mining

MDPI

Article

Application of White-Box Machine Learning Models for the Prediction of Blast-Induced Peak Particle Velocity

Mehrshad Samadi [1], Seyed Amir Konjkav-Sabzevari [2] and Zohreh Sheikh Khozani [3,*]

[1] Department of Civil Engineering, Iran University of Science and Technology, Tehran 13114-16846, Iran; mehrshad1364@gmail.com
[2] Faculty of Mining, Petroleum & Geophysics Engineering, Shahrood University of Technology, Shahrood 36199-95161, Iran; amirkonjkaav@gmail.com
[3] Paleoclimate Dynamics Group, Alfred Wegener Institute, Helmholtz Center for Polar and Marine Research, 27570 Bremerhaven, Germany
* Correspondence: y.sheikh2131@gmail.com

Abstract

Blast-induced ground vibration is a critical environmental hazard in open-pit mining operations, capable of causing severe damage to adjacent structures and infrastructure. Accurately predicting vibration intensity is universally quantified by the peak particle velocity (*PPV*) index. Among the various factors affecting blast-induced ground vibrations, the distance from the blast face to the monitoring point (*D*) and the charge weight per delay (*W*) are the most influential and controllable parameters in a specific mine site. Therefore, these variables were selected as inputs for *PPV* estimation. The present study develops advanced white-box machine learning (ML) models, including Multi-Expression Programming (MEP), Gene Expression Programming (GEP), Multivariate Adaptive Regression Splines (MARS), and Stronger Variable Creator Machines (SVCMs), for predicting *PPV*. The general explicit equation was derived from the developed ML models implemented in a spreadsheet program, which can be easily used to estimate *PPV*. The MEP model achieves the highest accuracy, with a correlation coefficient (CC) of 0.993177 and a root mean square error (RMSE) of 0.836337, followed by the MARS, GEP, and SVCM models. The results of the present study, supported by k-fold cross-validation and parametric analysis, confirmed the potential of the proposed models for *PPV* estimation.

Keywords: peak particle velocity; machine learning models; ground vibration; blasting

Check for updates

Academic Editor: Hossein Masoumi

Received: 27 June 2026
Revised: 21 July 2026
Accepted: 23 July 2026
Published: 27 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Blasting is the most common method for breaking rock in mining and civil excavation projects, such as tunneling and quarrying. However, the collateral damage it causes cannot be ignored. The blasting creates intense shock waves and high-pressure gases that fracture the rock mass. It can cause serious environmental impacts, such as fly rock, ground vibration, and seismic waves that damage nearby residential areas, underground structures, and mining and infrastructure facilities, and it can cause structural damage to adjacent facilities. Therefore, accurate prediction of blasting intensity in mining projects is crucial. Ground vibration intensity is quantified using the peak particle velocity (*PPV*) index. Therefore, accurate and reliable prediction of *PPV* values is essential for safe design, environmental protection, and the safety of the surrounding environment in blast mining projects. *PPV* prediction has always been a major research topic in blasting engineering. Early research on PPV prediction has widely used empirical and regression-based equations

due to their simplicity and practical applicability, particularly the relationships provided by the United States Bureau of Mines (USBM). In recent years, computational intelligence has significantly expanded for resolving highly complex engineering problems [1]. In the field of blasting engineering, although empirical models are useful, they have limitations in capturing the complex interactions among parameters that affect *PPV*. This limitation has motivated the widespread use of ML methods to simulate nonlinear relationships directly from measured blasting data and has led to the increasing adoption of a wide spectrum of ML techniques for estimating *PPV* [2].

The application of ML models utilizing only two variables, namely, the distance from the blast face to the monitoring point (*D*) and the charge weight per delay (*W*), together with relatively small datasets, is well-documented in the blast vibration literature. Previous studies have demonstrated that ML methods, including artificial neural networks (ANNs), support vector machines (SVMs), and fuzzy logic systems, can accurately predict PPV [3–8]. These investigations confirm that reliable PPV prediction models can be developed using limited field datasets when the most influential blasting parameters are considered. In addition, optimization algorithms are widely used to improve the predictive accuracy of ML models by tuning their hyperparameters to estimate PPV values. For example, Rong et al. [9] integrated a support vector regression (SVR) model with five evolutionary algorithms and found that the gray wolf optimization (GWO) was more efficient at optimizing SVR parameters. Ogunsola et al. [10] developed three ANNs: a standard Levenberg–Marquardt model (ANN-LM) and two hybrid models optimized by the grasshopper optimization algorithm (ANN-GOA) and the multiverse optimization algorithm (ANN-MVO). Their comparative analysis revealed that the ANN-GOA model achieved the highest accuracy. Khatti et al. [11] developed six hybrid least square support vector machine (LSSVM) models by combining them with various evolutionary algorithms, including artificial bee colony (ABC), genetic algorithm (GA), GWO, Harris hawks optimization (HHO), particle swarm optimization (PSO), and salp swarm optimization (SSO). Their analysis revealed that the HHO-LSSVM models outperform the other hybrid models. Niu et al. [12] simulated online blast monitoring by developing a hybrid whale optimization algorithm (WOA) to optimize the initial weights and biases of ANNs. Their hybrid ANN model achieved higher accuracy than a standalone ANN model. Zhao et al. [13] proposed a hybrid chaos game optimization with ANN (CGO-ANN) and compared its performance with ANN-PSO, ANN-GA, and a single ANN model. They reported that the CGO-ANN model significantly outperformed other models. Bhattacharjee et al. [14] developed a three-layered ANN configuration and employed three distinct training algorithms: Levenberg–Marquardt (LM), Bayesian regularization (BR), and scaled conjugate gradient (SCG). They compared the performance of these networks against a traditional multiple linear regression (MLR) model. Their results showed that the ANN with BR and LM training algorithms presented the most accurate predictions. Yuan et al. [15] proposed an approach that integrates variational mode decomposition (VMD) with the CatBoost algorithm. In addition, they tuned hyperparameters using the improved hippopotamus optimization (IHO) algorithm. Ma et al. [16] introduced the social network search (SNS) algorithm to enhance LightGBM performance by optimizing its hyperparameters.

Gaopale et al. [17] proposed and evaluated a Sparrow Search Algorithm (SSA-ANN) to optimize the ANN. Their hybrid approach achieved superior performance compared with the ANN and ANN-GA. Ge et al. [18] developed a stacking ensemble framework that integrated three gradient-boosted tree models: LightGBM, XGBoost, and CatBoost. They optimized the hyperparameters of these data-driven models using Bayesian Optimization (BO), GWO, and PSO algorithms. Fissha et al. [19] investigated the predictive capabilities of Gaussian process regression (GPR), decision tree (DT), and SVR. They reported that the

GPR model achieved the highest accuracy, with an exceptionally low error rate. Ogunsola et al. [20] evaluated ANN-LM, an adaptive neuro-fuzzy inference system (ANFIS), GPR, and DT. They improved the predictive capability of the GPR approach, with its integration with BO (GPR-BO) and the differential evolution (DE) algorithm (GPR-DE). Their results indicated that the GPR-BO model outperformed the other standalone and hybrid approaches. Hasanipanah et al. [21] developed an SVR model to predict blast-induced ground vibrations at the Bakhtiari Dam in Iran. The comparative analysis demonstrated that the SVR method achieved significantly higher prediction accuracy and generalization performance than conventional empirical approaches.

Armaghani et al. [22] compared the performance of the hybrid ANN-PSO algorithm with that of a standard ANN. The findings confirmed the superior accuracy of ANN-PSO compared to the ANN. Saubi et al. [23] developed an ANN-GA model for the Jwaneng mine in China. The accuracy of the ANN-GA approach was higher than that of other hybrid models, such as ANN-PSO, ANN-ABC, and ANN-ICA. Zhou et al. [24] used Gene Expression Programming (GEP) and also implemented Monte Carlo simulations to assess and control the probabilistic risks of blasting operations. They demonstrated that combining GEP with the Monte Carlo approach enables engineers to accurately determine safe blasting perimeters. Bayat et al. [25] developed an ANN-based approach combined with the firefly algorithm (FA) and concluded that ANN-FA provided a powerful and practical methodology for blasting-induced ground vibrations. Behzadafshar et al. [26] developed hybrid ANN models with three variants of the imperialist competitive algorithm (ICA) to predict blast-induced ground vibrations at the Shur River dam in Iran. They constructed ICA-linear, ICA-power, and ICA-quadratic models and rigorously evaluated them against standard ANNs and empirical models. Their findings revealed that the ICA-based models consistently exhibited greater generalization performance than both the ANN and empirical models. Particularly, the ICA-linear model proved to be the most accurate predictive tool.

