Artificial Intelligence Applications And Innovations Vol. 2 No. 1 (2025)

Article history:
Received 8 September 2025
Revised 31 December 2025
Accepted 9 September 2026
Published online 1 April 2025

A Novel AI-Driven Framework for Predicting Fragmentation and Cost-Oriented Design Optimization Using a Hybrid XGBoost–NSGA-II Approach

Hassan. Hosseinzadeh[1*], Gholamreza. Nowrouzi[1], Elham. Rafieenia[2]

1. Department of Mining Engineering, Faculty of Engineering, University of Birjand, Birjand, Iran
2. Department of Mining Engineering, Faculty of Engineering, University of Sistan & Baluchestan, Sistan & Baluchestan, Iran

*** Corresponding author email address:** hosseinzadeh.hassan70@gmail.com

Article Info

Article type:
Original Research

How to cite this article:
Hosseinzadeh, H., Nowrouzi, G., & Rafieenia, E. (2025). A Novel AI-Driven Framework for Predicting Fragmentation and Cost-Oriented Design Optimization Using a Hybrid XGBoost–NSGA-II Approach. *Artificial Intelligence Applications and Innovations, 2*(2), 85-99.
https://doi.org/10.61838/jaiai.2.2.6

ABSTRACT

Modeling and optimizing the blasting process in mining has always been one of the fundamental challenges in mining engineering due to its multidimensional and nonlinear nature. This study aimed to develop a hybrid framework based on artificial intelligence for predicting rock fragmentation and simultaneously optimizing blasting cost and quality. In the first stage, the XGBoost machine learning algorithm was employed to model and predict the mean fragmentation size (P50) based on blasting design parameters. The results indicated that the proposed model could accurately capture complex relationships among variables, achieving R^2 values of 0.97 for the training dataset and 0.92 for the testing dataset, thus demonstrating remarkable predictive performance. Variable importance analysis revealed that the specific charge (q) and the burden distance (B were the most influential factors controlling fragmentation, while the rock mass quality (T) also played a decisive role in altering the fracture mechanism. In the next step, the NSGA-II evolutionary algorithm was applied for multi-objective optimization between blasting costs and the P50 index. The optimization outcomes generated a set of Pareto solutions, allowing engineers to flexibly select blasting patterns in accordance with either economic or operational priorities. The findings demonstrate that the proposed hybrid framework not only enhances prediction accuracy compared to empirical models but also provides a powerful data-driven decision-making tool for blast design. By introducing an integrated approach based on XGBoost and NSGA-II, this study makes a significant contribution to the blasting engineering literature and paves the way for developing sustainable and efficient blast designs in various mining operations.

Keywords: *XGBoost, Blast Fragmentation, Multi-objective Optimization, NSGA-II algorithm, Data-driven Modeling.*

1. Introduction

Rock blasting in both surface and underground mines, as the most common and cost-effective method for primary rock fragmentation, has consistently played a central role in the mining value chain. The quality of fragmentation achieved through blasting not only affects the efficiency of subsequent operations such as loading, hauling, and secondary crushing but also directly influences the overall extraction costs[1]. Quantitative fragmentation indices, particularly the mean particle size known as P_{50}, are among the key metrics for evaluating blasting efficiency, as this index is directly related to the performance of loading machinery, material handling systems, and the energy consumption required for secondary crushing. Therefore, precise control and accurate prediction of parameters such as P_{50} have always been considered essential objectives in blast design [2].

Despite the importance of this issue, modeling and predicting rock fragmentation remains a major challenge for mining engineers and researchers due to the complex, multidimensional, and nonlinear nature of blasting. Blasting design parameters—such as hole diameter, burden distance (B), spacing between rows (S), stemming length, sub-drilling depth, and specific charge (q)—each exert nonlinear and intricate effects on the fracture pattern and fragmentation distribution. Furthermore, geological conditions including rock strength, jointing degree, regional tectonics, and groundwater presence make blasting outcomes highly unpredictable. As a result, traditional models based on empirical or semi-empirical relationships, although useful as initial frameworks for understanding the phenomenon, fail to deliver accurate and generalizable predictions under diverse geological and operational conditions [3-5].

In recent decades, the emergence of data-driven approaches and artificial intelligence has opened new horizons for modeling complex mining phenomena. Among machine learning algorithms, XGBoost, a gradient boosting method based on decision trees, has gained widespread applications in various engineering domains due to its strong capability in modeling nonlinear relationships, preventing overfitting, and achieving outstanding predictive accuracy. Applying such algorithms for predicting blasting fragmentation indices enables overcoming the limitations of empirical models and achieving higher reliability and precision [6-10].

However, prediction of fragmentation represents only part of the blast design problem. In practice, mining engineers require approaches that, in addition to predicting fragmentation quality, also allow for simultaneous optimization of multiple objectives. Among the most critical of these objectives are minimizing blasting costs (including explosive consumption, drilling operations, and hole preparation) and achieving desirable fragmentation. This inherently constitutes a multi-objective optimization problem, since there exists a trade-off between fragmentation quality and costs. For example, increasing the specific charge (q) can improve fragmentation but simultaneously increases explosive costs and may also raise additional costs due to ground vibrations or fly-rock. In such circumstances, multi-objective evolutionary algorithms such as NSGA-II are powerful tools for solving optimization problems in blasting engineering, thanks to their ability to explore complex design spaces and provide a Pareto front of optimal solutions [11, 12].

Combining an accurate predictive model based on machine learning (for predicting P_{50}) with a multi-objective evolutionary algorithm (for optimizing cost and fragmentation) can thus provide a comprehensive and innovative framework for blast design that ensures both scientific accuracy and industrial applicability.

Accordingly, the present study aimed to develop an AI-driven framework for predicting P_{50} and optimizing blast design with respect to cost. Within this framework, the XGBoost algorithm was first used to establish a predictive model for P_{50} based on blasting design parameters. Then, by defining a comprehensive cost function that incorporates the economic aspects of blasting operations, the NSGA-II algorithm was employed to simultaneously optimize cost and fragmentation quality. The results, while offering novel scientific insights into the complex relationships between blasting variables and fragmentation, show that the proposed hybrid framework can serve as an effective tool for achieving blasting patterns with reduced costs and improved fragmentation quality under real mining conditions. The main novelty of this research lies in presenting an integrated, AI-based framework that simultaneously enables accurate prediction of fragmentation (P_{50}) and cost-oriented blast design optimization. Unlike most previous studies that either focused solely on developing empirical or data-driven models for fragmentation prediction or examined cost optimization in isolation, this research integrates the predictive power of XGBoost in capturing nonlinear relationships among blasting parameters and P_{50} with the capability of NSGA-II in handling multi-objective optimization. Defining a comprehensive cost function tailored to real-world blasting operations and embedding it within the optimization process ensures that the results of this study possess not only scientific significance but also high practical value. Therefore, this research bridges the domains of data-driven modeling and multi-objective optimization in blasting engineering, opening new horizons for cost-effective and efficient blast design in mining operations.

2. Literature Review

Classical fragmentation prediction models—particularly the Kuz–Ram family, which combines the Kuznetsov equation, the Rosin–Rammler distribution, and Cunningham's uniformity index—have long served as the dominant framework for blast design. However, their sensitivity to the so-called rock factor and the reduction of complex geomechanical and operational realities into a few aggregate parameters have limited their accuracy and generalizability in real-world scenarios. Recent studies, while revisiting and improving individual components of the model (e.g., data-driven estimation of the rock factor or modified size distribution functions), consistently emphasize that the nonlinear and multifactorial nature of fragmentation cannot be fully captured by static empirical relationships [13].

