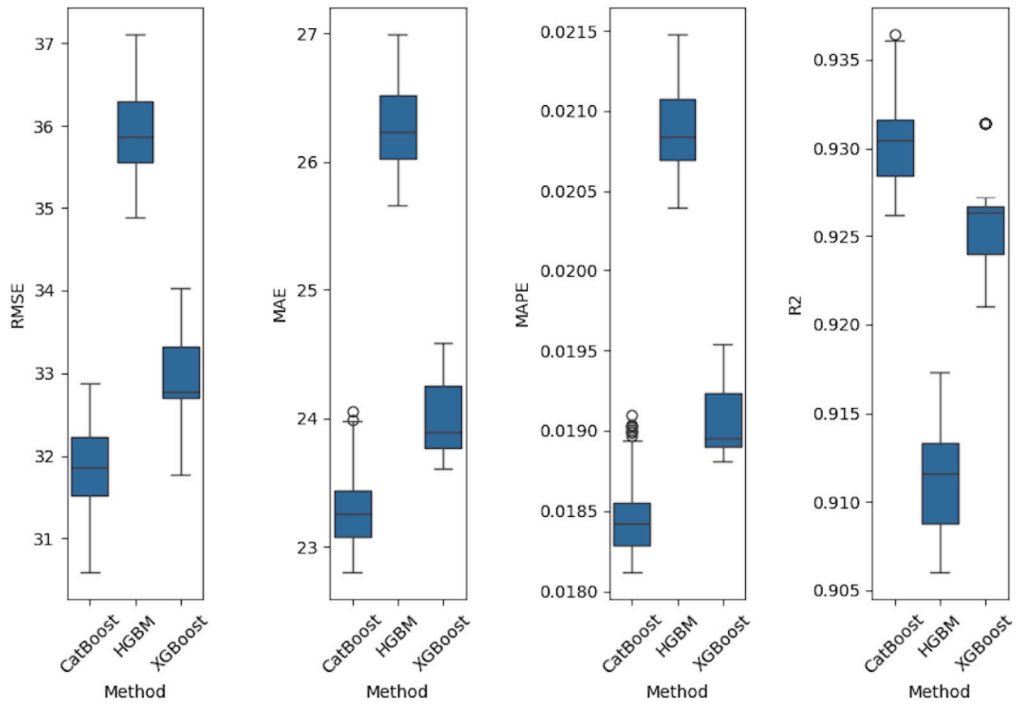


Fig. 5. Performance metrics for mill throughput prediction using CatBoost, HGBM, and XGBoost methods.

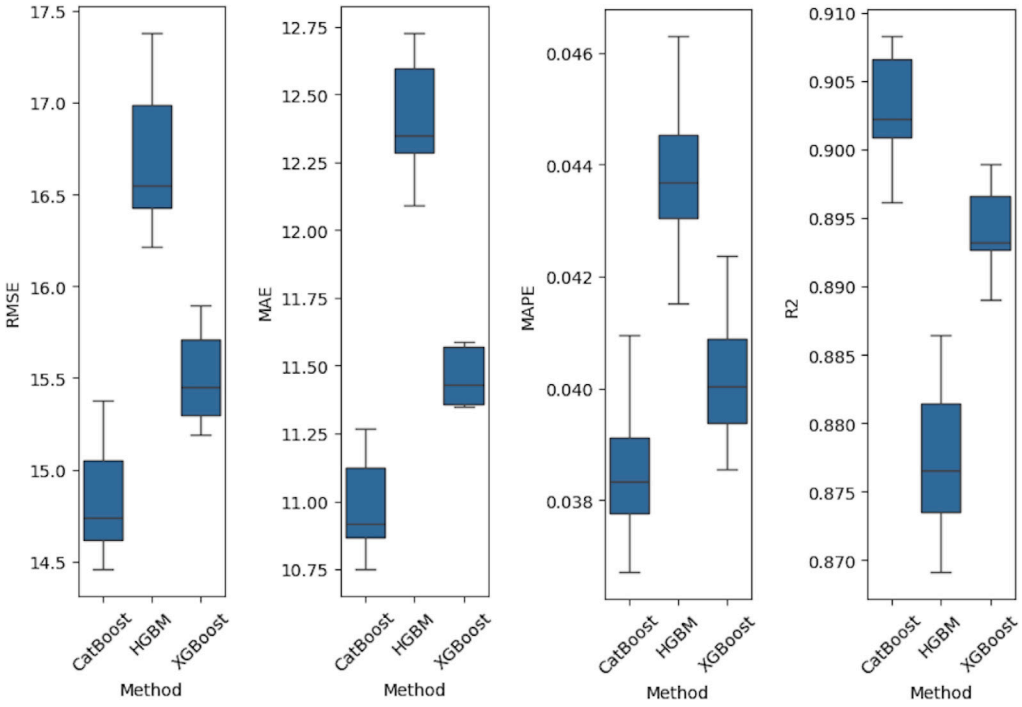


Fig. 6. Performance metrics for circulating load prediction using CatBoost, HGBM, and XGBoost methods.

presented in Table 5. Based on these results, CatBoost emerged as the most accurate model, with the lowest average rank and statistically significant differences from the other models, as indicated by p-values below the significance threshold ($\alpha = 0.05$). This confirms CatBoost's superiority in predicting both SAG mill throughput and circulating load.