The reported findings highlighted the capability of ML models to model *PPV*. Despite these advancements, many studies used black-box ML models, including SVR and the ANN, to estimate *PPV*. Therefore, the applications of novel ML methods for predicting blasting vibration continue to evolve. To develop reliable, accurate predictive models that support safer, more efficient blasting operations is an interesting subject and essential for engineers. Moreover, the results of the aforementioned studies also provide a strong basis for extending the analysis to novel white-box ML methods for modeling *PPV*.

Research Gap and Objectives

Although the literature review has successfully demonstrated the superiority of ML for predicting *PPV*, several critical research gaps remain. While highly accurate and widely used ML approaches, such as ANNs and SVR techniques, are used to estimate *PPV*, they are fundamentally black-box ML models. These methods do not yield explicit mathematical equations that are easy to use in practical engineering applications [27]. Recent studies highlight the potential and capability of white-box ML algorithms, such as Multi-Expression Programming (MEP), Gene Expression Programming (GEP), Multivariate Adaptive Regression Splines (MARS), and Stronger Variable Creator Machines (SVCMs), which have gained considerable attention due to their robustness, generalization ability, and improved predictive performance for modeling complex problems in civil engineering projects [28–34]. Despite the promising results of these white-box ML methods, limited research has examined their use, which addresses the drawback of black-box ML methods by generating explicit mathematical formulas for estimating *PPV* based on effective variables. Furthermore, there is a lack of a comprehensive assessment of these white-box

algorithms and of comparisons between them. Therefore, the present study conducted a direct performance comparison of these white-box algorithms under identical conditions and with the same input and output variables to provide a fair assessment of white-box methods. These algorithms provide general mathematical equations for estimating *PPV*. Quantitative assessments of generated formulas were conducted employing commonly used statistical metrics, including the root mean square error (RMSE), Mean Absolute Error (MAE), and correlation coefficient (CC). In addition, a qualitative evaluation of the performance of ML methods was conducted using graphical plots, including Taylor, violin, and scatter diagrams, to present the outcomes. Finally, the study provided explicit mathematical equations for implementation in the Excel spreadsheet, which is likely easy to use for future applications of the proposed models to estimate PPV. To the best of the authors' knowledge, this is the first study to simultaneously develop explicit MEP, GEP, MARS, and SVCM equations for *PPV* prediction and to translate all models into spreadsheet-compatible forms for direct engineering application.

The remainder of this manuscript is organized as follows. Section 2 describes the dataset and presents the theoretical background of the selected machine learning algorithms. Section 3 introduces the statistical performance metrics used to evaluate the models. Section 4 details the model development procedures and presents the explicit mathematical formulations derived from the developed models. Section 5 provides a comprehensive quantitative and visual assessment of the results, supported by rigorous five-fold cross-validation and parametric analyses. Section 6 compares the outcomes of the present study with those reported in previously published *PPV* prediction research. Section 7 discusses the limitations of the current research and outlines recommendations for future work. Finally, Section 8 presents the main conclusions of the study.

2. Materials and Methods

2.1. Dataset Used

The present study used a reliable, previously published dataset from Iphar et al. [7]. This dataset contains field measurements of blast-induced ground vibration records originally measured at an open-pit magnesite mine located west of Eskisehir, Turkey. This mine was established in 1963 and has been one of Turkey's major magnesite producers, extracting about 300,000 tons of raw magnesite annually from its open-pit mines. A total of 44 blasting events were recorded. For each blasting event, two primary input parameters were recorded: *D*, denoting the direct distance (in meters) between the blast face and the monitoring station, and *W*, representing the maximum explosive charge weight per delay (in kilograms). These two variables are considered the most influential and controllable factors affecting ground vibration and are widely used in traditional and ML-based prediction models for modeling *PPV*, which represents the intensity of ground vibration in millimeters per second (mm/s). The functional form of input/output variables can be written as follows:

$$PPV = f(D, W) \quad (1)$$

The available dataset covers a wide range of blasting conditions. The input variables, including *D* and *W*, varied between 90 m and 1280 m, and between 74.5 kg and 1208 kg, respectively. The *PPV* values ranged between 2.03 mm/s and 39.6 mm/s. These dataset variations capture both near-field and far-field vibration behaviors, as well as different blasting intensities. Therefore, the dataset is suitable for ML modeling because it represents real-world blasting operations and includes key influencing variables (*D* and *W*) that have a measurable impact on *PPV*. The histogram of the involved variables is shown in Figure 1.

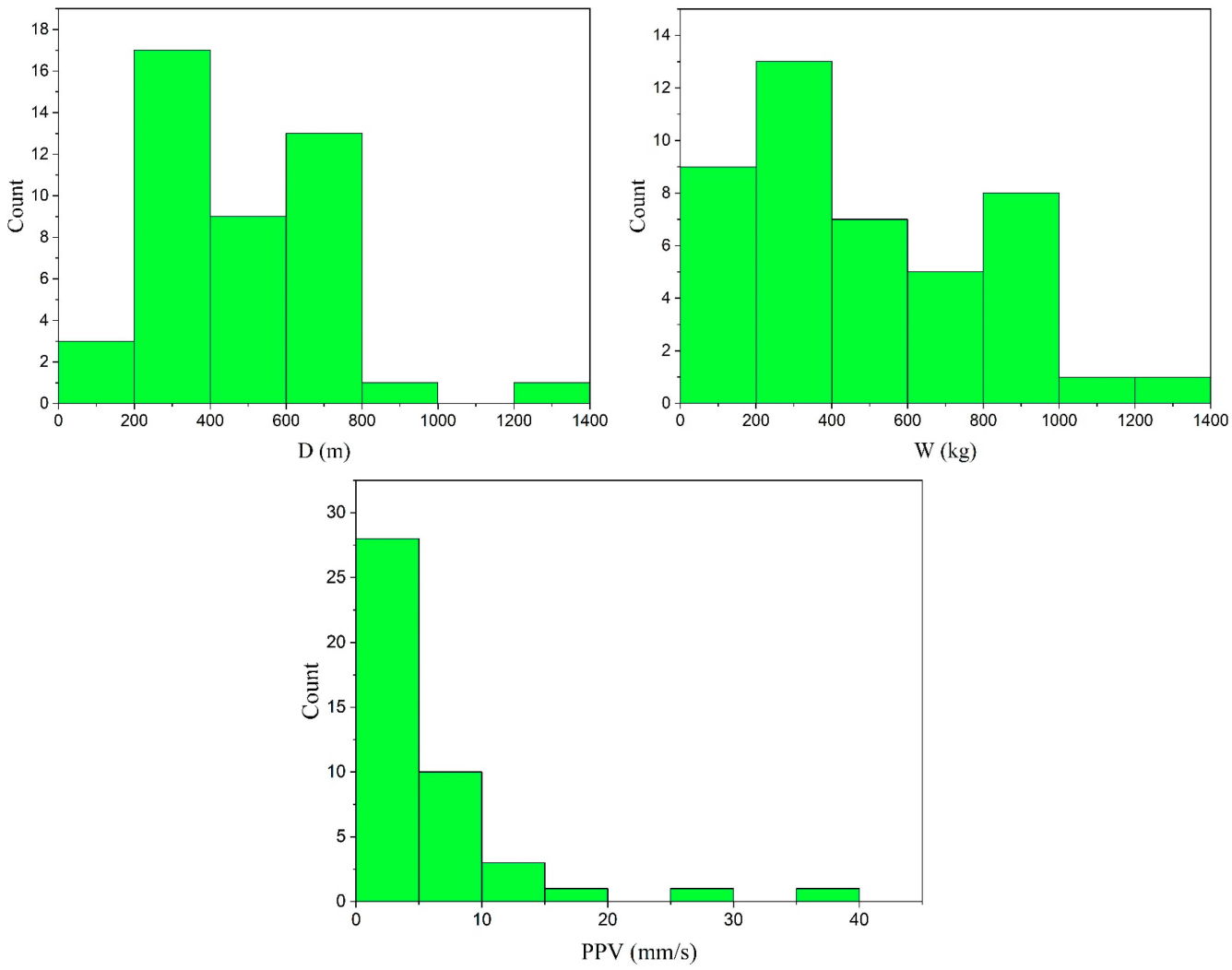


Figure 1. Histogram of involved variables [10].

It is important to note that in geotechnical and blasting engineering applications, especially in blast-induced ground vibration studies, high-quality field-measured datasets are inherently limited due to the high cost, operational constraints, and safety considerations associated with controlled blasting experiments. In addition, the rock mass properties naturally influence wave propagation, and geological conditions often remain relatively uniform within specific mining sites. Therefore, this study employs the two most controllable operational parameters, *D* and *W*, to estimate *PPV* at an open-pit magnesite mine, as is well-documented in previous research.