Over the past decade, data-driven and machine learning approaches have been increasingly applied to predict fragmentation indices such as P_{50} and P80. Comparative results suggest that XGBoost often outperforms classical regression methods and even certain neural networks in modeling nonlinear relationships between blasting parameters and fragmentation outcomes. Moreover, hybrid frameworks—integrating ML models with metaheuristic optimizers for hyperparameter tuning—have produced significant improvements in prediction accuracy. This trend has been corroborated by recent review papers and peer-reviewed studies, where Bayesian and evolutionary optimizers have been widely adopted to enhance the performance of RF, ANN, SVR, and XGBoost models [14-16].

In parallel, a separate body of research has focused on the optimization of drilling–blasting costs without direct integration with fragmentation prediction models. These studies typically formulate the problem mathematically—using integer or stochastic programming—or employ metaheuristic algorithms such as Firefly, PSO, and GA to minimize explosive, drilling, and operational costs. However, fragmentation indices are either oversimplified or treated through out-of-loop empirical models. Even the most recent examples, despite innovations in risk modeling or operational constraints, have largely emphasized cost reduction, while incorporating machine learning-based P_{50} prediction within multi-objective optimization remains limited [17].

To address the inherent cost–fragmentation quality trade-off, the use of multi-objective algorithms—particularly NSGA-II (Non-dominated Sorting Genetic Algorithm II)—has become increasingly common in mining. Nevertheless, the dominant focus in this literature has been on problems such as ore blending, production scheduling, or operational planning, rather than blast design explicitly linked to predicted fragmentation. These works demonstrate the ability of NSGA-II to manage multi-objective trade-offs through Pareto fronts and solution selection mechanisms, yet they generally lack the component of machine learning-driven P_{50} prediction within the optimization loop [18-20].

More recently, several attempts have been reported to bridge prediction and optimization in blasting. These include studies applying LSSVM or ANN models in conjunction with metaheuristics to predict fragmentation and then optimize blasting parameters, as well as hybrid investigations that jointly address cost and fragmentation indices such as X_{50} or P_{50}. Nevertheless, most of these studies either rely on predictive models other than XGBoost, or adopt multi-objective frameworks outside the NSGA-II family, or define cost functions and operational constraints in a limited or fragmented manner. This gap highlights the innovation of the present framework: leveraging XGBoost for accurate prediction of P_{50} and integrating it with NSGA-II for simultaneous optimization of blasting cost and fragmentation-related parameters, supported by a comprehensive cost function and realistic operational constraints [2, 21, 22].

3. Data and Methods

3.1. Data

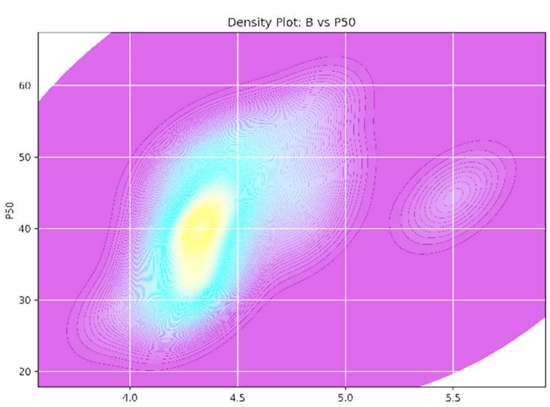
This study was conducted based on a field dataset comprising 37 records from blasting operations in an open-pit iron ore mine located in Hunan Province, China. Each record represents a real blast design pattern in which multiple input parameters and a key output variable were documented. The input parameters included row spacing (S), burden distance (B), specific charge (q), sub-drilling depth (H), and rock type (T). Row spacing (S) refers to the horizontal distance between adjacent blastholes in the same row, which plays a decisive role in energy distribution and fragmentation uniformity. Burden (B), defined as the distance between the first blasthole and the nearest free face, directly influences the throw direction of rock after blasting. The specific charge (q), expressed as the ratio of explosive mass to the blasted rock volume, serves as a key indicator of energy utilization efficiency. Sub-drilling (H) represents the extra drilling depth beyond the designed bench height, applied to avoid the presence of unblasted toe material. In addition, rock type (T) was considered as a categorical variable, classified into two groups—intact rock and jointed rock—which significantly affect energy transmission and, consequently, the fragmentation pattern [2]. Table 1 presents the descriptive statistics of the parameters used in this study.

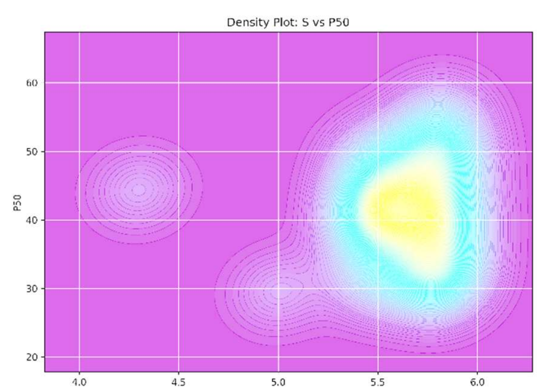
Table 1: Descriptive Statistics of the Parameters Used in the Study

Variable	Count / Frequency	Mean / Average	Median	Standard Deviation	Min	Q1	Q3	Max
S	36	5.533	5.8	0.368	4.3	5.5	5.8	5.8
B	36	4.408	4.3	0.305	4	4.3	4.5	5.5
q	36	0.431	0.4	0.054	0.32	0.39	0.47	0.54
H	36	1.756	1.5	0.598	0.5	1.5	2	2.5
P_{50}	36	42.98	40.21	7.28	29.54	35.05	48.06	55.78

The output variable was the mean particle size, or P_{50}, representing the median dimension of the fragmentation size distribution after blasting. P_{50} is widely recognized as a standard indicator for assessing blasting efficiency since it reflects fragmentation quality and is directly linked to material loading and hauling performance. Hence, P_{50} holds particular importance in the optimization of mining operations [23]. The range of recorded values for the input parameters covers a wide variety of real blasting conditions, enabling analysis of interactions between design variables and rock characteristics. The combination of quantitative and qualitative variables within this dataset also provides a suitable basis for evaluating the capability of machine learning algorithms in modeling complex phenomena and highlights the potential of data-driven approaches for reducing reliance on empirical judgments and improving the precision of blast design.

Analysis of the density distribution of blasting design parameters—B, S, q, and H—in relation to fragmentation index (P_{50}) reveals that the behavior of each parameter reflects the interplay of geomechanical, geometric, and operational factors during blasting (Figure 1).