This strong performance is largely due to CatBoost's implementation of several techniques specifically designed to prevent overfitting. One of the key features is the use of the innovative "Ordered Boosting" algorithm, which introduces randomness during the tree-building process, helping to reduce overfitting. This approach enhances the model's

Table 4
Summary of evaluation metrics results for circulating load using CatBoost, HGBM, and XGBoost models.

Model	Metrics	Min	Max	Mean	Median	STD
CatBoost	RMSE	14.4567	15.3742	14.8256	14.7358	0.2690
	MAE	10.7492	11.2689	10.9826	10.9171	0.1486
	MAPE	0.0367	0.0409	0.0385	0.0383	0.0011
	R^2	0.8961	0.9083	0.9031	0.9022	0.0034
HGBM	RMSE	16.2144	17.3810	16.6653	16.5490	0.3361
	MAE	12.0912	12.7280	12.4012	12.3482	0.1856
	MAPE	0.0415	0.0463	0.0437	0.0437	0.0013
	R^2	0.8691	0.8865	0.8776	0.8765	0.0049
XGBoost	RMSE	15.1932	15.8950	15.5001	15.4516	0.2218
	MAE	11.3481	11.5872	11.4573	11.4300	0.0922
	MAPE	0.0386	0.0424	0.0401	0.0400	0.0011
	R^2	0.8890	0.8989	0.8941	0.8932	0.0030

Table 5
Friedman test and T-test results (accuracy = R^2 values) for mill throughput and circulating load.

Method	Mill throughput		Circulating load	
	Average rank	P-value (T-test)	Average rank	P-value (T-test)
CatBoost	1.0	–	1.0	–
XGBoost	2.0	6.09E−71	2.0	1.08E−73
HGBM	3.0	3.44E−123	3.0	6.25E−110

ability to generalise to new, unseen data. As a result, CatBoost outperforms other ensemble methods, showing greater resilience against overfitting and delivering superior predictive accuracy. These CatBoost models for predicting mill throughput and circulating load will be used as surrogate models within optimisation evolutionary algorithms in the next step.

4.2. Optimisation results

This study utilised four well-established evolutionary multi-objective optimisation algorithms including NSGA-II, MOEA/D, RVEA, and SPEA2, with parameters as represented in Table 6. The CatBoost models for predicting mill throughput and circulating load, serve as the fitness functions across all algorithms. A population size of 100 and a maximum of 50 generations are employed for each algorithm, while all other parameters are maintained at their default configurations.

The objective is to simultaneously maximise mill throughput and minimise circulating load while adhering to the given operational constraints. The mathematical formulation of this optimisation problem is expressed as follows:

$$\max_{\mathbf{x}} \quad f_1(x_1, x_2, \ldots, x_{14}) \quad \text{Mill Throughput}$$
$$\min_{\mathbf{x}} \quad f_2(x_1, x_2, \ldots, x_{14}) \quad \text{Circulating Load}$$

Subject to:

$$\text{Lower Bound}_i \leq x_i \leq \text{Upper Bound}_i, \quad \forall i = 1, 2, \ldots, 14$$

and

$$\sum_{j=1}^{10} \%PL_j = 100$$

In this formulation parameter j represents each particle size-related feature, starting from j = 1 for %PL 13.2 mm to j = 10 for 212–300 mm. Among the 17 variables listed in Table 1, three variables including CV006, CV007, and CV008 are disturbances (non-controllable) and are therefore fixed at their median values to ensure the optimisation reflects typical operating conditions. The remaining 14 variables are included in the optimisation process, with the lower and upper bounds set to the minimum and maximum values of each feature, as presented in Table 1".

Table 6
Parameters of multi-objective optimisation algorithms.

Algorithm	Default parameters
NSGA-II	sampling = float random crossover = Simulated Binary Crossover (SBX) (eta = 15, prob = 0.9) mutation = Polynomial Mutation (PM) (eta = 20)
MOEA/D	sampling = float random crossover = DifferentialEvolutionCrossover (F = 0.5, CR = 1.0) mutation = Polynomial Mutation (PM) (eta = 20) prob_neighbours = 0.9
RVEA	sampling = float random crossover = Simulated Binary Crossover (SBX) (eta = 15, prob = 0.9) mutation = Polynomial Mutation (PM) (eta = 20) alpha = 2
SPEA2	sampling = float random crossover = Simulated Binary Crossover (SBX) (eta = 15, prob = 0.9) mutation = Polynomial Mutation (PM) (eta = 20)

The constraints ensure that the sum of the particle size distribution is 100%, and each variable remains within its defined operational limits.

As outlined in Section 3.2, multi-objective optimisation methods do not yield a single optimal solution. Instead, they produce a set of Pareto-optimal solutions, each representing different trade-offs between the objectives. Figs. 7 to 10 illustrates the distribution of objectives across various generations for a single run, along with the Pareto fronts generated by the applied multi-objective optimisation algorithms at selected generations.

