SciEn

Proceedings of 8th International Conference on Mechanical Engineering and Renewable Energy 2025
10-12 December, 2025, Chattogram, BANGLADESH

SciEn Conference Series: Engineering Vol. 4, 2026, pp 117-122 https://doi.org/10.38032/scse.2026.4.60

Hybrid GWO-XGBoost Framework for Blast-Induced Ground Vibration Prediction: Integration of Genetic Algorithm-Optimized Scaled Distance

Arnob Deb[1,], Tahmid Zaman Raad[2], Pusmita Roy[1], Md Tauhidur Rahman[3]*

[1]Department of Petroleum and Mining Engineering, Military Institute of Science & Technology, Dhaka-1216, Bangladesh
[2]Department of Industrial and Production Engineering, Military Institute of Science & Technology, Dhaka-1216, Bangladesh
[3]Bob L. Herd Department of Petroleum Engineering, Texas Tech University, 2500 Broadway, Lubbock, TX 79409, USA

ABSTRACT

Blast-induced ground vibrations, commonly measured in terms of Peak Particle Velocity (PPV), can affect adjoining communities, damage structures, and undermine the integrity of rock masses. Therefore, accurate PPV prediction is a prerequisite for making blasting sustainable and safe. This paper introduces a novel hybrid machine learning method that integrates Extreme Gradient Boosting (XGBoost) and Grey Wolf Optimization (GWO) in predicting PPV using 88 datasets obtained at the Akdaglar Quarry, Istanbul. Nine important input parameters were taken into account, including geological aspects and blasting design. In comparison to the traditional scaled distance, a new parameter modified scaled distance (MSD), was derived through a genetic algorithm, and showed better correlation with PPV (r = –0.869). Model performance was improved by optimal tuning of XGBoost hyperparameters through the utilization of the GWO method. The GWO-XGBoost model performed better than SVM and stand-alone XGBoost, with R = 0.9916, R^2 = 0.9748, RMSE = 1.0291 mm/s, and MAPE = 3.1105%. The monitoring distance and scaled distance are the most influential predictors, according to SHAP (Shapley Additive Explanations) analysis. The results demonstrate that evolutionary optimization algorithms substantially enhance PPV prediction accuracy, offering practical implications for pre-blast vibration assessment and sustainable quarry design.

Keywords: Peak Particle Velocity (PPV), Ground vibration, Gray Wolf Optimization (GWO), Extreme Gradient Boosting (XGBoost).

1. Introduction

Blasting, an encyclopedically recognized, cost-effective, and efficient step for rock fragmentation in mining engineering. This method employs explosives to disintegrate rock masses into smaller segments, speeding the separation as well as processing of the materials for subsequent use [2],[3]. A major segment of explosive energy unleashed throughout blasting, about 20–30%, is utilized for breaking and displacing the rock mass, while the remainder is lost to adverse effects such as noise, flyrock, ground vibrations, overbreak, and backbreak. These negative outcomes can have a detrimental impact on the surrounding environment.
Optimizing blasting operations and accurately predicting blast-induced ground vibrations can help alleviate negative impacts. The primary determinants for predicting vibration outcomes include frequency, duration, and PPV recognized and extensively utilized as a practical index predictions of vibration values. Research has shown that many predictive models predominantly rely on the quantity of explosives and the distance to establish an empirical relationship, which is determined by calculating the maximum charge per delay (MC) and the distance from the blast location (Di) [19]. Furthermore, researchers have introduced semi-empirical, statistical, and graphical methods for forecasting peak particle velocity (PPV) resultant blasting. However, these approaches have proven to be inadequate due to the limited number of parameters taken into account, as PPV is also affected by various other factors related to blast design [4],[5]. In current years, machine learning (ML) approaches have been increasingly employed to develop effective solutions for various challenges in mining engineering. Table 1 presents several studies related to the prediction of PPV.