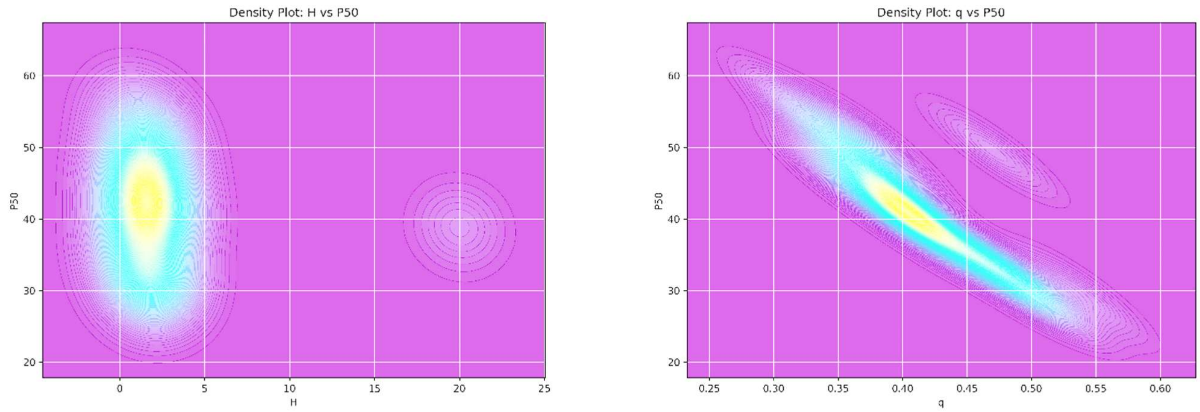


Figure 1: Density Plot of the Parameters Used in the Study vs. P_{50}

For B, the clustering of data in the 4.2–4.6 m range indicates efforts to achieve stability and uniformity in blast design. However, observed values extending up to 7.5 m reflect local adjustments in response to varying lithological conditions or bench geometries. While such deviations allow adaptation to geological heterogeneity, at a larger scale they increase the dispersion of explosive energy and reduce fragmentation homogeneity, which ultimately decreases the efficiency of loading and hauling operations. The distribution of S shows a similar pattern. Concentration of data between 3.5 and 5.8 m indicates a prevailing design, while the overall variation range (3.9–6.1 m) clearly reflects the flexibility of blasthole patterns in dealing with geomechanical heterogeneity. Although such flexibility is necessary, excessive variability can lead to fluctuations in energy distribution and reduce the predictability of fragmentation outcomes. For q, a relatively consistent and interpretable relationship with P_{50} is observed. An increase in q leads to a reduction in P_{50}, showing a clear negative correlation between the two. Concentration of data in the range of q = 0.38–0.42 kg/m³ with corresponding P_{50} values of 35–40 mm suggests conditions close to optimal, where explosive energy is effectively utilized for rock fragmentation and operational efficiency is maximized. This underscores the role of q as one of the most critical and sensitive parameters in blast design. By contrast, H exhibits the greatest variability. Although most values are less than 5 m, cases exceeding 20 m indicate abnormal conditions that may result from operational errors, variations in rock strength and hardness, or equipment limitations. Such extreme variations not only impose additional costs but also lead to undesired rock damage, reduced controllability of fragmentation, and increased uncertainty in blast outcomes.

In summary, the behavior of blasting design parameters is not governed solely by numerical standards but rather by the complex interplay of geomechanical conditions, operational constraints, and engineering judgment. This inherent variability and uncertainty indicate that traditional design approaches alone are insufficient for achieving the required precision and predictability. Therefore, data-driven models and machine learning algorithms are indispensable, as they can analyze multiple variables and nonlinear patterns simultaneously, ensuring optimized blast designs and improved efficiency under real mining conditions.

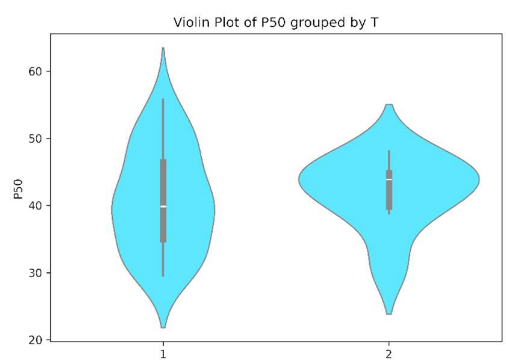


Figure 2: Distribution of Fragmentation Index (P_{50}) with Respect to Rock Mass Condition (T)

Figure 2 illustrates the distribution of fragmentation index P_{50} according to rock mass conditions. The results reveal that jointed rock masses (T = 1), due to structural heterogeneity and numerous fractures and joints, exhibit substantial variability in P_{50} values. The range of this index spans from approximately 30 to 55 mm, reflecting the nonlinear and inconsistent response of fractured rock to blast wave propagation and attenuation. The wide spread of the violin plot in this case indicates that explosive energy is absorbed or dissipated along multiple pathways, producing fragments of heterogeneous sizes with high variability. The clustering of data into several distinct groups further confirms that the response of fractured rock is mainly governed by pre-existing discontinuities. In contrast, intact rock masses (T = 2) display more consistent and predictable mechanical behavior. The data are concentrated within a narrower range of 40–48 mm, with the violin plot exhibiting a compact and uniform shape. This high concentration suggests that the coherence and continuity of intact rock facilitate relatively uniform energy distribution, and induced fractures largely follow the blast design. As a result, fragmentation in intact rock is less affected by random factors, producing more stable and homogeneous size distributions. Accordingly, it can be concluded that rock mass heterogeneity is the most critical factor contributing to uncertainty in fragmentation outcomes. In fractured rock, networks of discontinuities with varying strengths cause energy dispersion and consequently produce highly variable fragmentation, whereas in intact rock, geomechanically uniformity allows for greater controllability and predictability in blast design. These findings underscore the necessity of accurately evaluating rock mass conditions and incorporating geomechanically indices into the blast design process, particularly in heterogeneous and fractured environments, to achieve desirable fragmentation and reduce uncertainty.

3.2. Theoretical Framework of Methodology

The methodological framework of this study is built upon the synergistic integration of advanced machine learning and multi-objective evolutionary optimization. In this approach, the XGBoost algorithm serves as the predictive engine for the fragmentation index (P_{50}), while NSGA-II is employed to achieve an optimal balance among conflicting blast design parameters. The aim of this integration is to simultaneously enhance prediction accuracy and provide practical decision support for blast design optimization in complex mining environments.

3.2.1. XGBoost Machine Learning Algorithm

Extreme Gradient Boosting (XGBoost) is one of the most advanced tree-based machine learning frameworks, specifically designed for computational efficiency and scalability. XGBoost operates on the principle of gradient boosting, where a series of shallow and weak decision trees—each with limited standalone predictive ability—are combined sequentially to build a powerful and accurate model. In this iterative process, each new tree is trained to correct the residual errors of the preceding models, thereby improving accuracy step by step and enabling the model to capture increasingly complex patterns within the data. One of the key innovations of XGBoost compared to similar algorithms is its use of regularization mechanisms, which control model complexity and mitigate overfitting. This feature is particularly critical in noisy or highly variable datasets—such as blasting data in mining—because it ensures not only high accuracy in training but also reliable generalization to unseen data. Additionally, XGBoost incorporates computational optimizations and parallelization, making it both fast and scalable, thus suitable for analyzing large-scale industrial and research datasets.

In this study, XGBoost is chosen as the primary predictive model to capture nonlinear and complex relationships between blast design parameters—such as burden (B), spacing (S), specific charge (q), and subdrilling (H)—and the fragmentation index (P_{50}). The algorithm's gradual learning process, ability to regulate model complexity, and robustness against noise make it superior to traditional methods and many alternative algorithms, ensuring accurate and reliable outputs that can be effectively integrated into blasting optimization and engineering decision-making [24, 25].

3.2.2. NSGA-II Multi-objective Optimization Algorithm

Although XGBoost provides accurate fragmentation prediction, blast design inherently involves trade-offs among multiple conflicting objectives. For instance, improving fragmentation quality (reducing P_{50}) may increase explosive consumption, whereas lowering specific charge (q) may compromise fragmentation uniformity. To address such trade-offs, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) is employed as the optimization layer.