As observed in these figures, NSGA-II and SPEA2 exhibit higher variation in performance across generations, while RVEA shows moderate variation, and MOEA/D displays the least variation. For all four methods, as expected, mill throughput consistently increases while circulating load decreases as the optimisation progresses. The Pareto fronts shift towards higher mill throughput and lower circulating load across successive generations, indicating that the optimisation process is progressing as expected. MOEA/D demonstrates the fastest convergence, with minimal changes in objective values beyond generation 40. Moreover, the final Pareto front generated by MOEA/D shows a narrower spread of solutions compared to the other algorithms. In terms of extreme objective values achieved, NSGA-II produces the highest mill throughput in the final generation, while SPEA2 achieves the lowest circulating load.

Due to the stochastic nature of these algorithms, slight variations in results can occur across multiple runs. To mitigate this variability, each algorithm was executed ten times. Hypervolume (HV) (Guerreiro et al., 2021) is selected as the performance metric for comparing these algorithms. The hypervolume indicator is a widely used performance metric in multi-objective optimisation, designed to assess the quality of the Pareto front solution set. It quantifies the volume of the objective space that is dominated by the set of solutions and bounded by a defined reference point. To compute the hypervolume indicator, the objective space dominated by each solution is first determined relative to the reference point. The individual contributions are then aggregated to obtain the total hypervolume of the solution set. A larger hypervolume indicates a more favourable solution set.

Fig. 11 presents a box plot comparison of hypervolume values for the four multi-objective optimisation algorithms used in this study: MOEA/D, NSGA-II, SPEA2, and RVEA. Each box plot illustrates the distribution of hypervolume values obtained over 10 independent runs per algorithm, offering a clear view of their overall performance.

Based on these results, it is evident that the MOEA/D algorithm exhibits the lowest median hypervolume and the largest variability, as indicated by its wide interquartile range (IQR). This suggests that

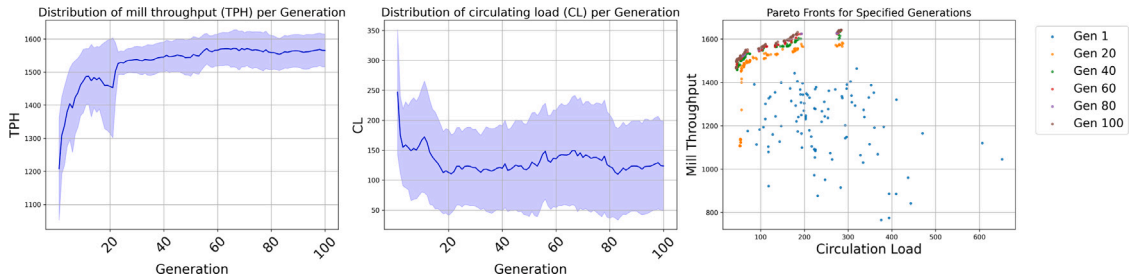


Fig. 7. Fitness distribution of TPH, CL, and Pareto fronts for selected generations using NSGAII algorithm.

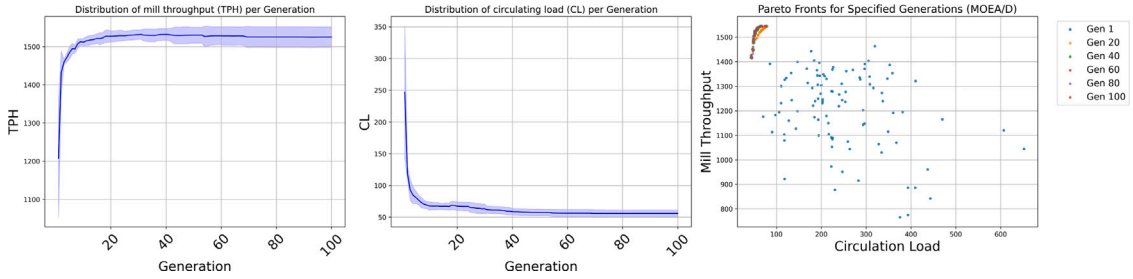


Fig. 8. Fitness distribution of TPH, CL, and Pareto fronts for selected generations using MOEA/D algorithm.

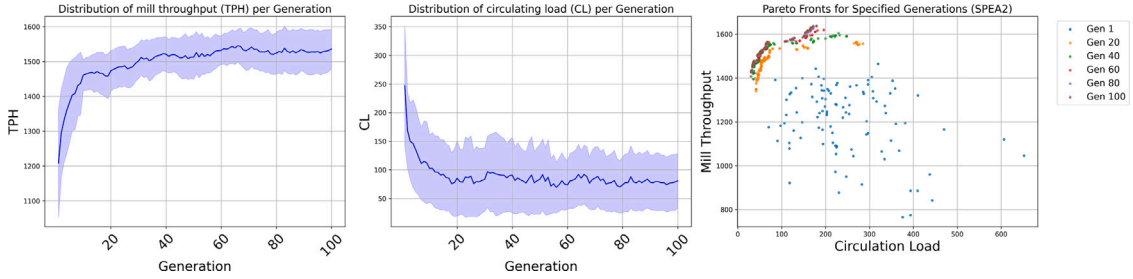


Fig. 9. Fitness distribution of TPH, CL, and Pareto fronts for selected generations using SPEA2 algorithm.