2.2. White-Box ML Methods

This study used four white-box ML methods to model PPV, including MEP, GEP, MARS, and SVCM approaches. The main feature of these methods is that they can produce explicit formulas for approximating the output parameter without assuming a predefined regression form. In addition, the derived formulas are attractive for engineering practice because these models can be implemented directly in spreadsheets or design sheets. The brief description of the main concept of each proposed ML technique is as follows.

2.2.1. MEP

MEP is an evolutionary symbolic regression technique developed as an extension of Genetic Programming (GP) for deriving explicit mathematical models. In MEP, candidate

solutions are encoded as linear chromosomes composed of multiple genes, where each gene may represent either a terminal variable or a mathematical operator. Unlike conventional GP, MEP can store multiple mathematical expressions within a single chromosome, and the most suitable expression is selected based on a predefined fitness criterion. This structure improves both the search capability and computational efficiency of the algorithm in identifying optimal nonlinear relationships between input and output variables [35].

2.2.2. MARS

The MARS modeling technique generates a regression model through two main stages: forward selection and backward pruning. In the forward stage, a wide array of basis functions is created to maximize accuracy on the model. Then, in the backward-pruning stage, the least significant basis functions are removed using the Generalized Cross-Validation (GCV) criterion to avoid model overfitting [36]. Finally, the MARS algorithm produces an explicit mathematical equation that represents a linear combination of the selected basis functions to estimate the output parameter.

2.2.3. GEP

GEP is an evolutionary computation and symbolic regression technique developed based on the principles of GA and GP. The GEP algorithm generates mathematical models by evolving populations of chromosomes through biological-inspired operators such as mutation, recombination, and selection. In GEP, chromosomes are encoded as linear strings of genes, which are subsequently expressed as nonlinear mathematical expression trees, called sub-expression trees (Sub-ETs). This structure enables the algorithm to automatically discover complex functional relationships between input and output variables without requiring predefined mathematical assumptions [37].

2.2.4. SVCM

The SVCM is an innovative ML algorithm that enhances predictive accuracy by systematically generating and screening transformed mathematical features. The algorithm isolates the most significant nonlinear transformations through a rigorous three-step procedure. First, in the feature generation phase, synthetic variables are created by applying diverse mathematical functions to the original inputs. Next, during target correlation screening, only newly generated variables that exhibit a higher coefficient of determination (R^2) with the target variable than their parent variables are retained. Finally, collinearity filtering prevents multicollinearity by finalizing only those variables that have an $R^2 < 0.5$ with their original parent variables. By successfully isolating features with high target correlation but low parent collinearity, the SVCM utilizes multiple linear regression to combine these optimized parameters into an explicit predictive equation [27].

It is important to distinguish the SVCM method from conventional feature engineering approaches such as polynomial feature expansion and kernel-based methods. Traditional polynomial feature generation typically constructs higher-order terms (e.g., x^2, xy, y^2) by exhaustively combining input variables up to a predefined degree. Although this approach can capture nonlinear relationships, it often results in a rapid increase in the number of generated features and introduces substantial multicollinearity, which may adversely affect model stability and interpretability. In contrast, the SVCM employs a more systematic and adaptive feature-generation framework. Rather than relying solely on polynomial transformations, it utilizes a diverse set of nonlinear operators, including logarithmic, exponential, and trigonometric functions, to generate candidate variables. More importantly, the SVCM incorporates a multi-stage filtering mechanism that eliminates redundant features by removing newly generated variables that are highly correlated with their parent variables. As a result, only informative and non-redundant variables are retained for subsequent model

development, reducing multicollinearity while improving model robustness. This adaptive mechanism enables the SVCM to efficiently capture complex nonlinear relationships without generating an excessive number of unnecessary variables.

Furthermore, the SVCM fundamentally differs from kernel-based methods, such as those employed in support vector machines (SVMs). Kernel methods implicitly map input variables into high-dimensional feature spaces, where nonlinear relationships become linearly separable. While this mapping is generally highly accurate, it operates entirely as a black box, obscuring the underlying mathematical relationships. In contrast, the SVCM explicitly generates, evaluates, and retains transformed variables, allowing the final model to be expressed as a transparent closed-form mathematical equation. This characteristic significantly enhances reproducibility and practical applicability in engineering practice.

3. Performance Metrics

The following statistical criteria were used to evaluate the ML model. The performance of the predictive models is evaluated using the correlation coefficient (CC), root mean square error (RMSE), and Mean Absolute Error (MAE). The equations of statistical criteria are defined as follows:

$$CC = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\widehat{y}_i - \bar{\widehat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(\widehat{y}_i - \bar{\widehat{y}})^2}} \quad (2)$$

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

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y}_i| \quad (4)$$

where y_i represents the actual (measured) value, $\widehat{y}_i$ represents the predicted value, $\bar{y}$ is the mean of the actual values, $\bar{\widehat{y}}$ is the mean of the predicted values, and n is the total number of data samples.

4. Modeling PPV Using White-Box ML Methods

For the modeling process, the dataset was randomly partitioned into two subsets: 80% of the data was used for model development, while the remaining 20% was reserved for testing of ML models. In addition to ensuring rigorous, robust evaluation of the developed white-box models, a k-fold cross-validation strategy was implemented. Therefore, the five-fold cross-validation approach was utilized for this purpose. In this approach, the entire dataset is randomly partitioned into five equally sized subsets, which are called folds. Next, the white-box models are trained on four folds, each containing approximately 35 samples, and validated on the remaining unseen fold, which contains approximately nine samples. This process is repeated five times to ensure that each fold is used exactly once as the testing set. Finally, the statistical performance metrics are calculated across all five iterations. This concept is a highly reliable and unbiased assessment of the predictive ML models.

4.1. MEP Model Development

The MEP algorithm was implemented to estimate *PPV* using two input variables, *D* and *W*. Unlike the standard GP approach, which encodes a single solution per chromosome, MEP encodes multiple mathematical expressions within a single linear chromosome. This concept enables a more comprehensive exploration of the search space, significantly reducing the likelihood of premature convergence. The model evolved over successive generations, selecting the optimal combination of arithmetic operations, trigonometric functions, and logarithmic scaling to minimize the predictive error. The development of

the MEP model requires adjusting several internal evolutionary parameters that influence the predictive performance of the symbolic expressions generated by the MEP algorithm. The population size represents the total number of candidate programs generated in each evolutionary cycle. Genetic diversity within the MEP model is controlled by evolutionary operators. The crossover and mutation probabilities control the likelihood of applying crossover and mutation operators during the evolutionary process, respectively. In this study, the uniform crossover strategy was employed, in which genes are randomly exchanged between parent chromosomes to generate new offspring solutions. In addition, the function set specifies the mathematical operators and functions used to construct the model's symbolic expressions. The optimal MEP parameter configuration was determined through trial and error to achieve the highest prediction accuracy for PPV estimation. Based on this optimization process, the final MEP model was developed with a population size of 250, a code length of 500, a uniform crossover, a crossover probability of 0.9, and a mutation probability of 0.01. The MEP algorithm utilized a comprehensive function set consisting of basic arithmetic operators ($+$, $-$, $\times$, $\div$) and advanced mathematical functions ($Sqrt$, Exp, Ln, Sin, Cos, Atan, $1/x$) to construct the mathematical expressions. The MEP algorithm provided an explicit expression for PPV estimation as follows:

$$\begin{aligned} PPV = \quad & \cos\left(\tfrac{1}{K_9}\right) + \tfrac{K_3 K_8}{K_3 - W\cos(W) + D - 1/K_9} + \cos\left(\tfrac{D}{K_3} + \cos\left(2D^2\right)\right) \\ & + \exp(K_8) + K_8 + \exp(K_6) + \arctan\left(\tfrac{1}{K_2}\right) - K_6^2 \sin(K_{11}) \\ & + 2K_{10} - \tfrac{1}{K_5} + \cos(D + \cos D) \cdot \sin\left(-\ln\left(K_4\sqrt{2D^2}\right)\right) \end{aligned} \tag{5}$$

where

$$K_1 = W - \sin(D)$$

$$K_2 = \arctan\left(K_1\sqrt{\ln(D^2)} - \sqrt{2D^2}\right)$$

$$K_3 = W + K_2$$

$$K_4 = 2D^2 + W - \frac{1}{W\sin(D)}$$

$$K_5 = \sin\left(2\sqrt{W} - K_1\right) + W\sin(D) - \sqrt{\ln(D^2)} + W$$

$$K_6 = \frac{K_3}{D - \sqrt{W}}$$

$$K_7 = \frac{W(\cos D - D - \cos W)}{D}$$

$$K_8 = \cos(K_7)$$

$$K_9 = \sin(\ln K_4)$$

$$K_{10} = \cos\left(\sqrt{W} - K_1\right) - K_9$$

$$K_{11} = \sin\left(W + \frac{1}{\ln K_5}\right) + \ln K_4 - \frac{1}{W\sin(D)}$$

As observed in the MEP formulation, the nested and sequential nature of the MEP evolutionary process makes the explicit mathematical formulation of PPV appear structurally complex. However, the complexity of this MEP equation is mathematically justified. The traditional models used to estimate PPV rely on simplistic power laws. However, the propagation of blast waves involves complex inelastic damping and depends on the distance of the blast. The nested exponential and trigonometric functions generated by the MEP algorithm account for the nonlinear relationships among the involved variables. In

addition, the MEP equation is visually dense and relies on a limited number of numerical constants (low degrees of freedom) compared to the ANN model's black-box weight matrices. Therefore, the MEP model is mathematically appropriate for limited datasets. Furthermore, by embedding the MEP formulation directly into a standard Excel spreadsheet function, as provided in the Supplementary Materials, we ensure the simple, immediate, practical application of the derived MEP model for estimating *PPV* in engineering practice.