Table 1 Analysis of associated studies in forecasting PPV using ML approaches during the last decade

References	Datasets Number	Superior Model	R^2	RMSE
[3]	252	HGS–ANN	0.922	1.761
[13]	192	ANFIS	0.98	~
[6]	191	ANN	0.948	0.0008
[8]	169	MLPNN	0.954	0.03
[15]	166	GOA–ELM	0.9105	2.855
[1]	154	GEP	0.91	5.78
[10]	14	GRNN	0.999	0.0001

*Corresponding Author Email Address: arnobdeb9147@gmail.com

2. Research Objective

Traditional methods for predicting these vibrations frequently depend on empirical models, which may struggle to take into consideration the complicated interrelationships between variables, leading to less accurate predictions[12],[14]. This study aims to achieve a primary objective by addressing a significant gap in existing research by utilizing advanced machine learning techniques, specifically XGBoost, to analyze ground vibrations for this dataset. Additionally, gray wolf optimization (GWO) is employed to optimize the hyperparameters of the XGBoost model. Furthermore, this research introduces a novel input parameter, the modified SD determined by a genetic algorithm, which exhibits a more potent correlation with PPV than the traditional SD, thereby offering a fresh approach to this domain.

3. Methods

3.1 Extreme Gradient Boosting (XGBoost)

T Chen developed XG-Boost [20], an artificial intelligence method characterized by its ability to generate boosted trees in a parallel manner for classification and regression problems, efficiently combining novel algorithms with GBDT approaches acting as a soft computing library.[9], [11], [16]. An overview of the XG-Boost algorithm is - A dataset, consisting number of k samples and p parameters denoted (as $H = \{(x_i, y_i)\}$ ($|H| = k$, $x_i \in R^p$, $y_i \in R$), an additive function (M) predicts the resulting values of a tree collaborative model as follows,

$$\hat{y}_i = \phi(x_i) = \sum M_{m=1} f_m(x_i),\ f_m \in S \quad (1)$$

where $f_m \in$ S and S is the set of regression trees. It can be computed like:

$$S = \{f(x) = \omega_{b(x)}\} \quad (2)$$

Where b: (b: $R^p \to T$, $\omega \in R^T$), the minimization function is utilized to reduce errors in the trees, where b signifies the tree framework, T symbolizes how many leaves are present in the tree, and f_m is the product of b and ω as the leaf weight. The minimization factor is computed in the XGBoost model:

$$L^{(t)} = \sum_{k=1}^{k} l\left(y_{i,}\hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \quad (3)$$

The minimization factor (L) is determined using the discriminant function of a convex objective for determining a deviation between the forecasted ($\hat{y}_i$) and the real values (y_i), iterations are denoted by t. The regression tree algorithms' effectiveness is as follows:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda \parallel w \parallel^2 \quad (4)$$

3.2 Gray Wolf Optimization (GWO)

The GWO algorithm is grounded in a hierarchical structure [17]. which optimizes problems by employing four key processes inspired by the wolf's natural hunting strategy: ranking, tracking, surrounding, and killing prey. To initiate the optimization process, a stochastic population is created within the selected space and divided into four distinct categories, labeled α, β, δ, and ω, with the ω group following the others as depicted in Fig. 1. These groups represent the three most effective strategies in each period that serve as the foundation of GWO's entire optimization process [16].

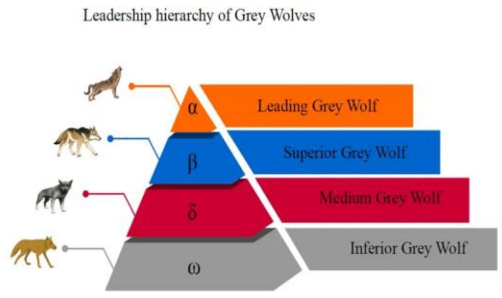

Fig. 1 Natural Supervision structure of gray wolves

Throughout the hunting process, gray wolves commonly encircle their quarry, as demonstrated by the numerical model utilized to replicate their behavior. The following equation represents this process:

$$\boldsymbol{Q}(t+1) = \boldsymbol{Q}_p(t) - (2\boldsymbol{a} \cdot \rho_1 - \boldsymbol{a})\,(2\rho_2 \cdot \boldsymbol{Q}_p(t) - \boldsymbol{Q}(t) \quad (5)$$

$$\boldsymbol{a} = 2 - 2 \cdot it/MAX - IT \quad (6)$$

Wherein, t signifies the current quantity of modifications, X_p(t) denotes the position vector regarding the target,
X(t) represents the positioning of the conventional gray wolf, and X(t+1) represents the location of the gray wolf in the subsequent modification. The value of 'a' can be computed with equation (6), and it declines gradually from 2 to 0. In this context, 'it' denotes a given quantity of modification, while MAX-IT indicates the highest quantity of modification. ρ1 and ρ2 are both arbitrary values between 0 and 1.

4. Materials

4.1 Site Description

The study used XGBoost and GWO-XGBoost models to predict ground vibrations, utilizing 88 datasets collected by Hudaverdi from the Akdaglar Quarry. To predict PPV, the machine learning models incorporated input variables such as the ratio of bench height to burden (H/B), burden to hole diameter (B/D), spacing to burden (S/B), sub-drilling to burden (U/B), stemming to burden (T/B), distance between monitoring stations (D), charge per delay (W), power factor (PF) and the modified scaled distance (SD).

4.2 Modified SD Developed by Genetic Algorithm (GA)

The term "scaled distance" refers to the amount of energy generated by explosives and the impact of distance on ground wave attenuation, specifically in air shock generation and seismic waves [15]. The more fundamental formula that is usually utilized in blasting regulations and fundamental blasting circumstances is the square root scale [7]. Though this SD is used in predicting PPV in the empirical approach, nevertheless the SD is also utilized as an input variable in ML models [8]. In this study, SD is employed as an input variable. Instead of applying traditional SD, we employed a modified version of SD which is developed by genetic algorithm (GA). Similar to traditional SD, D, and W are used in this development. The modified SD is defined as:

$$\text{Modified SD} = \frac{-2.438 * \text{W}^{0.2744}}{41.63765 * \text{D}^{0.528}} \quad (7)$$

The following equation pertains to analyzing the relationship between two variables, D (in meters) and W (in kilograms), which represent the distance and charge per delay, respectively. The modified SD is developed to enhance the correlation with the output variable, PPV portrayed in Fig. 2. The traditional SD and the modified SD are compared, revealing a stronger correlation with PPV using the modified SD, which has a value of -0.869. Consequently, this modified SD has significant implications for predicting PPV, and a new approach has been introduced that employs a GA to incorporate this parameter as an input variable.

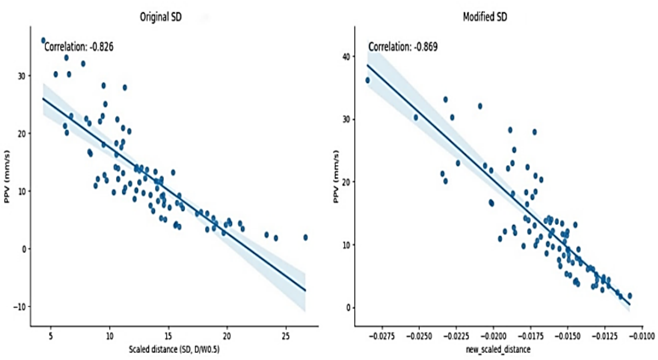

Fig. 2 Correlation comparison: SD Vs Modified SD

4.3 Data Visualization

Violin plots and box plots are shown in Fig. 3 and 4 for data visualization of each input and output parameter, where the distribution of the dataset is analyzed. A violin plot's primary benefit over a box plot is its ability to indicate density; a broader plot indicates a higher density. For example, in the violin plot for H/B in Figure 3, the highest and lowest values, demonstrated by the top and bottom of the violin plot, are 2.4 and 4.4, respectively. The black dot represents the average, which is 3.2. The violin plot is widest around 3.2, indicating the highest density for this dataset. However, high values have a small density at 4.4. Fig. 5 shows a Pearson correlation matrix, highlighting the correlations between PPV and various parameters by illustrating the pairwise relationships, with correlation coefficients shown for all variables. Certain parameters show a moderately strong correlation, with PPV being particularly strongly correlated with D and modified SD, having correlation values of -0.78 and -0.87, respectively.