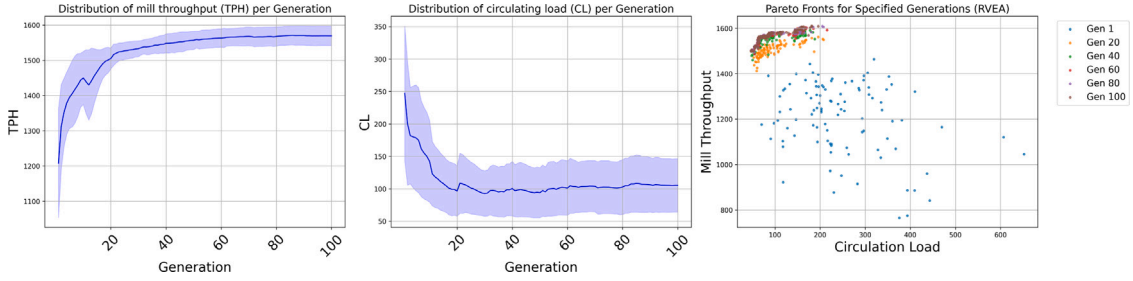


Fig. 10. Fitness distribution of TPH, CL, and Pareto fronts for selected generations using RVEA algorithm.

MOEA/D's performance is less consistent across different runs compared to the other methods. In contrast, NSGA-II and SPEA2 demonstrate higher median hypervolume values and narrower IQRs, indicating better overall performance and reduced variability, making them more reliable in consistently achieving favourable results. The RVEA algorithm has a relatively lower median compared to NSGA-II and SPEA2, but it outperforms MOEA/D and shows moderate variability, with an outlier reflecting occasional poorer performance.

To provide a more rigorous comparison, Table 7 reports the results of the Friedman test, which ranks the four multi-objective optimisation algorithms based on their hypervolume values. The average ranks indicate the relative performance of each algorithm, with a lower rank signifying better performance. Additionally, pairwise comparisons were carried out using the T-test, with a significance level of $\alpha = 0.05$, to assess whether the observed differences between the methods are statistically significant.

The results show that NSGA-II obtains the lowest average rank (1.6), suggesting it performs the best overall in terms of hypervolume. SPEA2 follows closely with an average rank of 1.7, and the T-test reveals that the difference between NSGA-II and SPEA2 is not statistically significant ($p = 0.806354$). RVEA ranks third, with a statistically significant difference compared to NSGA-II ($p = 0.008011$). Finally, MOEA/D has the highest average rank (3.9), and the T-test indicates it is significantly outperformed by all other algorithms, particularly when compared to NSGA-II ($p = 0.000022$). Based on the obtained results, NSGA-II and SPEA2 demonstrate the best overall performance in terms of hypervolume, with NSGA-II performing slightly better. MOEA/D, on the other hand, is the least effective method. Therefore, NSGA-II is

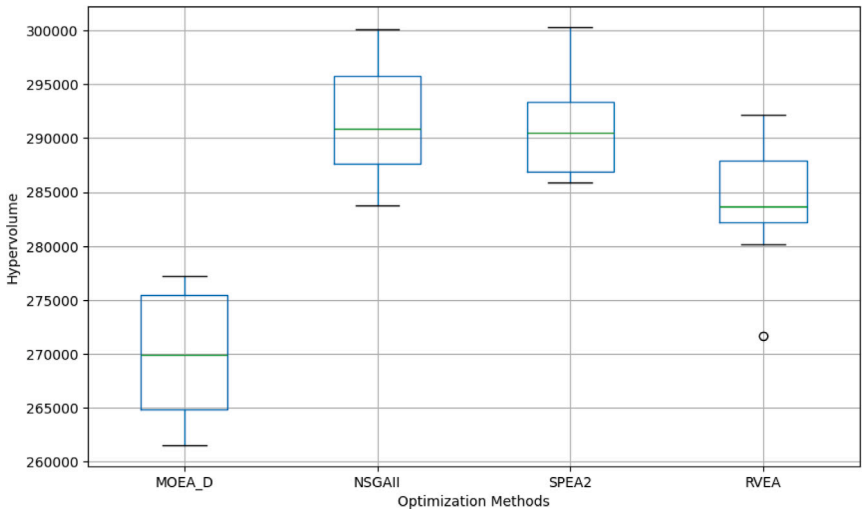


Fig. 11. Comparison of hypervolume values across different optimisation methods (MOEA/D, NSGA-II, SPEA2, and RVEA) over 10 independent runs.

Table 7
Friedman test and T-test results for hypervolume comparison across optimisation methods.

Method	Average rank (Friedman test)	P-value (T-test, $\alpha = 0.05$)
NSGA-II	1.6	–
SPEA2	1.7	0.806354
RVEA	2.8	0.008011
MOEA/D	3.9	0.000022

selected as the best-performing multi-objective optimisation method for maximising mill throughput and minimising circulating load.

4.3. Sensitivity analysis and optimum set points

NSGA-II, as the best-identified optimisation method, embedding CatBoost models as surrogate models for predicting mill throughput and circulating load, is able to propose optimal set points for controllable variables that maximise mill throughput and minimise circulating load.