The theoretical foundation of NSGA-II lies in the principle of Pareto optimality, which asserts that a solution is Pareto-optimal if no objective can be improved without deteriorating at least one other. NSGA-II efficiently generates the Pareto front using evolutionary strategies and key mechanisms such as:

1. Fast Non-dominated Sorting to rank solutions based on Pareto dominance,
2. Crowding Distance to preserve population diversity and ensure proper distribution along the Pareto front,
3. Elitism to retain the best solutions across successive generations.

Through these mechanisms, NSGA-II efficiently constructs the Pareto front and provides decision-makers with a set of non-dominated (optimal) blast design alternatives [11, 12, 20, 26].

The integration of XGBoost and NSGA-II forms a prediction–optimization framework. First, XGBoost learns the predictive relationship between blast design parameters and P_{50}. Then, NSGA-II uses this predictive model as a fitness function to explore the design space, searching for solutions that minimize P_{50} while simultaneously controlling operational costs such as drilling, explosive consumption, and labor. Theoretically, this approach represents a data-driven optimization paradigm: empirical blasting data are transformed into predictive knowledge (via XGBoost), which is then operationalized for optimization (via NSGA-II). Practically, this framework aims to reduce uncertainty, improve blasting productivity, and enhance resource efficiency in mining operations.

3.3. *Cost Function*

In order to develop a comprehensive multi-objective optimization framework, this study incorporates not only the fragmentation index (P_{50}) but also the blasting operation cost (C) as the second optimization objective (Equation 1). The cost function is designed to encompass all major components influencing blasting costs, including drilling costs, explosive consumption, detonators, and labor for drilling and loading. Its mathematical structure is expressed as follows:

$$C = \frac{L_m.(H + H_0)}{S.B.H.\rho_r} + \frac{q.E_m}{\rho_r} + \frac{N_D.D_m}{S.B.H.\rho_r} + R_1 + R_2 \tag{1}$$

Among the variables, L_m denotes the unit cost of drilling, ρ_r represents the ore density, Em indicates the unit price of explosives, N_D is the average number of electronic detonators consumed per blast hole, D_m represents the unit price of detonators, R_1 denotes drilling labor costs, and R_2 represents loading labor costs. Constant parameters are excluded from numerical detailing here, as they serve only auxiliary roles and do not affect the variability within the scope of this study. Analytically, the first term of the equation accounts for drilling costs as a function of hole diameter and depth. The second term represents explosive costs proportional to specific charge and rock density. The third term incorporates detonator costs, while the final term reflects labor expenses for drilling and loading. Collectively, this cost model provides a comprehensive quantitative representation of blasting expenses, enabling simultaneous technical–economic analysis. Thus, within the multi-objective optimization framework, this cost function is paired with P_{50} as the dual basis for optimization.

3.4. *Model Evaluation Metrics*

The predictive performance of the XGBoost model is evaluated using three widely recognized statistical indicators: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2) [27].

MAE measures the average magnitude of prediction errors, regardless of their direction:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

This reflects the overall accuracy of the model by quantifying the average deviation between observed values (y_i) and predicted values ($\hat{y}_i$).

RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

Unlike MAE, RMSE penalizes larger errors more severely due to the squared term, making it a sensitive indicator for assessing model stability and its ability to capture large variations in the data.

R^2 indicates the proportion of variance explained by the model relative to the total variance of observations:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \qquad (4)$$

Values close to 1 represent high explanatory power and strong correlation between actual and predicted values, whereas values near zero or negative indicate weak generalization capability.

In summary, MAE captures average absolute error, RMSE highlights the influence of larger deviations, and R^2 quantifies explanatory power. Together, these metrics provide a robust and comprehensive assessment of the predictive performance of XGBoost [27].

4. Modeling

4.1. Data Preparation and Preprocessing

Unlike distance-based algorithms such as SVM or kNN, tree-based algorithms like XGBoost do not strictly require normalization or standardization, since their learning process is based on recursive binary splits and threshold optimization. Consequently, variable scale does not directly influence tree construction [24]. Nevertheless, practical applications recommend rescaling to prevent numerical instability, particularly in the presence of features with widely varying ranges, and to facilitate interpretability of feature importance.

For this purpose, Min–Max linear normalization was applied in this study, scaling all variables to the standard range [0,1] [28]:

$$\acute{x}_i = \frac{x_i - min\,(x)}{max(x) - min\,(x)} \qquad (5)$$

where x_i is the original value and x'_i is the normalized value.

Following preprocessing, the dataset was randomly divided into training (80%) and testing (20%) subsets. This partitioning ensures sufficient data for model learning while allowing fair evaluation of predictive performance on unseen data, thereby enhancing the model's generalization capability to real-world conditions.

4.2. Model implementation

To implement machine learning algorithms and perform the modeling process, Google Colab was employed as a cloud-based computational platform with parallel processing capabilities, along with the Python programming language due to its flexibility and the availability of powerful libraries in the field of machine learning and data mining.

In the first stage, the rock fragmentation index P_{50} was predicted using the XGBoost model, recognized as one of the most robust and efficient gradient boosting algorithms. To obtain the optimal hyperparameters of this model, the K-Fold Cross Validation method was applied, which not only improved accuracy but also prevented overfitting. The optimal values are reported in Table 2.

Table 2: Optimized Parameters of the XGBoost Algorithm

Parameters	Optimized value
n_estimators	169
max_depth	5
learning_rate	0.368
subsample	0.715
colsample_bytree	0.805
min_split_loss	3.814
gamma	0.858

After training the model with the training dataset, its performance was evaluated on the test data. For this purpose, comparative plots were generated between the actual and predicted values of P_{50} in both the training and test sets. The results demonstrated that the XGBoost model effectively captured the complex relationships between blasting parameters and the fragmentation index P_{50}. In the training data, the close alignment of points with the one-to-one correlation line, along with favorable statistical indicators (R^2 = 0.97, MAE = 0.86, RMSE = 1.17), confirmed the model's high accuracy and strong ability to extract nonlinear patterns embedded in the dataset. In the testing data, although a relative decline in performance was observed (R^2 = 0.92, MAE = 1.57, RMSE = 1.89), the model still explained more than 90% of the variance, reflecting substantial consistency between actual and predicted values. The uniform distribution of residuals without clustering in any specific range further indicated that the model-maintained stability across the entire data domain. Overall, these results (Table 3, Figures 3 and 4) demonstrate that the developed model not only reproduces the training data with high accuracy but also generalizes effectively to independent datasets.

Table 3: Evaluation Metrics of the XGBoost Model for Training and Testing Data

Evaluation criteria	Train data	Test data
R^2	0.97	0.92
RMSE	1.17	1.89
MAE	0.86	1.57