This research developed a hybrid GWO-XGBoost model, which encompasses four key processes integral to the study: (1) preparation of the dataset; (2) development of the model; (3) validation and assessment; and (4) outcome evaluation. The 88 datasets, consisting of nine input variables, were randomly partitioned into training (80%) and testing (20%) subsets. An optimization technique was focused on adjusting the hyperparameters of the XGBoost model. In this context, this method served as the foundation for forecasting, with optimal hyperparameters identified using the GWO algorithm. The optimal XGBoost model is defined by maximizing the coefficient of determination (R^2). Ultimately, the model's performance was analyzed by exploiting the R, RMSE, MAPE, and R^2. The GWO-XGBoost approach emerged as the highest-performing model.

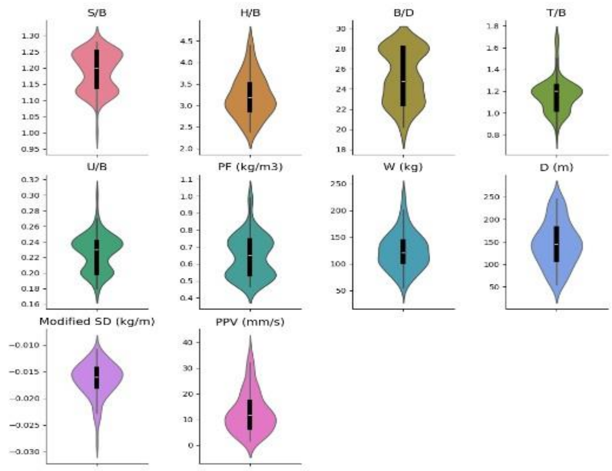

Fig. 3 Violin plots of parameters

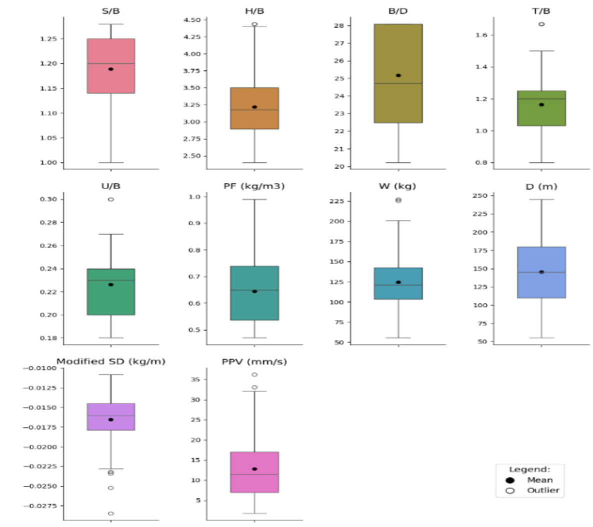

Fig. 4 Box plots of parameters used in ML modeling

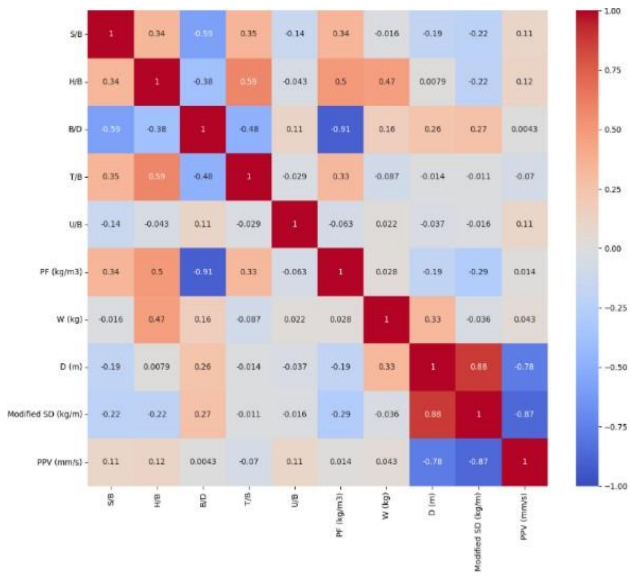

Fig. 5 Pearson correlation between parameters

4.4 Validating and Assessing Model Performance

To comprehensively assess the dependability of hybrid models, this study integrates four performance metrics, namely r, MAPE, R^2, and RMSE [6],[18],[19] The performance metrics are expressed mathematically as follows:

Root Mean Square Error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\mathrm{PPV}_i - \widehat{\mathrm{PPV}}_i\right)^2} \tag{8}$$

Pearson's coefficient of correlation (r):

$$r = \frac{\sum_{i=0}^{n-1}\left(\left(\mathrm{PPV}_i-\overline{\mathrm{PPV}}\right)*\left(\widehat{\mathrm{PPV}}_i-\overline{\widehat{\mathrm{PPV}}}\right)\right)}{\sqrt{\sum_{i=0}^{n-1}\left(\mathrm{PPV}_i-\overline{\mathrm{PPV}}\right)^2}*\sqrt{\sum_{i=0}^{n-1}\sum\left(\widehat{\mathrm{PPV}}_i-\overline{\widehat{\mathrm{PPV}}}\right)^2}} \tag{9}$$