An important consideration is the robustness of the identified optimal set points. This section conducts a sensitivity analysis by systematically mutating each feature within its allowable range to observe how the Pareto-optimal solutions respond. This provides insights into the robustness of the optimal solutions and the impact of changes in different input features. In this analysis, Pareto-optimal solutions from the final generation of the best-performing NSGA-II algorithm, selected from 10 independent runs, are analysed. Each input feature is systematically perturbed with a uniform distribution of values between the minimum and maximum bounds, while the other features are held constant. For each Pareto-optimal solution, ten perturbed variations were generated. The resulting changes in two objectives, mill throughput and circulating load, are then plotted to visualise the sensitivity of each Pareto-optimal solution to changes in each input feature. The plots depict the relationship between the original Pareto-optimal solutions and their perturbed counterparts, represented by circles. These points are connected by dotted lines of the same colour, with each colour corresponding to a specific Pareto-optimal solution (see Fig. 12).

The sensitivity analysis results for selected input features are presented in Fig. 12. As can be seen, there is a diverse range of impacts across different features, highlighting the complex interplay between input variables and the mill's performance. Some features exhibit significant sensitivity, acting as critical control points where even minor adjustments can lead to substantial changes in mill performance. In contrast, other features show less pronounced effects. Notably, certain features primarily influence one objective; for instance, water changes predominantly affect mill throughput rather than circulating load. In contrast, features like mill speed have a greater impact on circulating load. The fresh ratio impacts both objectives. The sensitivity of the objectives to changes also varies; for example, while altering some particle size distribution features, such as the 150–212 mm, shows minimal effect, other features induce more substantial variations. Regarding particle size distribution features, the sensitivity of optimal solutions is more pronounced for finer particles, such as the 13.2 mm fraction, compared to larger particles like the 150–212 mm fraction. The impact on the two objectives differs; for instance, changes in the 13 mm fraction predominantly affect mill throughput, whereas other features, such as the 37–53 mm fraction, influence both mill throughput and circulating load.

A crucial observation from the sensitivity analysis is that the mutated solutions generally do not improve both objectives simultaneously. Specifically, no mutated solution achieves both a higher mill throughput and a lower circulating load compared to its corresponding original Pareto-optimal solution, meaning that the original solutions are not dominated. This lack of dominance indicates that the original solutions are already well-optimised with respect to the trade-offs between circulating load and mill throughput. The mutated solutions either maintain or degrade the performance metrics compared to the original Pareto front. The fact that the mutated solutions could not improve over the original solutions highlights the robustness of the original Pareto front. It suggests that the process has been optimised effectively within the given constraints. The inability of the mutated solutions to dominate or improve upon the original solutions suggests that the optimisation process has already identified a strong set of solutions.

Fig. 13 illustrates the Pareto-optimal solutions from the final generation of NSGA-II across 10 independent runs as a parallel coordinate plot. The first 14 vertical axes represent the optimised input features, and the last two axes show the predicted circulating load and mill throughput. All of the solutions are represented as lines joining points on the parallel axes, where each point represents a variable's value in that solution.

Analysis of the top-performing solutions provides recommendations for increasing mill throughput and minimising circulating load. Some of these recommendations include maintaining an inlet water flow rate of 480–508 m^3/h and a fresh feed ratio above 98%. The optimal range for turning speed is between 9.4 and 9.9 RPM, with the most optimal points concentrated in the narrower range of 9.4 to 9.6 RPM. For particle size distribution, the recommended proportions are: 19–26.5 mm at 3–4.5%, 53–75 mm at 8–10.5%, 75–106 mm at 15–19.5%, and