4.2. GEP Model Development

GEP is an evolutionary algorithm that evolves computer programs of varying sizes and shapes to discover the optimal mathematical relationship between input and output variables. The distinct advantage of GEP is its ability to produce explicit mathematical equations, making it highly practical for field engineers. The evolutionary process was driven by a fitness function. Through iterative processes of mutation, transposition, and recombination, the algorithm converged on an optimal chromosomal architecture composed of various sub-expression trees (Sub-ETs). For modeling *PPV* and accurately capturing the highly nonlinear relationships between involved variables, the fundamental arithmetic operators ($+$, $-$, $\times$, $\div$) were combined with advanced nonlinear functions (x^2, x^3, $\sqrt{x}$, $x^{1/3}$, $\exp(x)$, $\ln(x)$, $\sin(x)$, $\cos(x)$, and $\arctan(x)$).

The evolutionary process was conducted by minimizing the RMSE, which served as the designated fitness function. The GEP algorithm used a population of 30 individuals evolving over 100,000 generations. Each candidate chromosome was structured with three genes, utilizing a head length of 7. The Sub-ETs generated by these three genes were ultimately aggregated using an addition ($+$) linking function. To facilitate genetic diversity and optimize the search space, several evolutionary operators were applied, including mutation (probability = 0.044), one-point crossover (probability = 0.3), two-point crossover (probability = 0.3), and gene recombination (probability = 0.1). The final optimized GEP model, with three resulting Sub-ETs, is shown in Figure 2.

As illustrated in the provided Sub-ETs, the translation into explicit mathematical equations can be expressed as follows for estimation of PPV:

$$PPV = \text{Sub-ET}_1 + \text{Sub-ET}_2 + \text{Sub-ET}_3 \tag{6}$$

where the individual sub-expression trees are mathematically expressed as:

$$\text{Sub-ET}_1 = \left(\exp\left[(\arctan(\cos(D + W + G_1C_1)))^3\right]\right)^3$$

$$\text{Sub-ET}_2 = \left(\exp\left[\cos\left(\arctan\left(\frac{D}{W}\right) + \sqrt[3]{D}\right)\right]\right)^2$$

$$\text{Sub-ET}_3 = \frac{2 \cdot G_3C_1 \cdot W - \exp(G_3C_0)}{\exp\left(\sqrt[3]{D}\right)}$$

The terms G_1C_1, G_3C_0, and G_3C_1 represent the specific numerical constants randomly generated and optimized by the GEP algorithm during the evolutionary process for Sub-ETs. The constant values related to Sub-ETs are $G_1C_1 = -1.247253$, $G_3C_0 = 6.205597$, and $G_3C_1 = 6.467834$. The resulting GEP formulation demonstrates the ability of evolutionary algorithms to identify highly nonlinear and complex relationships among blasting parameters.

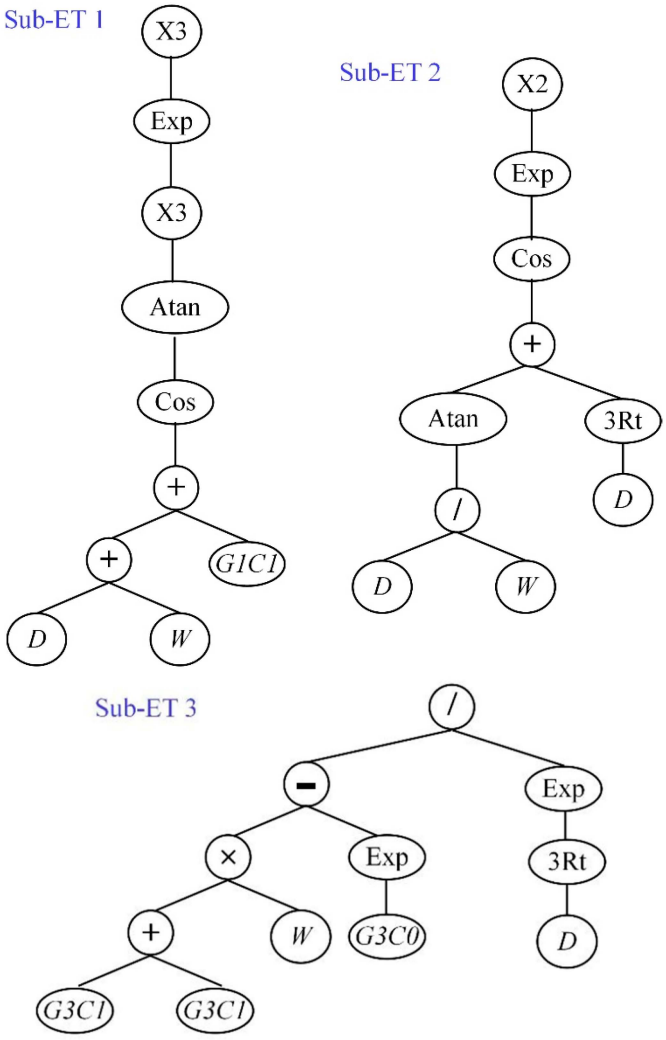


Figure 2. Sub-ETs for estimation of *PPV*.

4.3. MARS Model Development

MARS constructs the prediction equation for PPV by combining a set of piecewise basis functions and accounting for localized interactions among the input variables, including W and D. This approach increases the flexibility of the MARS equation, in contrast to conventional regression models that impose a fixed global functional form. In the present model, five basis functions were generated to describe the nonlinear relationship between the blasting parameters and PPV. The basis functions (*BFs*) of the MARS expression are obtained as follows:

$$BF_1 = \max(0, 269.72 - W)$$

$$BF_2 = \max(0, W - 74.5\ )\max(0, 320 - D)$$

$$BF_3 = \max(0, W - 377.5\ )\max(0, 730 - D)$$

$$BF_4 = \max(0, 377.5 - W)\max(0, 730 - D)$$

$$BF_5 = \max(0, W - 486\ )\max(0, 730 - D)$$

Using these basis functions, the final MARS predictive equation for *PPV* was formulated as:

$$PPV = 2.46557 - 0.107455\ BF_1 + 0.00056262\ BF_2 + 0.000118257\ BF_3 + 0.000143226\ BF_4 - 0.000121683\ BF_5 \quad (7)$$

By substituting these basis functions into the final expression (Equation (7)), the *PPV* values were calculated.

4.4. SVCM Model Development

In this study, the SVCM approach was used to develop an explicit predictive model to estimate PPV induced by blasting operations. The SVCM is a recently introduced ML technique designed to improve predictive performance by generating stronger, more informative variables from the original input variables. Unlike conventional regression models that directly use the original variables, the SVCM method enhances predictive capability by constructing transformed variables using various mathematical operators and then selecting the most effective variables based on statistical correlation criteria. The method aims to generate variables that exhibit stronger relationships with the target output while maintaining lower redundancy with the original variables. In the present study, the original blasting parameters included *D* and *W*. Following SVCM feature engineering and selection, the final predictive equation for *PPV* was obtained via linear regression. Based on the SVCM algorithm, the explicit predictive equation for *PPV* is constructed as a linear combination of the optimal transformed variables. The mathematical expression derived from the SVCM algorithm is defined as follows:

$$\begin{aligned} PPV = \; & 1841.640224 - 0.55894472D + 0.006244063W \\ & + 64.55362125\sqrt{|D|} - 476.2331224\ln(D) \\ & - \tfrac{24{,}180.43771}{D} - 8.34932 \times 10^{-5}W^2 \\ & + 5.24808 \times 10^{-8}W^3 - 0.929882359\sin(W) \\ & + 16.85197095\left(\tfrac{W}{D}\right) \end{aligned} \tag{8}$$

5. Results and Discussions

5.1. Quantitative Evaluation of White-Box ML Models for Estimation of PPV

The statistical performance of the four developed ML models was evaluated on the training, testing, and total datasets for *PPV* prediction, and the results are summarized in Table 1.