Coefficient of determination (**R^2**):

$$R^2 = \frac{\sum_{i=1}^{r}(\mathrm{PPV}-\overline{\mathrm{PPV}})^2-\sum_{i=1}^{r}(\mathrm{PPV}-\widehat{\mathrm{PPV}}_i)^2}{\sum_{i=1}^{r}(\mathrm{PPV}-\overline{\mathrm{PPV}})^2} \tag{10}$$

Mean absolute percentage error (MAPE):

$$R^2 = \frac{\sum_{i=1}^{r}(\mathrm{PPV}-\overline{\mathrm{PPV}})^2-\sum_{i=1}^{r}(\mathrm{PPV}-\widehat{\mathrm{PPV}}_i)^2}{\sum_{i=1}^{r}(\mathrm{PPV}-\overline{\mathrm{PPV}})^2} \tag{11}$$

Here, n indicates the quantity of samples in the testing period, $\widehat{\overline{\mathrm{PPV}}}$ = the forecasted PPV value, PPV = the true PPV value, $\overline{\mathrm{PPV}}$ = the average PPV value, and Var = variable.

5. Results and Discussion

Fig. 6 displays the various agent values of the GWO-XGBoost model. It is significant to note that the performance of the different agent values was exceptionally high, making it difficult to determine the most effective number of agents. The initial step in developing the hybrid model focused on implementing the XGBoost technique, which aimed to optimize the parameters for enhanced outcomes. The hyperparameter values of the GWO-XGBoost model are shown in Table 2.

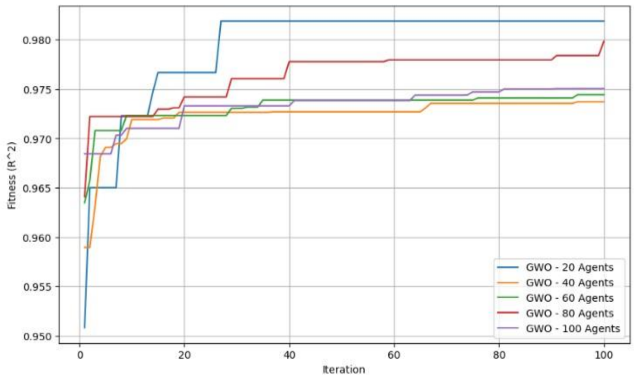

Fig. 6 Iteration curve of the GWO algorithm

Table 2 Optimal hyperparameters for GWO-XGBoost Model

Optimal parameters	Value
Max Depth	5
Min Child Weight	4
Subsample	0.7860
N Estimators	545
Colsample Bytree	0.6882
Learning Rate	0.012
Agents	100

The assessment of the fitted models' approximation capabilities was performed using statistical performance metrics, evaluated based on the test dataset. The results are exhibited in Table 3.