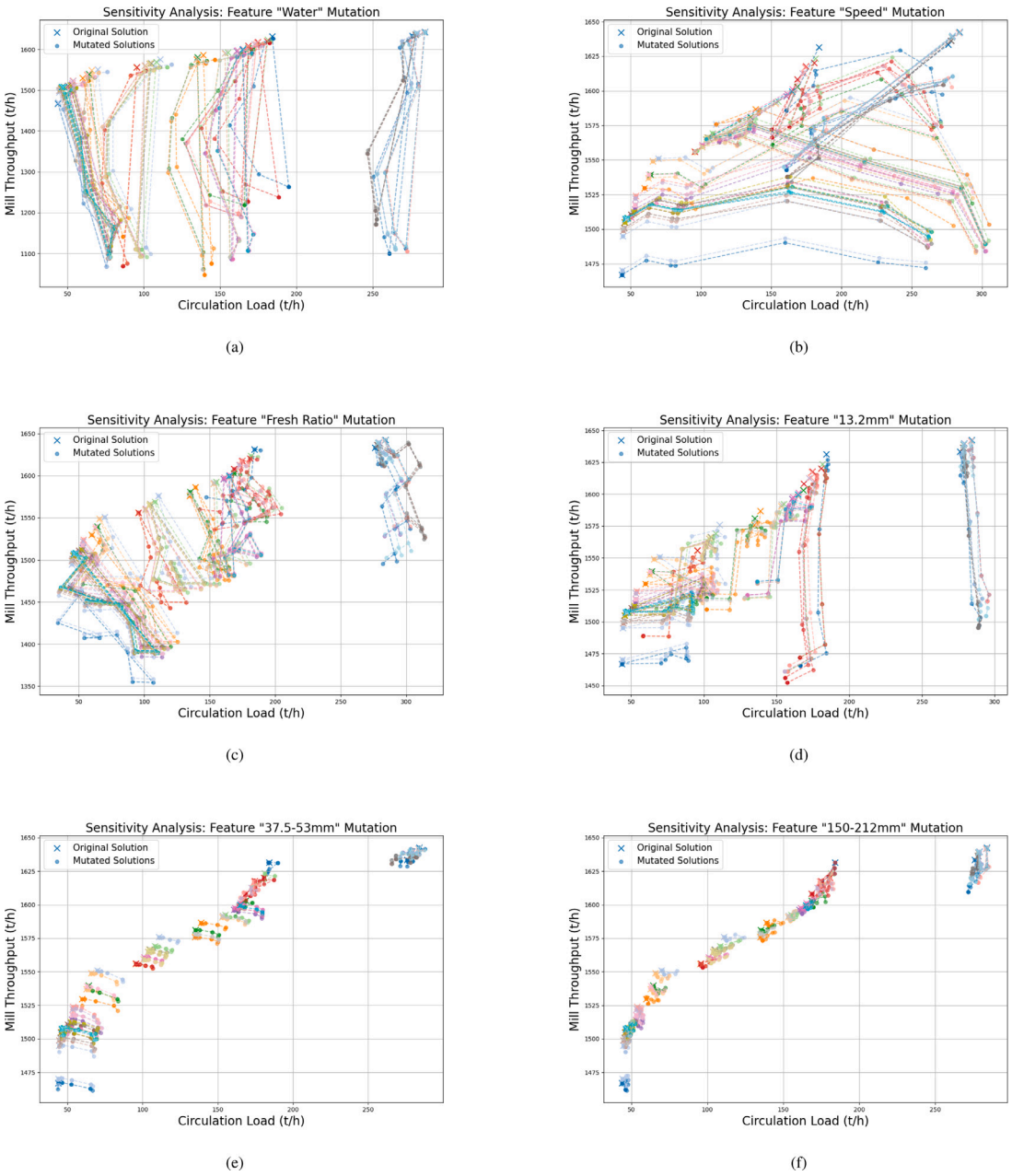


Fig. 12. Sensitivity analysis results for: (a) Water, (b) Speed, (c) Fresh Ratio, (d) %PL 13.2 mm, (e) %PL 37.5–53 mm, and (f) %PL 150–212 mm.

106–150 mm at 18.5–23.5%. Implementing these adjustments is expected to increase mill throughput to over 1500 t/h and reduce the circulating load to below 190 t/h, representing a substantial improvement over the initial values shown in Table 1 and Fig. 2. In this study, circulating load is reported in tonnes per hour (t/h). However, it is important to note that circulating load can be reported as a percentage of mill throughput. While the two representations are related, the same circulating load percentage can correspond to different t/h values depending on the throughput level. Presenting circulating load as a percentage may offer more intuitive insights for operators, particularly when comparing different suggested operating set points.

It is worth mentioning that in this research, all input variables, except for CV006, CV007, and CV008, are considered controllable, allowing the model to search for their optimal values in order to explore their potential impact on the objectives. However, setting some of these parameters exactly as identified by the optimisation algorithm and driving the circuit to the best-found configuration may not be straightforward in practice. For example, the current method for adjusting particle size distribution is through modifying feed ratios from stockpiles, which is difficult to set precisely. However, this study demonstrates how changes in input features, even those that are currently difficult to control, can influence the objectives and contribute to achieving higher mill throughput and lower circulating load. Based on the operational needs of a specific mining complex, it is also possible in future work to redefine the controllability of certain variables, for example, considering some of the currently controllable inputs as uncontrollable, and re-run the models accordingly to derive adjusted set points. Furthermore, when applying these optimally found set points in practical operations, there might be some practical considerations to take into account. For example, adjusting the recommended inlet water flow to achieve the required percent solids in the mill. These optimal results will be proposed to operational experts, who will select and apply the most appropriate set points based on current operational requirements. These experts will also leverage their extensive experience and knowledge of the process to do sanity checks. The expert