Table 1. The values of statistical metrics for estimation of *PPV* using MEP, GEP, MARS, and SVCM models using training, testing, and total datasets.

Model	Data Category	CC	RMSE	MAE
MEP	Training	0.995155	0.80736	0.617947
	Testing	0.991083	0.843628	0.683506
	Total	0.993177	0.836337	0.670096
GEP	Training	0.969092	1.95312	1.37739
	Testing	0.882016	1.985978	1.705516
	Total	0.963658	1.959886	1.444507
MARS	Training	0.980115	1.470885	1.113374
	Testing	0.956695	1.530882	1.121231
	Total	0.977011	1.518803	1.11498
SVCM	Training	0.953391	1.689337	1.36766
	Testing	0.912151	2.322062	1.518289
	Total	0.950785	2.207445	1.487479

As observed in Table 1, the MEP model exhibited the highest predictive accuracy. The MEP model yielded the highest CC values of 0.9952, 0.9911, and 0.9932 for the training, testing, and total datasets, respectively, along with the lowest RMSE values of 0.8074,

0.8436, and 0.8363, and the lowest MAE values of 0.6179, 0.6835, and 0.6701. These results indicate the MEP model's high accuracy in predicting PPV. The MARS model achieved CC values of 0.9801 and 0.9567 for the training and testing datasets, respectively. Moreover, the RMSE values are 1.4709 for the training and 1.5309 for the testing dataset. For the total dataset, the MARS model produced CC, RMSE, and MAE values of 0.9770, 1.5188, and 1.1150, respectively. Its RMSE and MAE values were also relatively low compared with those of the GEP and SVCM models, confirming that MARS provided accurate predictions, although its accuracy remained lower than that of the MEP model. In contrast, the GEP and SVCM models exhibited weaker performance. The GEP model produced CC values of 0.9691, 0.8820, and 0.9637, while the SVCM model yielded CC values of 0.9534, 0.9122, and 0.9508 for the training, testing, and total datasets, respectively. In addition, both models showed higher RMSE and MAE values than MEP and MARS, indicating their lower prediction accuracy. Overall, the statistical metrics indicated that the MEP model was the most accurate for estimating PPV, followed by MARS, whereas GEP and SVCM were less accurate. The superior performance of the MEP and MARS models indicated their strong ability to capture the nonlinear relationship between the input parameters and PPV values. The obtained results demonstrate that all developed models provided acceptable predictive capability.

The superior performance of the MEP model depends on its ability to evolve highly flexible symbolic mathematical structures that represent complex nonlinear interactions in the blasting dataset. The strong performance of the MARS approach demonstrates the effectiveness of piecewise linear regression methods for modeling PPV while preserving relatively simple mathematical formulations. The GEP model showed acceptable predictive performance. Although its error indices were slightly higher than those of the MARS and MEP models, the GEP approach still successfully modeled the nonlinear relationship between PPV and other variables through symbolic evolutionary regression. The SVCM model achieved the lowest predictive accuracy among the ML models. Nevertheless, the SVCM approach still provided reasonable results in predicting *PPV* values.

To eliminate uncertainty and random fluctuations associated with relatively small test sets, the performance of the MEP, MARS, GEP, and SVCM models was evaluated using five-fold cross-validation. The five-fold CV results related to each fold are presented in Figure 3.

As observed, the MEP model achieved the highest average CC of 0.9334 and the lowest average RMSE of = 2.2399 across all folds. The MARS model maintained its rank as the second-best algorithm, with an average CC of 0.9163 and an average RMSE = 3.0047, followed by the SVCM, with an average CC = 0.9132 and an average RMSE of 2.8442, and the GEP, with an average CC of 0.8891 and an average RMSE of 3.1688, which exhibited lower predictive accuracies. The results of the five-fold cross-validation showed that the developed white-box ML models maintained satisfactory performance across different data partitions. The results of cross-validation demonstrate that the performance of the proposed models is not dependent on a single random data split and that the developed models exhibit acceptable generalization capability for PPV prediction. Moreover, the cross-validation outcomes were generally consistent with the results obtained from the conventional training-testing split method.

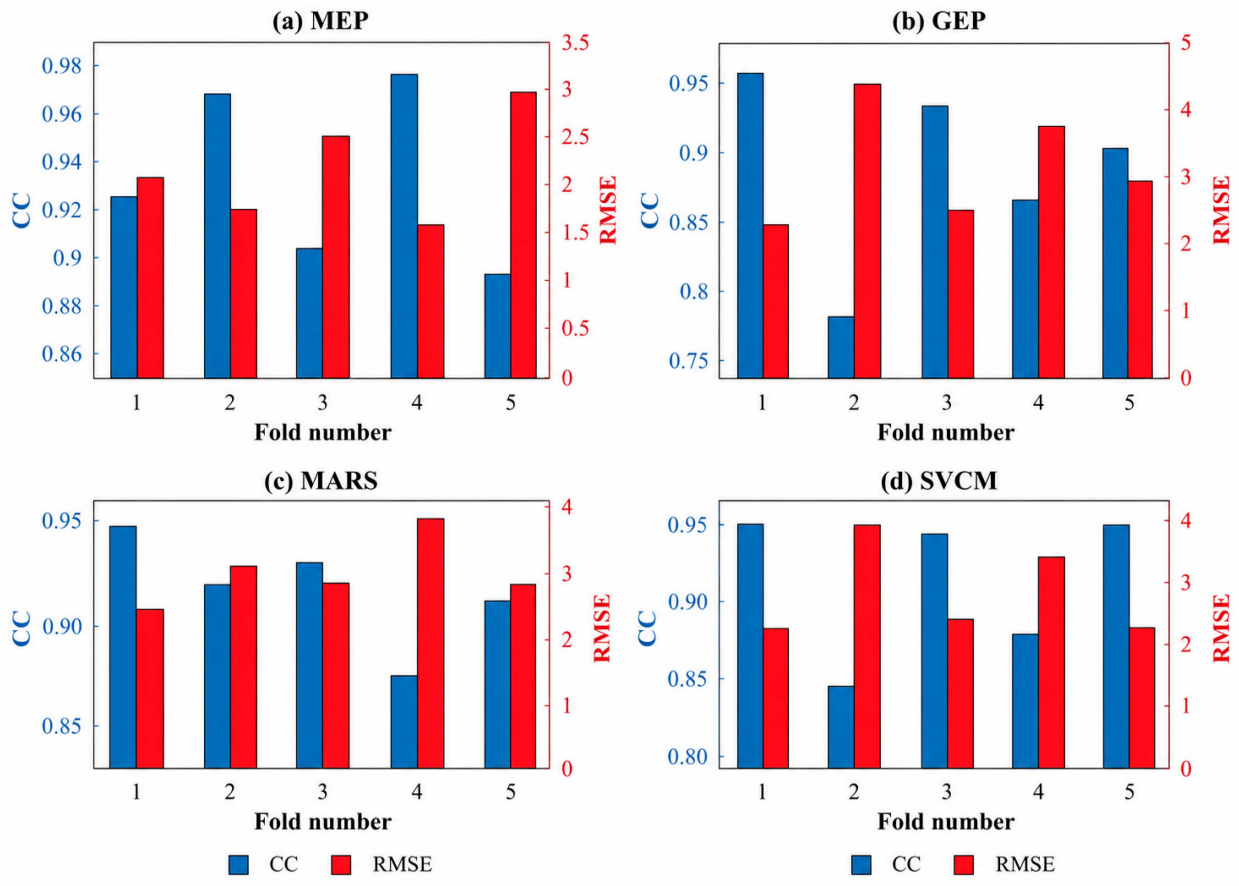


Figure 3. Results of five-fold cross-validation for (**a**) MEP; (**b**) GEP; (**c**) MARS; (**d**) SVCM.

5.2. Visualization Assessment of White-Box ML Models

To comprehensively evaluate the developed ML models, a graphical assessment of the predicted PPV values is presented using scatter, Taylor, and violin plots for the entire dataset. The scatter plots used to estimate PPV values are shown in Figure 4.

The scatter plots in Figure 4 compare measured PPV values with corresponding predictions from the MEP, GEP, MARS, and SVCM models. The predictive accuracy of each model can be visually assessed by the closeness of the scatter points to the ideal line. The MEP results (Figure 4a) demonstrate the highest predictive capability. The blue data points are tightly clustered along the ideal line across the entire data range, with minimal deviation even at higher PPV values. The results of the MARS model, shown as green data points (Figure 4c), align well with the ideal line. However, compared to the MEP model, there is slightly more scatter around the ideal line. The results of the GEP model (Figure 4b), shown as red data points, exhibit greater dispersion around the ideal line than the MARS and MEP models. The SVCM model (Figure 4d) exhibits the widest dispersion among the ML algorithms. Therefore, the scatter plots provide strong visual evidence that the MEP model achieved the highest predictive accuracy. In contrast, the SVCM model exhibited comparatively lower predictive performance.