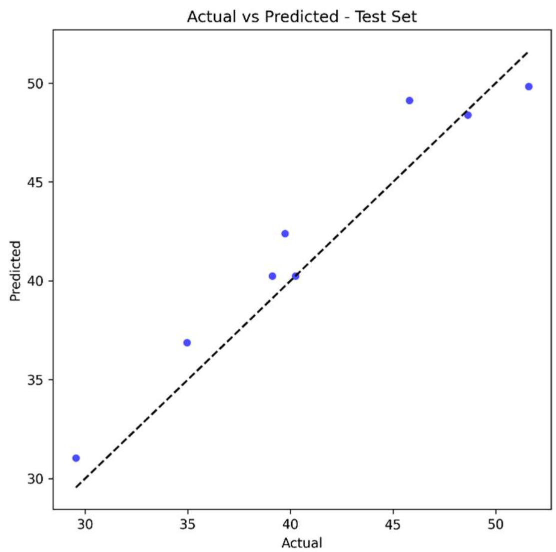


Figure 3: Scatter Plot of Actual vs. Predicted Values for the Test Data

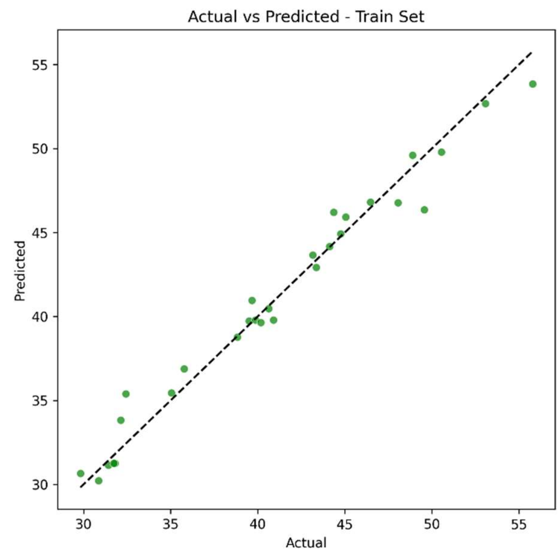


Figure 4: Scatter Plot of Actual vs. Predicted Values for the Train Data

To investigate prediction uncertainty, error distributions and residual plots were constructed for the test data (Figures 5 and 6). The residual plot derived from the XGBoost model precisely illustrates the distribution of prediction errors for the test samples. As shown, the scatter points span a horizontal range of predicted values (approximately 31–50) and a vertical range of residuals (approximately –3 to 2).

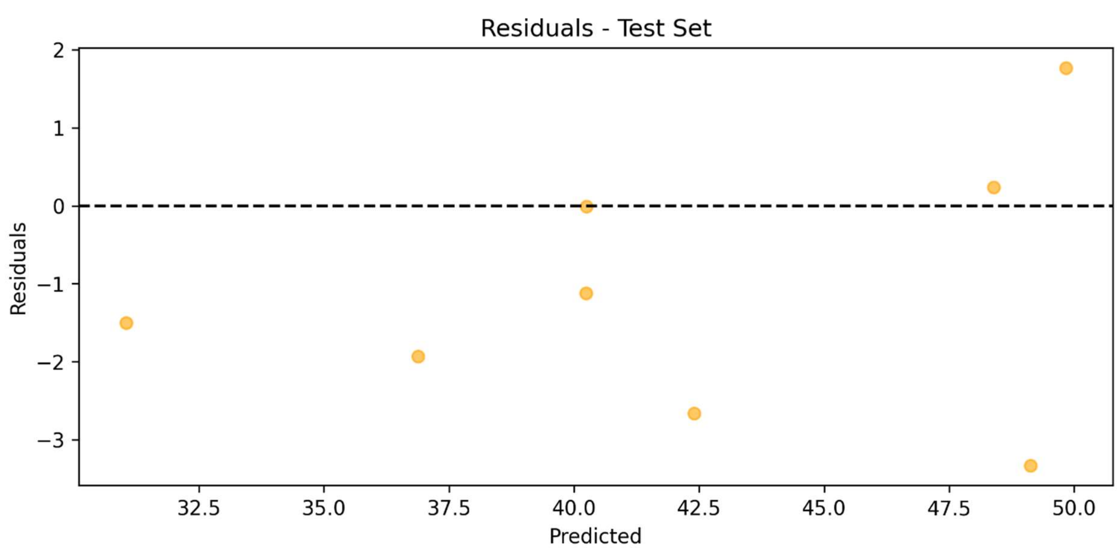


Figure 5: Residual Plot from the XGBoost Model for the Test Data

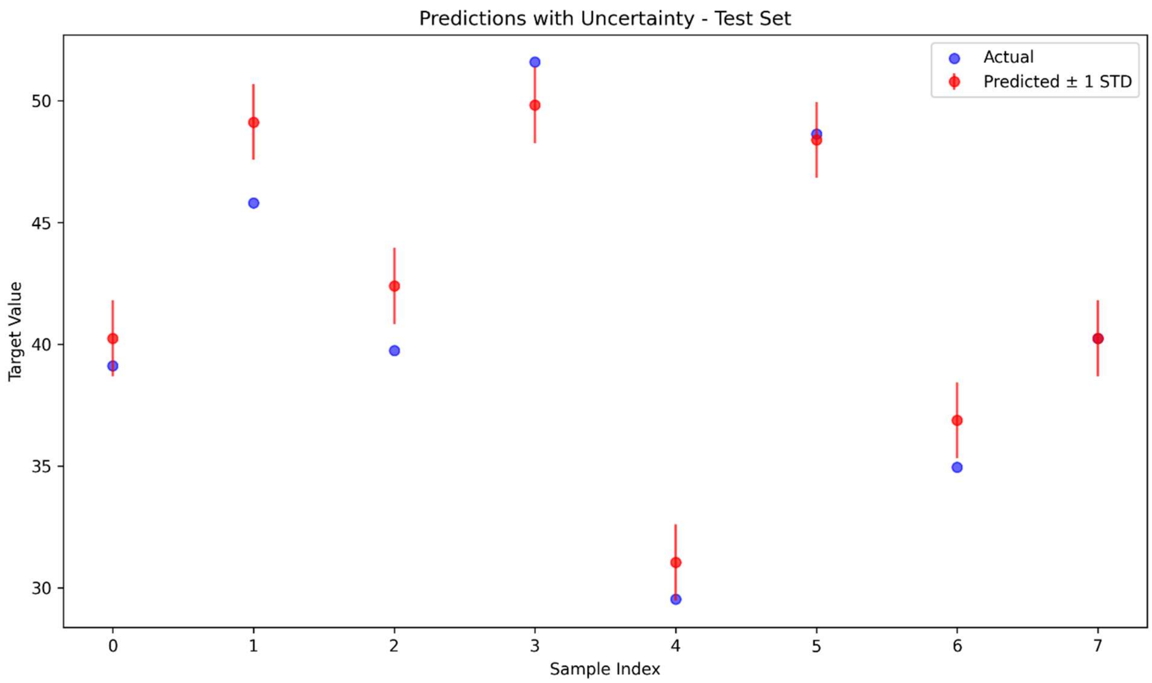


Figure 6: Uncertainty Analysis Plot for the Test Data

From an analytical perspective, the absence of any distinct pattern in the distribution of points indicates desirable model performance. The residuals are randomly scattered around the zero line (represented by a virtual horizontal axis), suggesting that the model does not exhibit significant systematic errors across different predicted values. This property is highly favorable, as it demonstrates the model's capability for generalization to unseen test data. A closer examination of the plot reveals that the vast majority of residuals are concentrated within ±1.5 units, reflecting acceptable predictive accuracy. However, the presence of a few outliers with larger negative residuals (around –3) indicates that the model faced greater difficulty in predicting some specific samples. This could be attributed to the unique characteristics of these samples or heterogeneity in the training data. From a practical standpoint, this residual distribution is reassuring, as it shows that model errors are uniformly distributed across the prediction range and that the model does not display noticeable bias in any specific interval. Such behavior is crucial for engineering applications, as it increases confidence in the reliability of the model's predictions under varying conditions.