Table 3 An outline of the models' respective performance metrics

Various Models	Evaluation Criteria			
	R	R²	RMSE	MAPE
SVM	0.9336	0.8504	3.0614	27.9074
XGBoost	0.9731	0.9426	1.8958	5.8514
GWO-XGBoost	0.9916	0.9748	1.0291	3.1105

According to the data presented in Table 3, the hybrid XGBoost models exhibited superior performance relative to the support vector machine (SVM) and XGBoost techniques concerning various error metrics, including R, R^2, RMSE, and MAPE. The accompanying graphs in Fig. 9, Fig. 8, and Fig. 7 show the connection between measured and anticipated points and the PPV prediction outcomes.

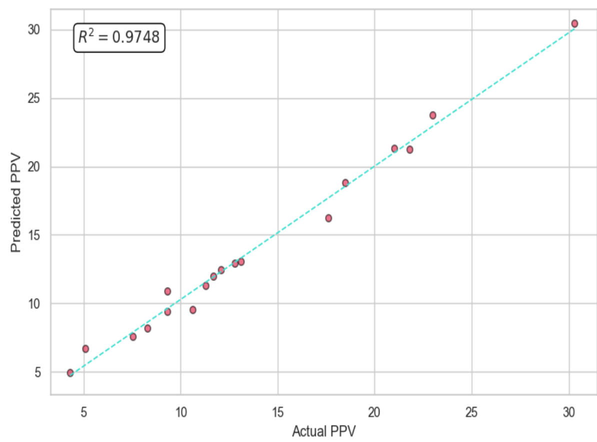

Fig. 7 Actual vs. GWO-XGBoost anticipated PPV

Fig. 10 integrates feature effects and relevance, utilizing SHAP values to show each feature's weight distribution on the output of the model. The graphic emphasizes the importance of both new SD and D characteristics for predicting the output variable. High values of both new SD and D negatively impact prediction accuracy.