The Taylor diagram simultaneously considers the CC, standard deviation (SD), and root mean square deviation (RMSD) between the predicted and observed PPV values. In a Taylor diagram, the optimal model is the one closest to the reference point representing the measured PPV data, as it exhibits high correlation, low RMSD, and a standard deviation similar to that of the observed dataset. Figure 5 illustrates the Taylor diagram for a comprehensive assessment of the accuracy of the developed ML models in estimating PPV values.

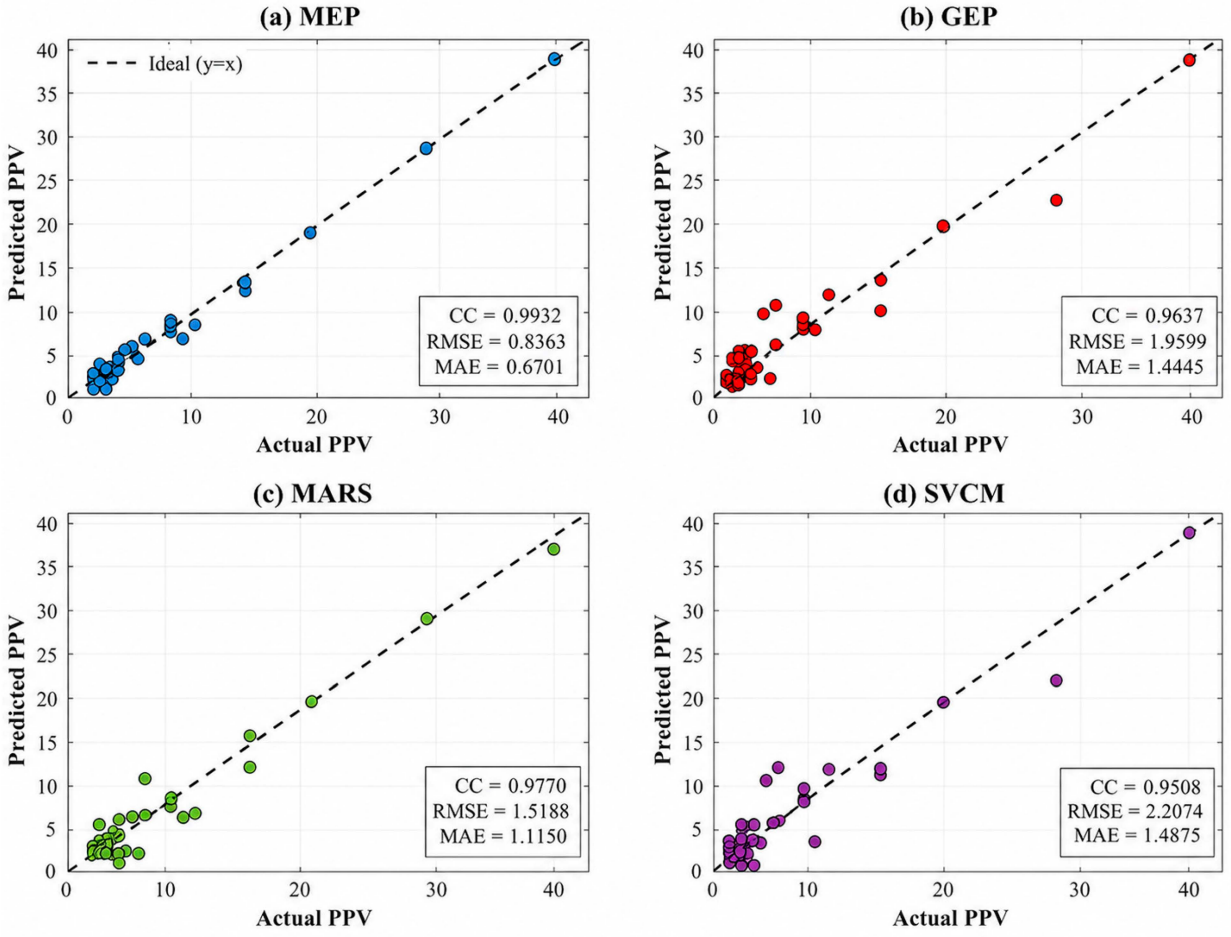


Figure 4. Scatter plots for estimation of PPV values using white-box ML methods: (**a**) MEP, (**b**) GEP, (**c**) MARS, (**d**) SVCM.

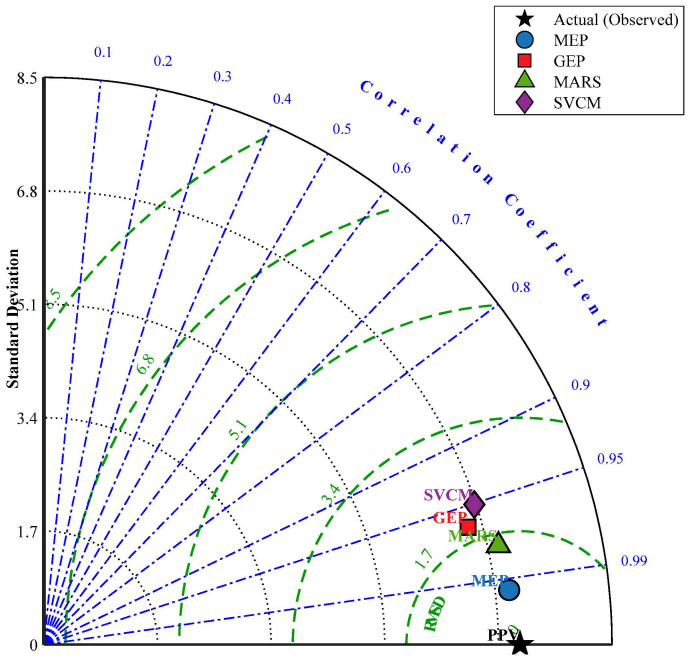


Figure 5. Taylor diagram for estimation of *PPV*.

As observed in Figure 5, the spatial distribution of the ML models on the Taylor diagram clearly shows that the MEP model has the highest accuracy, followed by MARS,

GEP, and SVCM. The MEP model is positioned closest to the actual observation point. The MARS model demonstrates competitive performance with the MEP model. It is positioned slightly farther from the reference point than the MEP model. The GEP model is farther from the actual point than MEP and MARS. The SVCM is positioned furthest from the actual data reference point. Consequently, the SVCM achieves lower accuracy than other ML models in estimating *PPV*.

The violin plot provides a comprehensive visual statistical comparison of the predicted PPV against the actual measured values. Figure 6 displays a violin plot of the distribution of outcomes from ML models estimating *PPV*.