The uncertainty analysis plot (Figure 6) provides a detailed comparison between actual P_{50} values (blue points) and predicted values from the XGBoost model with uncertainty intervals (red points with error bars) for the test dataset. As observed, in the majority of cases, the actual target values fall within ± one standard deviation of the predictions. This confirms that the model not only provides accurate forecasts but also delivers a conservative estimate of its potential error range. Analytically, the widths of the uncertainty intervals are not uniform across all samples. In some cases, the intervals are narrow, indicating high confidence in those predictions. Conversely, for a few samples, the intervals are wider, reflecting greater uncertainty regarding prediction accuracy. This behavior illustrates the model's adaptability to heterogeneity in the dataset and its sensitivity to the inherent uncertainties of geological and blasting processes. Such an ability to reflect confidence levels adds significant value for practical decision-making. Specifically, in scenarios where the cost of error is high, decision-makers can adopt more conservative strategies based on wider uncertainty ranges. On the other hand, where narrower ranges are observed, predictions can be relied upon with greater confidence. Overall, the results confirm that the XGBoost model not only delivers high predictive accuracy but also excels in estimating the magnitude of its own errors. This makes it a powerful tool for industrial applications and multidimensional decision-making in blasting and rock fragmentation engineering.

Finally, the relative importance of input variables in the XGBoost model was evaluated using the Feature Importance metric (Table 4). In this study, the Gain criterion was applied, which quantifies the average improvement in the loss function when a feature is used to split a decision tree node. In other words, Gain indicates the relative contribution of each feature to reducing prediction error, and features with higher Gain values play a more influential role in enhancing model performance. The distribution of Gain values among input variables thus reveals the hierarchy of parameter importance and allows for the identification of the most critical factors in predicting fragmentation index P_{50}.

The XGBoost analysis showed that specific charge (q), with a contribution of 47%, was the most influential parameter in predicting rock fragmentation P_{50}, as it directly controls the amount of explosive energy applied per unit volume of rock, thereby influencing rock breakage intensity and particle size distribution. Burden (B), with a 34% contribution, was the second most significant parameter, as it defines the blast geometry and the transmission of explosive energy into the rock mass. Together, q and B accounted for more than 80% of the model's predictive power. Rock type (T), with 14%, played a secondary role, while sub-drilling (H) and spacing (S), with only 5%, were the least influential due to their limited operational variation in the training dataset.

Operational results indicate that precise control of q and B is the key to accurate prediction of P_{50}, and the XGBoost model is capable of extracting the physical relationships between explosive energy, blast geometry, and fragmentation.

Table 4: Importance of Parameters Used in Modeling by the XGBoost Model

Parameter	Raw Gain Score	Relative Importance based on Gain (%)
q	6.66	44.6
B	4.83	32.4
T	1.97	13.2
H	0.79	5.3
S	0.67	4.5

In addition to the data-driven XGBoost model, the empirical Kuz–Ram model (Equation 6) was also employed in this study for predicting P_{50}, in order to enable a comparative assessment of predictive performance (results presented in Table 5). The Kuz–Ram model is an empirical approach for estimating the mean fragment size of blasted rock, developed based on relationships among explosive energy, blast-hole geometry, and rock properties. It incorporates parameters such as the rock factor (K), the volume of rock broken per hole (V_0), the explosive energy (Q), the relative strength of the explosive (E), and the maximum allowable fragment size (X_m) to predict P_{50}, thereby providing blasting engineers with a practical tool for preliminary blast design and evaluation [5, 13].

$$P_{50} = K.\sqrt{X_m.\frac{V_0^{0.8}}{Q}}.(Q)^{\frac{1}{6}}.(\frac{115}{E})^{\frac{19}{30}} \tag{6}$$

The mathematical formulation of the model is given in Equation (6), and the input parameters are as follows:

- K (Rock factor): depends on rock type and jointing, determined empirically. Typical values include 7 for medium rocks, 10 for hard jointed rocks, and 13 for hard massive rocks, with a direct influence on P_{50}.
- V_0 (Volume of rock broken per hole): usually approximated as the product of Spacing (S), Burden (B), and Bench height (Hb). Its unit is cubic meters, and larger volumes generally increase the average fragment size.
- Q (Explosive energy per hole): calculated as the product of the linear charge density (q) and the explosive column height (h). Higher Q values typically yield finer fragmentation.
- E (Relative strength of explosive): benchmarked against TNT, with TNT = 115 and ANFO = 100, accounting for effective energy delivery.
- X_m (Maximum allowable fragment size): defines the operational upper limit of fragment size, often taken as 0.7 m in open-pit iron mines.

Table 5: Evaluation Metrics of the Kuz–Ram Model for Training and Testing Data

Evaluation criteria	Train data	Test data
R^2	0.85	0.74
RMSE	3.25	3.91
MAE	2.71	3.44

As the results demonstrate, the data-driven XGBoost model significantly outperformed the empirical Kuz–Ram model, due to its capability of capturing complex nonlinear interactions and combined effects of blasting parameters and geological conditions. In contrast, while Kuz–Ram offers simplicity and rapid applicability as an empirical reference, it falls short in accuracy owing to its inherent limitations, such as the inability to represent nonlinear complexities and incomplete adaptability to diverse geological settings.

4.3. Multi-objective Optimization Using NSGA-II

Following the prediction of rock fragmentation index (P_{50}) using the XGBoost model, the next step involved a multi-objective optimization process targeting two conflicting objectives: minimizing P_{50} and minimizing blasting cost (C). To this end, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) was applied, which is well-known for its robustness in exploring multi-dimensional search spaces and identifying Pareto-optimal solutions.

In the optimization process:

- The cost function (introduced in the previous section) was incorporated into the algorithm to account for expenses associated with drilling, charging, explosives, and manpower.
- The trained XGBoost model was integrated into NSGA-II to ensure that fragmentation quality (P_{50}) was simultaneously considered alongside cost.

The outcome was a set of Pareto-optimal solutions representing trade-offs between cost and fragmentation. These results were visualized in a Pareto front chart, where each point denotes an optimal trade-off solution. For each non-dominated solution, the corresponding optimal values of blast design parameters—including Burden (B), Spacing (S), Sub-drilling (H), and Specific charge(q)—were extracted and stored in an Excel file. This provides engineers with actionable insights for analyzing and selecting the most suitable parameter combinations for practical mining operations.