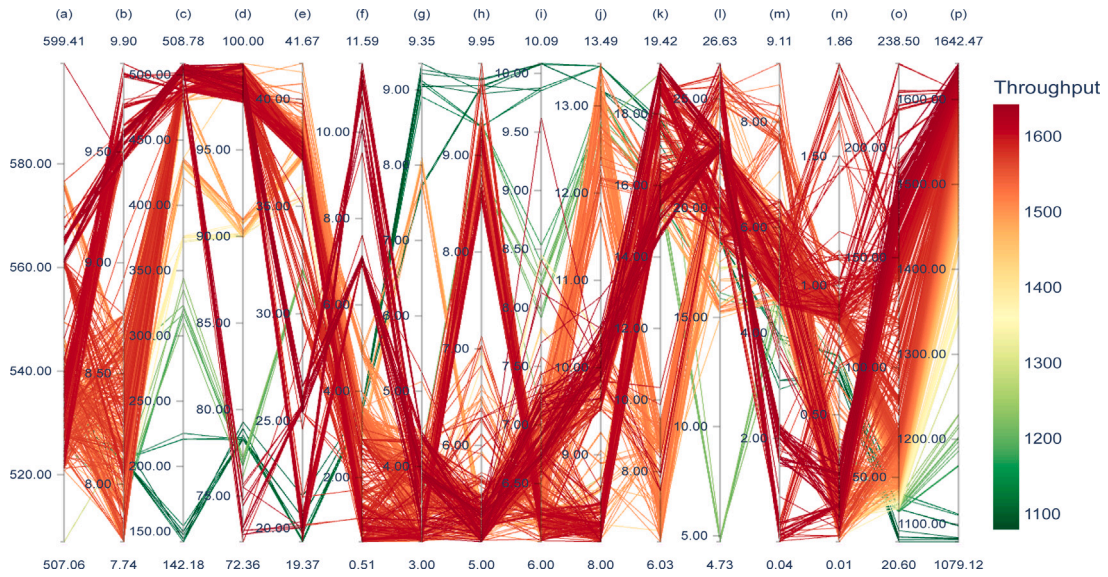


Fig. 13. Parallel plot illustrating Pareto optimal solutions generated by NSGAII across 10 independent runs: (a) Weight, (b) Speed, (c) Water, (d) Fresh Ratio, (e) 13.2 mm, (f) 13.2–9 mm, (g) 19–26.5 mm, (h) 26.5–37.5 mm, (i) 37.5–3 mm, (j) 53–5 mm, (k) 75–06 mm, (l) 106–50 mm, (m) 150–12 mm, (n) 212–00 mm, (o) Circulation load, (p) Mill throughput.

review process will involve checking the consistency of the optimal solutions with expected real-world behaviour, taking into account logical relationships between input variables and objectives.

The advantage of multi-objective optimisation in this case is that it does not provide a single optimal solution, but rather a set of Pareto-optimal solutions. This enables experts to choose the most suitable setting based on real operational conditions.

5. Conclusions

This research addresses the optimal parameter setting for the grinding step in mineral processing plants with the simultaneous objectives of maximising mill throughput and minimising circulating load. An integrated framework comprising machine learning models and evolutionary algorithms is developed for this purpose. Three gradient boosting methods are utilised to model mill throughput and circulating load. CatBoost outperformed the other models for both objectives, achieving an average R^2 of 0.93 for mill throughput and 0.90 for circulating load.

For the multi-objective optimisation, the performance of four algorithms including NSGA-II, MOEA/D, RVEA, and SPEA2 are compared. Both NSGA-II and SPEA2 represented superior results, with NSGA-II showing a marginally higher hypervolume. Sensitivity analysis of the proposed optimal parameter setting by NSGA-II, demonstrated the impact of varying input features on the optimal solutions and also validated their robustness.

This integrated framework, which combines CatBoost models and NSGA-II effectively provides set points that maximise mill throughput while minimising circulating load. The best-performing solutions are analysed, and suggestions for process parameter setting are provided. A key advantage of the multi-objective optimisation framework developed in this study is that it generates a set of Pareto-optimal solutions, enabling experts to select the best settings that are more appropriate for real-time operational conditions.

For future work, incorporating more process parameters, ore properties, and especially equipment-related factors such as liner and grate conditions could enhance the overall predictive performance. Additionally, performing hyperparameter tuning on both the prediction and optimisation models could further refine their performance. Investigating hybrid model approaches that combine the unique strengths of each method also holds promising potential for achieving even better outcomes.

CRediT authorship contribution statement

Zahra Ghasemi: Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Mehdi Neshat:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Chris Aldrich:** Writing – review & editing, Supervision, Methodology, Investigation. **Max Zanin:** Writing – review & editing, Supervision. **Lei Chen:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by the Australian Research Council Integrated Operations for Complex Resources Industrial Transformation Training Centre (project number IC190100017) and funded by universities, industry and the Australian Government. The authors would also like to extend their gratitude to Mike Daniel for his valuable support and contributions to this work.

Data availability

The data that has been used is confidential.

References

Alruiz, O., Morrell, S., Suazo, C., Naranjo, A., 2009. A novel approach to the geometallurgical modelling of the collahuasi grinding circuit. Miner. Eng. 22 (12), 1060–1067.

Avalos, S., Kracht, W., Ortiz, J.M., 2020. Machine learning and deep learning methods in mining operations: A data-driven SAG mill energy consumption prediction application. Min. Met. Explor. 37 (4), 1197–1212.

Behnamfard, A., Roudi, D.N., Veglio, F., 2020. The performance improvement of a full-scale autogenous mill by setting the feed ore properties. J. Clean. Prod. 271, 122554.

Blank, J., Deb, K., 2020. Pymoo: Multi-objective optimization in python. Ieee Access 8, 89497–89509, URL https://pymoo.org.