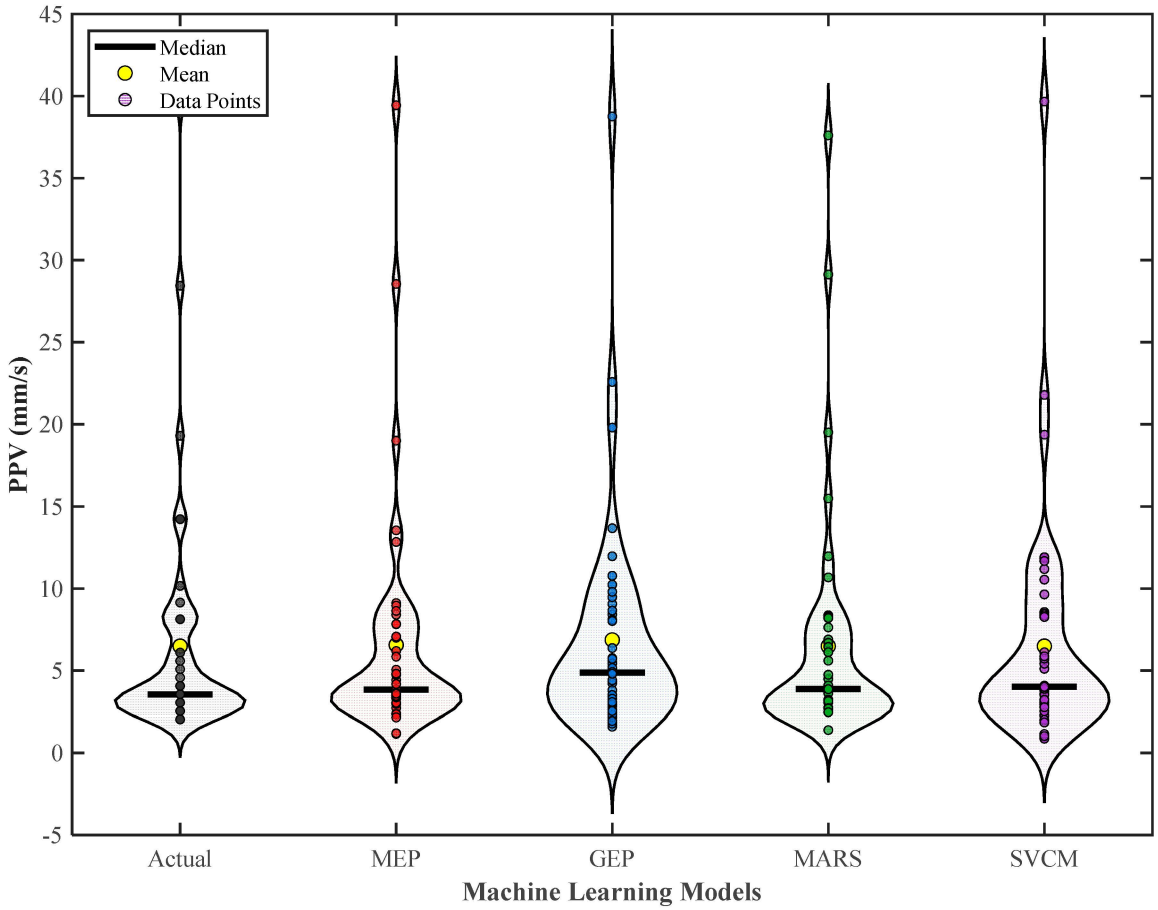


Figure 6. Violin plots comparing the distribution of actual and predicted PPV values across the different ML models. The solid, thick black line inside each violin indicates the median value; the yellow circle represents the mean value; and the scatter points correspond to individual observations.

As shown in Figure 6, the median value is represented by a thick black horizontal line, the mean value is indicated by a yellow marker, and the individual data points are displayed as scatter points within each violin distribution. It is clear that, among the developed ML models, the MEP model demonstrates the closest agreement with the actual PPV distribution. The shape of the MEP violin is highly consistent with the actual data, particularly in the lower and moderate PPV ranges where most observations are concentrated. In addition, the median and mean values of the MEP predictions are very close to those of the actual data, indicating minimal distributional bias. The dispersion of the MEP predictions is also relatively similar to that of the actual distribution, confirming its superior predictive capability. The MARS model also shows good agreement with the actual PPV distribution. This indicates that the MARS model can capture the dominant nonlinear behavior of PPV with acceptable accuracy, although its predictive variability is slightly

lower than that of the MEP model. The GEP model exhibits a broader distribution, with greater dispersion and higher variability, than the actual data. The violin shape indicates that the model tends to produce predictions with greater spread, especially in intermediate and high PPV regions. Although the GEP model captures the general trend in the data, the increased variability suggests reduced predictive accuracy relative to MEP and MARS. The SVCM model demonstrates the weakest agreement with the actual PPV distribution. Its violin plot reveals greater deviations in both the distribution's shape and spread relative to the actual data. The model predictions appear more dispersed and less concentrated around the dominant PPV ranges, resulting in a lower estimated PPV accuracy compared to others.

5.3. Parametric Study

The data-driven and symbolic ML algorithms operate primarily as data-driven curve-fitting tools. The primary advantage of white-box ML models lies in their ability to capture complex nonlinear interactions while maintaining explicit mathematical forms that can be used directly to estimate *PPV*. Unlike black-box ML models, such as ANNs and SVRs, the proposed white-box ML models provide fully explicit mathematical expressions, allowing engineers to inspect the influence of input variables and directly implement them in spreadsheet environments. A parametric study can simulate the behavior of the developed white-box models with respect to the input variables and compare the predicted trends with the measured *PPV* values.

A parametric study systematically investigates the individual effect of each input variable on the PPV. Therefore, each input variable varies over its relevant experimental range, and the trend in observed PPV values is compared with that of each ML model. In this analysis, each independent variable (*W* and *D*) was systematically varied across its full experimental range, and the resulting *PPV* predictions from all four white-box models were plotted against the measured field data. Figure 7a illustrates the results of the parametric study for the evaluated white-box models, isolating the effect of the distance variable (*D*) on the predicted PPV. Similarly, Figure 7b demonstrates the isolated effect of the explosive charge weight variable (*W*) on *PPV*.

(**a**)

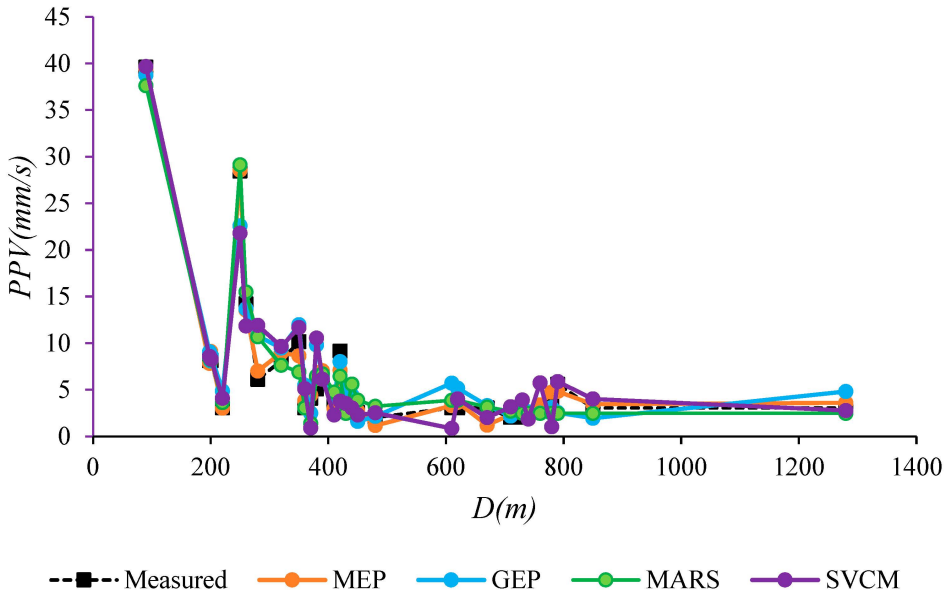


Figure 7. *Cont.*

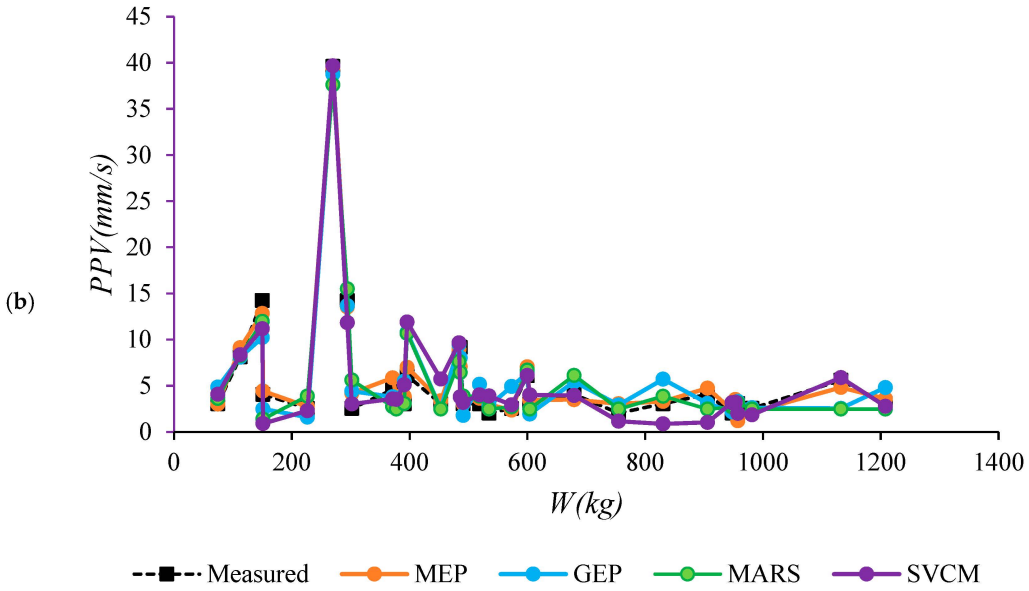


Figure 7. The variation of measured and predicted PPV values with changes in (**a**) D and (**b**) W variables.

As illustrated in Figure 7, the comparison between the estimated values from the ML models and the observed PPV values across the range of each independent variable indicates that the ML models proposed in this study agree well with field observation data. Therefore, the results demonstrated that the ML models correctly estimate the trend in PPV change by considering each independent variable and aligning with the actual values. The close agreement between the measured and estimated *PPV* values across the ranges of *D* and *W* demonstrates that the proposed models not only fit the available data but also reproduce the experimental behavior for each variable.