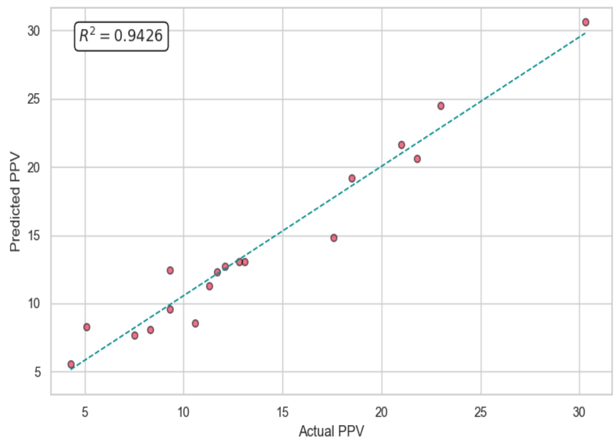

Fig. 8 Actual vs. XGBoost anticipated PPV

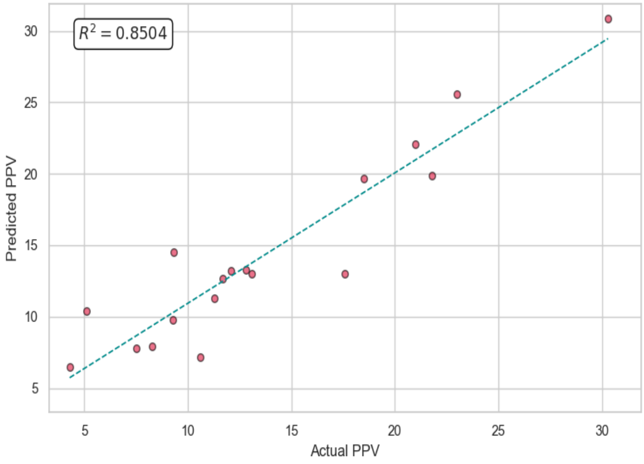

Fig. 9 Actual vs. SVM anticipated PPV

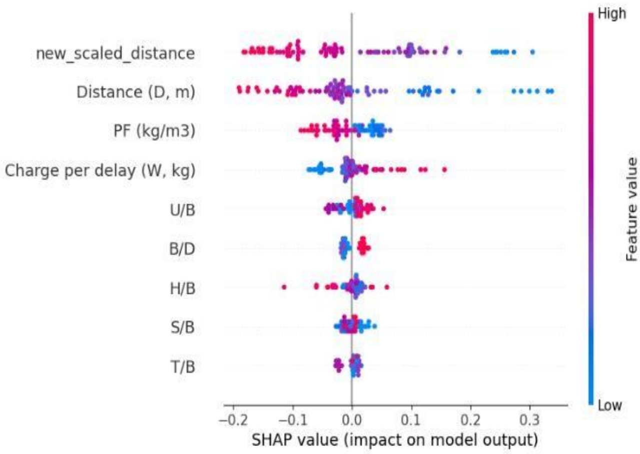

Fig. 10 Global explanation of the GWO-XGBoost model

The contrast between the true PPV value and the hybrid model's anticipated PPV strength, as shown in the accompanying Fig. 11, supports these conclusions.

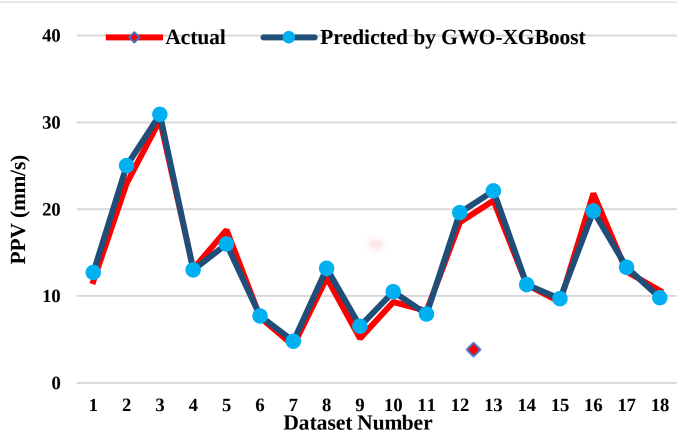

Fig. 11 Forecasted and actual PPV for testing datasets

6. Conclusion

Classical PPV prediction models exhibit limitations in accuracy, whereas ML approaches demonstrate superior capability in handling complex nonlinear parameters. Consequently, advanced ML techniques, particularly when enhanced by meta-heuristic optimization, offer a promising path for reliable pre-blast PPV estimation.

- Enhancing XGBoost performance necessitates the exploration and application of diverse meta-heuristic algorithms to effectively tune its numerous complex parameters.
- Future research must significantly expand training datasets (beyond the current limitation of 70 cases) and incorporate more significant blasting and geological characteristics to improve model robustness and generalizability.
- The effectiveness of the proposed optimized ML model and its general applicability require rigorous evaluation against other optimization techniques and diverse, larger datasets.

Accurate PPV forecasting is vital for optimizing blast design, minimizing environmental disturbances, and maximizing explosive energy utilization. Continued refinement through robust optimization and comprehensive data remains essential for practical deployment.

7. References

[1] P. Bayat, M. Monjezi, A. Mehrdanesh, and M. Khandelwal, "Blasting pattern optimization using gene expression programming and grasshopper optimization algorithm to minimise blast-induced ground vibrations," *Engineering with Computers*, vol. 38, no. 4, pp. 3341–3350, 2022.

[2] G. Chen and S. L. Huang, "Analysis of ground vibrations caused by open pit production blasts: A case study," *Fragblast*, vol. 5, no. 1–2, pp. 91–107, 2001.