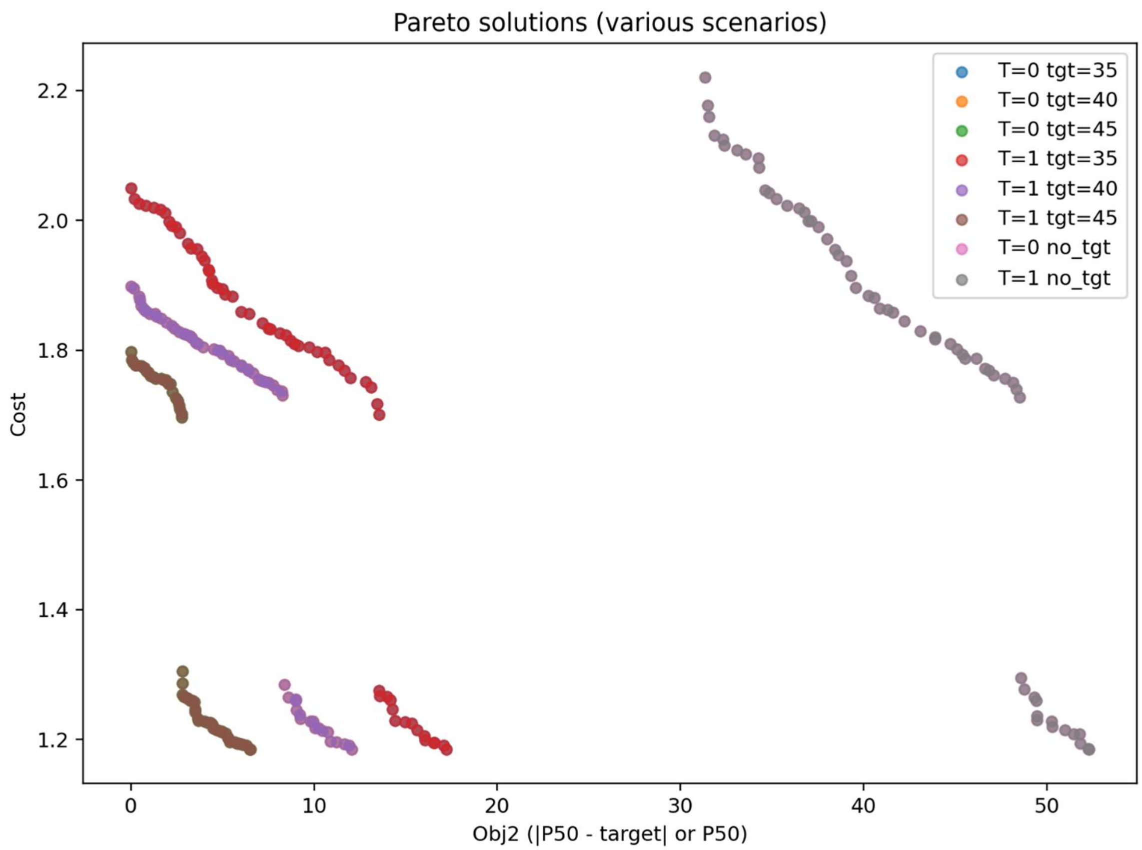


Figure 7: A set of Pareto fronts resulting from the simultaneous optimization of two objectives: minimizing total blast cost and minimizing crushing prediction error

Figure 7 illustrates the Pareto fronts obtained from the simultaneous optimization of blasting cost and P_{50} prediction error. The NSGA-II, when coupled with the XGBoost model, successfully generated a diverse set of non-dominated solutions, each reflecting different blast design trade-offs. As observed, the Pareto fronts exhibit varied distributions under different scenarios (e.g., based on rock type conditions, T=0 or T=1). For instance, in cases where the target P_{50} was defined around 35–40 mm, the optimal solutions clustered near the lower prediction error axis (Obj2 ≈ 0), while costs remained at moderate-to-low levels. This indicates that achieving target fragmentation with relatively low cost is feasible. Conversely, in scenarios with no explicit target fragmentation (no_tgt), the Pareto fronts were more dispersed, including some solutions with very low cost but extremely high error, which are operationally less desirable.

The key insight is that Pareto fronts empower decision-makers to select solutions aligned with their priorities:

- If cost reduction is the primary objective, solutions in the lower-cost region (C < 1.3) may be preferred, even at the expense of slightly higher fragmentation error.
- Conversely, if achieving precise fragmentation (P_{50}) is more critical, solutions closer to the vertical axis (low error, Obj2 ≈ 0) are more suitable, albeit at higher costs.

Overall, the results demonstrate that integrating NSGA-II with XGBoost achieves a dynamic balance between conflicting objectives—cost and fragmentation quality—thereby facilitating data-driven and optimized blast design.

5. Discussion

The results of this study demonstrated that the integrated framework—comprising the machine learning model XGBoost for predicting the fragmentation index (P_{50}) and the evolutionary multi-objective optimization algorithm NSGA-II—is capable of efficiently modeling and managing the multidimensional and nonlinear complexities of blasting processes. This framework not only captures the direct dependencies between blast design variables and fragmentation indices but also uncovers their nonlinear interactions. The variable importance analysis revealed that the specific charge(q) and the burden (B) are the primary controlling parameters influencing variations in P_{50}, contributing the largest share to the predictive capacity of the model. Specific charge represents the rate of energy release per unit volume of rock, whereas B governs the degree to which the released energy is transmitted to the free surface, thereby facilitating stress release and blast-wave propagation. Any deviation from the optimal ranges of q and B may compromise blasting efficiency and lead to substantial changes in the particle size distribution. In contrast, parameters such as spacing (S) and subdrilling (H) had relatively minor contributions. These variables serve secondary, corrective roles, and their effects become more pronounced only when q and B fall outside their optimal ranges. From a geological perspective, the rock mass condition (T) exhibited a notable influence on P_{50} variability. This parameter particularly distinguishes between two fundamental states of the rock mass: intact versus jointed rock. In jointed rocks, pre-existing discontinuities act as preferential paths for stress wave propagation and crack extension, resulting in heterogeneous fragmentation or undesired increases in P_{50}. Conversely, in intact rocks, the blast energy is primarily consumed in creating new fractures, producing a more uniform fragmentation pattern. Thus, even under optimal geometric design, the structural features of the rock mass play a decisive role in the final outcome.

In terms of modeling performance, XGBoost exhibited superior predictive ability. High coefficients of determination (R^2) in both training and testing datasets confirmed the model's ability to capture the nonlinear and complex interactions between blasting parameters and fragmentation. Residual analysis revealed that errors were randomly and evenly distributed around zero, with no systematic bias observed, which highlights the model's robustness and generalizability.

During the multi-objective optimization phase, NSGA-II successfully extracted a set of Pareto-optimal solutions that represented diverse trade-offs between cost and fragmentation quality. The Pareto front enabled flexible decision-making: in scenarios prioritizing cost reduction, low-cost solutions with slightly larger P_{50} values could be selected; whereas in cases emphasizing finer fragmentation, solutions near the quality axis (lower P_{50}) were preferable, despite higher costs. This flexibility provides mining engineers and managers with a strategic tool, aligning data-driven insights with operational and economic goals.

6. Conclusion

This research yielded several significant and practical contributions, which can be summarized as follows:

1. The XGBoost model effectively captured complex nonlinear relationships between blasting parameters and fragmentation index (P_{50}), outperforming classical empirical models and demonstrating strong generalization capability.

2. Feature importance analysis confirmed that q and B are the key determinants of fragmentation quality, underscoring the necessity of precise control over explosive energy per unit volume and blast geometry in blast design. These findings are consistent with established literature, reinforcing the critical role of these parameters.
3. The geological condition (T) was shown to exert a substantial influence on fragmentation outcomes. This finding emphasizes that even optimally designed blast patterns cannot guarantee desirable fragmentation without adequate understanding of the rock mass characteristics.
4. The evolutionary multi-objective algorithm NSGA-II enabled simultaneous optimization of conflicting objectives such as cost and fragmentation quality. The resulting Pareto front provides decision-makers with a range of solutions adaptable to real-world mining scenarios.
5. By integrating XGBoost with NSGA-II, this study introduced a novel framework that is relatively underexplored in the blasting engineering literature. Beyond its academic contribution, the framework has high practical applicability and can be directly implemented in real-world blast design.

Despite the promising results, certain limitations of this study must be acknowledged. The most notable is the relatively small dataset comprising 36–37 blasting records from a single mine. While the XGBoost model demonstrated strong predictive performance, the limited sample size inherently increases the risk of overfitting and reduces statistical power. Consequently, the current findings should be considered preliminary, and further validation with larger datasets is necessary before generalizing the conclusions to broader blasting conditions.