Bond, F.C., 1961. Crushing and grinding calculations, part I. Br. Chem. Eng. 6, 378–385.
Both, C., Dimitrakopoulos, R., 2021. Applied machine learning for geometallurgical throughput prediction—A case study using production data at the Tropicana Gold mining complex. Minerals 11 (11), 1257.
Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.
Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining. pp. 785–794.
Cheng, R., Jin, Y., Olhofer, M., Sendhoff, B., 2016. A reference vector guided evolutionary algorithm for many-objective optimization. IEEE Trans. Evol. Comput. 20 (5), 773–791.
Curilem, M., Acuña, G., Cubillos, F., Vyhmeister, E., 2011. Neural networks and support vector machine models applied to energy consumption optimization in semiautogeneous grinding. Chem. Eng. Trans. 25, 761–766.
Deb, K., 2011. Multi-objective optimisation using evolutionary algorithms: an introduction. In: Multi-Objective Evolutionary Optimisation for Product Design and Manufacturing. Springer, pp. 3–34.
Deb, K., Pratap, A., Agarwal, S., Meyarivan, T., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. 6 (2), 182–197.
Dietterich, T.G., 2000. Ensemble methods in machine learning. In: International Workshop on Multiple Classifier Systems. pp. 1–15.
Flores, L., Limitada, S.G.-M.E., 2005. Hardness model and reconciliation of throughput models to plant results at minera escondida ltda., Chile. Technical Bulletin 5.
Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. System Sci. 55 (1), 119–139.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
Fuerstenau, M.C., Han, K.N., 2003. Principles of Mineral Processing. SME.
Ghasemi, Z., Neshat, M., Aldrich, C., Karageorgos, J., Zanin, M., Neumann, F., Chen, L., 2024a. An integrated intelligent framework for maximising SAG mill throughput: Incorporating expert knowledge, machine learning and evolutionary algorithms for parameter optimisation. Miner. Eng. 212, 108733.
Ghasemi, Z., Neumann, F., Zanin, M., Karageorgos, J., Chen, L., 2024b. A comparative study of prediction methods for semi-autogenous grinding mill throughput. Miner. Eng. 205, 108458.
Guerreiro, A.P., Fonseca, C.M., Paquete, L., 2021. The hypervolume indicator: Computational problems and algorithms. ACM Comput. Surv. 54 (6), 1–42.
Hart, S., Valery, W., Clements, B., Reed, M., Song, M., Dunne, R., 2001. Optimisation of the cadia hill SAG mill circuit. In: SAG Conference.
Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2, Springer.
Hoseinian, F.S., Faradonbeh, R.S., Abdollahzadeh, A., Rezai, B., Soltani-Mohammadi, S., 2017. Semi-autogenous mill power model development using gene expression programming. Powder Technol. 308, 61–69.
Katoch, S., Chauhan, S.S., Kumar, V., 2021. A review on genetic algorithm: past, present, and future. Multimedia Tools Appl. 80, 8091–8126.
Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y., 2017. Lightgbm: A highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30.
Morrell, S., 2004. Predicting the specific energy of autogenous and semi-autogenous mills from small diameter drill core samples. Miner. Eng. 17 (3), 447–451.
Ou, T., Liu, J., Liu, F., Chen, W., Qin, J., 2023. Coupling of XGBoost ensemble methods and discrete element modelling in predicting autogenous grinding mill throughput. Powder Technol. 422, 118480.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
Powell, M., Van der Westhuizen, A., Mainza, A., 2009. Applying grindcurves to mill operation and optimisation. Miner. Eng. 22 (7–8), 625–632.
Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. Adv. Neural Inf. Process. Syst. 31.
Ruiz, M.A.V., Gonzales, J.A.V., Villalba, F.J.B., 2024. Multivariable predictive models for the estimation of power consumption (kW) of a semi-autogenous mill applying machine learning algorithms [modelos predictivos multivariables para la estimación de consumo de potencia (kW) de un molino semi-autógeno aplicando algoritmos de machine learning]. J. Energy Environ. Sci. 8 (1), 14–31.
Rybinski, E., Ghersi, J., Davila, F., Linares, J., Valery, W., Jankovic, A., Valle, R., Dikmen, S., 2011. Optimisation and continuous improvement of antamina comminution circuit. In: SAG Conference. p. 19.
Trivedi, A., Srinivasan, D., Sanyal, K., Ghosh, A., 2016. A survey of multiobjective evolutionary algorithms based on decomposition. IEEE Trans. Evol. Comput. 21 (3), 440–462.
Wills, B.A., Finch, J., 2015. Wills' Mineral Processing Technology: an Introduction to the Practical Aspects of Ore Treatment and Mineral Recovery. Butterworth-Heinemann.
Wills, B.A., Napier-Munn, T., 2006. An introduction to the practical aspects of ore treatment and mineral recovery. Wills' Miner. Process. Technol. 267–352.
Wolpert, D.H., 1992. Stacked generalization. Neural Netw. 5 (2), 241–259.
Zhang, Q., Li, H., 2007. MOEA/D: a multiobjective evolutionary algorithm based on decomposition. IEEE Trans. Evol. Comput. 11 (6), 712–731.
Zitzler, E., Laumanns, M., Thiele, L., 2001. SPEA2: Improving the strength Pareto evolutionary algorithm. TIK Rep. 103.