[3] K. Hino, "Fragmentation of rock through blasting," *Journal of the Industrial Explosive Society*, 1996.

[4] T. Hudaverdi and O. Akyildiz, "Prediction and evaluation of blast-induced ground vibrations for structural damage and human response," *Arabian Journal of Geosciences*, vol. 14, no. 5, p. 378, 2021.

[5] M. Khandelwal and T. N. Singh, "Evaluation of blast-induced ground vibration predictors," *Soil Dynamics and Earthquake Engineering*, vol. 27, no. 2, pp. 116–125, Feb. 2007.

[6] A. I. Lawal, S. Kwon, and G. Y. Kim, "Prediction of the blast-induced ground vibration in tunnel blasting using ANN, moth-flame optimized ANN, and gene expression programming," *Acta Geophysica*, vol. 69, no. 1, pp. 161–174, 2021.

[7] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014.

[8] H. Nguyen and X.-N. Bui, "A novel hunger games search optimization-based artificial neural network for predicting ground vibration intensity," *Natural Resources Research*, vol. 30, no. 5, pp. 3865–3880, 2021.

[9] H. Nguyen, X.-N. Bui, H.-B. Bui, and D. T. Cuong, "Developing an XG-Boost model to predict blast-induced peak particle velocity in an open-pit mine: A case study," *Acta Geophysica*, vol. 67, no. 2, pp. 477–490, 2019.

[10] P. Ragam and D. S. Nimaje, "Assessment of blast-induced ground vibration using different predictor approaches—a comparison," *Chemical Engineering Transactions*, vol. 66, pp. 487–492, 2018.

[11] X. Ren, H. Guo, S. Li, S. Wang, and J. Li, "A novel image classification method with CNN-XGBoost

model," in *Lecture Notes in Computer Science*, pp. 378–390, 2017.

[12] P. P. Roy, "Putting ground vibration predictions into practice," *Colliery Guardian*, vol. 241, no. 2, 1993.

[13] T. N. Singh, L. K. Dontha, and V. Bhardwaj, "Study into blast vibration and frequency using ANFIS and MVRA," *Mining Technology*, vol. 117, no. 3, pp. 116–121, 2008.

[14] S. A. Taqieddin, "Ground vibration levels: Prediction and parameters," *Mining Science and Technology*, vol. 3, no. 2, pp. 111–115, 1986.

[15] C. Yu, M. Koopialipoor, B. R. Murlidhar, A. S. Mohammed, D. J. Armaghani, E. T. Mohamad, and Z. Wang, "Optimal ELM–Harris Hawks Optimization and ELM–Grasshopper Optimization models to forecast peak particle velocity resulting from mine blasting," *Natural Resources Research*, vol. 30, no. 3, pp. 2647–2662, 2021.

[16] Z. Yu, X. Shi, J. Zhou, X. Chen, X. Miao, B. Teng, and T. Ipangelwa, "Prediction of blast-induced rock movement during bench blasting: Use of Gray Wolf Optimizer and support vector regression," *Natural Resources Research*, vol. 29, no. 2, pp. 843–865, 2020.

[17] J. Zhou, Y. Qiu, M. Khandelwal, S. Zhu, and X. Zhang, "Developing a hybrid model of Jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations," *International Journal of Rock Mechanics and Mining Sciences*, vol. 145, p. 104856, 2021.

[18] J. Zhou, Y. Qiu, S. Zhu, D. J. Armaghani, C. Li, H. Nguyen, and S. Yagiz, "Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate," *Engineering Applications of Artificial Intelligence*, vol. 97, p. 104015, 2021.

[19] M. Amiri, M. Hasanipanah, and H. Bakhshandeh Amnieh, "Predicting ground vibration induced by rock blasting using a novel hybrid of neural network and itemset mining," *Neural Computing and Applications*, vol. 32, no. 18, pp. 14681–14699, Sep. 2020.

[20] T. Chen, "XGBoost: Extreme gradient boosting," *R Package Version*, vol. 1, p. 1, 2015.