Based on the findings, future research is recommended to expand the dataset across different mines with varied geological conditions to enhance model generalizability. Moreover, incorporating other machine learning algorithms such as LightGBM or deep neural networks would allow comparative evaluation of predictive performance. Additionally, considering alternative fragmentation indices such as P80 and fines content within the optimization process would enrich the analysis. Finally, the development of multi-dimensional cost functions that integrate environmental impacts of blasting—such as ground vibrations, flyrock, and noise pollution—would contribute to more sustainable and holistic blast design optimization.

Authors' Contributions

All authors equally contributed to this study.

Declaration

In order to correct and improve the academic writing of our paper, we have used the language model ChatGPT.

Transparency Statement

Data are available for research purposes upon reasonable request to the corresponding author.

Acknowledgments

By researchers in Bioinformatics, Computational Biology, Artificial Intelligence and Medicine.

Declaration of Interest

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

Funding

According to the authors, this article has no financial support.

Ethical Considerations

Not applicable.

References

[1] B. Ke *et al.*, "Parameter optimization and fragmentation prediction of fan-shaped deep hole blasting in Sanxin gold and copper mine," *Minerals,* vol. 12, no. 7, p. 788, 2022.
[2] J. Guo, Z. Zhao, P. Zhao, and J. Chen, "Prediction and optimization of open-pit mine blasting based on intelligent algorithms," *Applied Sciences,* vol. 14, no. 13, p. 5609, 2024.
[3] B. Cahyaningsih, A. P. Putra, D. G. P. Harsono, and V. H. Avesina, "EVALUATION STUDY OF BLAST-INDUCED ROCK FRAGMENTATION USING STEMROCK AND BLASTBAG AIR DECK SYSTEMS WITH THE KUZ-RAM METHOD AT THE LIMESTONE QUARRY NAROGONG SITE OF PT SOLUSI BANGUN INDONESIA TBK," *Jurnal Sains dan Teknologi Reaksi,* vol. 23, no. 01, pp. 42-49, 2025.

[4] A. Hekmat, S. Munoz, and R. Gomez, "Prediction of rock fragmentation based on a modified Kuz-Ram model," in *Proceedings of the 27th international symposium on mine planning and equipment selection-MPES 2018*, 2019: Springer, pp. 69-79.

[5] O. Yilmaz, "Rock factor prediction in the Kuz–Ram model and burden estimation by mean fragment size," *Geomechanics for Energy and the Environment,* vol. 33, p. 100415, 2023.

[6] X. Geng, S. Wu, Q. Yan, J. Sun, Z. Xia, and Z. Zhang, "An optimized XGBoost model for predicting tunneling-induced ground settlement," *Geotechnical and Geological Engineering,* vol. 42, no. 2, pp. 1297-1311, 2024.

[7] I. Krop, T. Sasaoka, H. Shimada, and A. Hamanaka, "Optimizing mean fragment size prediction in rock blasting: a synergistic approach combining clustering, hyperparameter tuning, and data augmentation," *Eng,* vol. 5, no. 3, pp. 1905-1936, 2024.

[8] X. Liu, H. Yang, H. Jing, X. Sun, and J. Yu, "Research on intelligent risk early warning of open-pit blasting site based on deep learning," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects,* vol. 47, no. 1, pp. 6355-6371, 2025.

[9] S. Q. Liu *et al.*, "Machine learning for open-pit mining: a systematic review," *International Journal of Mining, Reclamation and Environment,* vol. 39, no. 1, pp. 1-39, 2025.

[10] Y. Zhang, Y. Qiu, K. Du, H. Nguyen, D. J. Armaghani, and J. Zhou, "Optimizing Flyrock Forecasting in Open-Pit Blasting Using Hybrid Machine Learning Models," *Rock Mechanics and Rock Engineering,* pp. 1-28, 2025.

[11] Y. Yusoff, M. S. Ngadiman, and A. M. Zain, "Overview of NSGA-II for optimizing machining process parameters," *Procedia Engineering,* vol. 15, pp. 3978-3983, 2011.

[12] S. Verma, M. Pant, and V. Snasel, "A comprehensive review on NSGA-II for multi-objective combinatorial optimization problems," *IEEE access,* vol. 9, pp. 57757-57791, 2021.

[13] C. Cunningham, "The Kuz-Ram fragmentation model–20 years on," in *Brighton conference proceedings*, 2005, vol. 4: European Federation of Explosives Engineer Brighton, UK, pp. 201-210.

[14] J. Silva, J. Amaya, and F. Basso, "Development of a predictive model of fragmentation using drilling and blasting data in open pit mining," *Journal of the Southern African Institute of Mining and Metallurgy,* vol. 117, no. 11, pp. 1089-1094, 2017.

[15] A. K. Raj, B. S. Choudhary, and G. W. Deressa, "Prediction of rock fragmentation for surface mine blasting through machine learning techniques," *Journal of The Institution of Engineers (India): Series D,* vol. 106, no. 1, pp. 641-661, 2025.

[16] J. Zhao, D. Li, J. Zhou, D. J. Armaghani, and A. Zhou, "Performance evaluation of rock fragmentation prediction based on RF-BOA, AdaBoost-BOA, GBoost-BOA, and ERT-BOA hybrid models," *Deep Underground Science and Engineering,* vol. 4, no. 1, pp. 3-17, 2025.

[17] E. Bakhtavar, R. Sadiq, and K. Hewage, "Optimization of blasting-associated costs in surface mines using risk-based probabilistic integer programming and firefly algorithm," *Natural Resources Research,* vol. 30, no. 6, pp. 4789-4806, 2021.

[18] J. Xiang, J. Chen, A. Zhang, X. Zhao, S. Zhuo, and S. Yang, "Multi-Objective Ore Blending Optimization for Polymetallic Open-Pit Mines Based on Improved Matter-Element Extension Model and NSGA-II," *Mathematics,* vol. 13, no. 11, p. 1843, 2025.

[19] Z. Chen, P. Song, R. Liu, Q. Wang, and Y. Shen, "Optimization of ore production scheduling strategy using NSGA-II-GRA in open-pit mining," *Scientific Reports,* vol. 15, no. 1, p. 10376, 2025.

[20] J. Wang, L. Xu, S. Sun, Y. Ma, and G. Yu, "Multi-objective optimization using improved NSGA-II for integrated process planning and scheduling problems in a machining job shop for large-size valve," *Plos One,* vol. 19, no. 6, p. e0306024, 2024.

[21] M. A. Cotrina Teatino *et al.*, "Optimization of Fragmentation and Operational Costs of Drilling and Blasting using Hybrid Machine Learning Models in an Open-Pit Mine in Peru," *Journal of Mining and Environment,* vol. 16, no. 4, pp. 1195-1219, 2025.

[22] J. Guo, P. Zhao, and P. Li, "Prediction and optimization of blasting-induced ground vibration in open-pit mines using intelligent algorithms," *Applied Sciences,* vol. 13, no. 12, p. 7166, 2023.

[23] X. Song *et al.*, "Quantitative classification evaluation model for tight sandstone reservoirs based on machine learning," *Scientific Reports,* vol. 14, no. 1, p. 20712, 2024.

[24] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794.

[25] P. Zhang, Y. Jia, and Y. Shang, "Research and application of XGBoost in imbalanced data," *International Journal of Distributed Sensor Networks,* vol. 18, no. 6, p. 15501329221106935, 2022.

[26] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE transactions on evolutionary computation,* vol. 6, no. 2, pp. 182-197, 2002.

[27] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," *Peerj computer science,* vol. 7, p. e623, 2021.

[28] S. Patro and K. K. Sahu, "Normalization: A preprocessing stage," *arXiv preprint arXiv:1503.06462,* 2015.