6. Compared with Previous Research

To further evaluate the efficacy of the proposed ML models, we directly compared our findings with those reported by Iphar et al. [7], who previously investigated PPV prediction using regression and ANFIS models on the same dataset of 44 blasting records. The previous study yielded relatively weak prediction performance for the conventional regression model compared to the ANFIS model, with CC, RMSE, and MAE values of 0.9008, 3.1914, and 1.9672, respectively. In contrast, all developed white-box ML models in the present study significantly outperformed the classical regression model. For example, the MEP model achieved CC = 0.9932, RMSE = 0.8363, and MAE = 0.6701, while the MARS model produced CC = 0.9770, RMSE = 1.5188, and MAE = 1.1150. Similarly, the GEP and SVCM models showed noticeable improvements over the regression approach. These results clearly indicate that advanced white-box ML formulations are considerably more efficient at capturing the nonlinear relationship between blasting parameters and PPV than traditional regression equations. Moreover, Iphar et al. [7] reported that the ANFIS models achieved CC = 0.9940, RMSE = 0.7843, and MAE = 0.6705, whereas the MEP model obtained CC = 0.9932, RMSE = 0.8363, and MAE = 0.6701 across all datasets. The differences between the statistical indicator values of the ANFIS and MEP models are relatively small, indicating that the proposed MEP approach yields results comparable to the sophisticated ANFIS technique. Although the ANFIS model produced slightly lower RMSE values, the MEP model showed nearly identical MAE and correlation coefficient values.

It is worth noting that the white-box ML models developed in the present investigation have a significant practical advantage over the ANFIS model. The ANFIS technique is gener-

ally considered a black-box model that requires specialized software environments, training procedures, and expert knowledge in intelligent systems and programming for implementation and application. In contrast, the proposed ML formulations were transformed into explicit mathematical equations and implemented directly in Microsoft Excel spreadsheets. Consequently, engineers and practitioners can easily estimate PPV values without requiring advanced programming skills or detailed knowledge of ML algorithms. This significantly improves the practical usability of the developed models for field applications. However, among the developed ML models, the MARS model offers an additional advantage due to its relatively simple mathematical structure and ease of implementation compared with more complex symbolic formulations, such as MEP and GEP. Therefore, although ANFIS slightly outperformed the developed models in statistical accuracy, the proposed white-box ML formulations provide acceptable, competitive accuracy relative to ANFIS and, when implemented in Excel, improve ease of application and engineering applicability.

Furthermore, using the same benchmark dataset, Iphar et al. [7] reported a site-specific empirical regression model with a correlation coefficient (R) of 0.90 and an RMSE of 3.19. In comparison, all four proposed white-box machine learning models outperformed the classical empirical regression equation. Specifically, the developed models achieved correlation coefficients ranging from 0.951 to 0.993 and RMSE values ranging from 0.836 to 2.207, all of which represent substantial improvements over the regression model. In particular, the best-performing MEP model reduced the RMSE by approximately 74% while increasing the correlation coefficient from 0.90 to 0.993. These results further confirm that white-box machine learning models provide superior predictive capability compared with the conventional empirical equation.

7. Limitations and Future Works

Although the present study demonstrates the effectiveness of white-box machine learning models and practical engineering utility for predicting PPV, several limitations should be acknowledged. First, the dataset used in this study comprises 44 field observations from a previously published benchmark study reported by Iphar et al. [7]. While such dataset sizes are common in blast vibration research due to the high cost, logistical constraints, and safety considerations associated with controlled blasting experiments, generalizing the results to broader geological and operational conditions beyond those represented in the dataset should be done cautiously. However, five-fold cross-validation was implemented to rigorously mitigate the risks of overfitting and to confirm the stability of the proposed models across different data partitions.

Second, the current models were developed using only two primary controllable parameters: D and W. These two input parameters are widely recognized as dominant factors controlling PPV. However, future research can consider other influential geological and geomechanical factors such as rock mass properties, discontinuity conditions, burden, spacing, and explosive characteristics.

Third, although the developed white-box models generate explicit mathematical expressions that enhance practical usability, some of the derived expressions may not have a direct physical interpretation in terms of blast wave propagation mechanics. However, the parametric study demonstrates that the proposed models accurately capture the trend in PPV for each input variable.

Despite these limitations, the present study provides a systematic comparative assessment of four white-box ML approaches and demonstrates their potential for practical PPV prediction in mining engineering applications. To address these limitations, future research should focus on validating and updating the proposed explicit equations using larger, more

diverse, multi-site blasting datasets and incorporating additional geological variables to further enhance their global applicability.

8. Summary and Conclusions

This research investigated the application of four advanced white-box ML algorithms, namely, MEP, GEP, MARS, and SVCM, to predict blast-induced peak particle velocity (*PPV*). Based on a field dataset of 44 blasting records comprising distance from the blast face (*D*) and maximum explosive charge weight per delay (*W*), explicit mathematical formulations were successfully developed and evaluated. The main conclusions drawn from this research are as follows:

- The MEP approach demonstrated the highest predictive accuracy, with a CC of 0.9932, an RMSE of 0.8363, and an MAE of 0.6701 for the entire dataset. Furthermore, visual assessments using scatter, Taylor, and violin plots confirmed that the MEP estimations were most consistent with the actual PPV distribution.
- The MARS model ranked second in predictive accuracy, with CC = 0.9770 and RMSE = 1.5188. MARS predicted PPV with high accuracy using a relatively simple, piecewise-linear mathematical formulation. The GEP and SVCM models also demonstrated acceptable predictive performance, although they were lower than that of the MEP and MARS models. However, all ML models significantly outperformed the traditional empirical regression model developed in the previous study.
- A major contribution of this study is the generation of explicit, closed-form mathematical equations for PPV estimation. Unlike traditional black-box ML models, such as ANNs or ANFIS, the explicit equations generated by these white-box models were successfully translated into standard spreadsheet functions. This allows field engineers to instantly implement these high-performance models in Microsoft Excel without requiring advanced machine learning expertise.
- In comparison to the previously published ANFIS model utilizing the same dataset, the MEP model achieved highly competitive statistical results. While the ANFIS model provided a marginally lower RMSE (0.7843 vs. 0.8363), the MEP model achieved a nearly identical MAE (0.6701 vs. 0.6705) and CC (0.9932 vs. 0.9940). However, all white-box ML models are implemented in Excel spreadsheets, resulting in greater practical engineering applicability than conventional black-box techniques.
- Although all developed white-box models achieved high predictive correlation coefficients (CC > 0.95), the superiority of the MEP algorithm is not merely based on marginal metric differences from a single testing dataset. A robust cross-validation strategy was used to rigorously evaluate the models across multiple independent data folds. The five-fold cross-validation results indicated that the MEP model achieved optimal, statistically robust performance.
- The statistical metrics from the conventional train–test split showed strong consistency between the training and testing datasets, confirming that no overfitting occurred in the proposed models. In addition, five-fold cross-validation further supported this generalizability.
- The developed MEP model may appear more complex in its expression for the estimation of PPV compared to the simple structure formulation, such as the MARS model. However, the resulting formulation remains fully explicit and can be implemented directly in Excel (see the attached Supplementary Materials), enabling practical engineering use for estimation of PPV.
- The proposed equations should be applied only within the calibrated data range, corresponding to $90 \leq D \leq 1280$ m and $84.5 \leq W \leq 1208$ kg. Application outside these ranges constitutes extrapolation and may lead to unreliable predictions.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/mining6030056/s1, Table S1: Manual Guide and Results of the Proposed ML Models.

Author Contributions: Conceptualization, M.S.; methodology, M.S.; software, M.S.; validation, M.S. and Z.S.K.; formal analysis, M.S.; investigation, M.S. and Z.S.K.; data curation, M.S.; writing—original draft preparation, M.S.; writing—review and editing, M.S., S.A.K.-S. and Z.S.K.; visualization, M.S.; supervision, M.S., Z.S.K.; project administration, Z.S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset used in this study is provided as Supplementary Materials in Excel format. It includes 44 records of blast-induced peak particle velocity (PPV), the corresponding input variables (distance from the blast face, D, and charge weight per delay, W), and the predicted PPV values generated by the developed white-box machine learning models. The original field measurements were obtained from Iphar et al. [7]. To further enhance transparency and reproducibility, the explicit mathematical equations derived from the developed models, along with their spreadsheet implementations, are also provided.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Silewu, K.; Kahanji, C.; Simwanda, L.; Sykora, M. Intelligent Data Driven Ensemble Approaches for Bending Strength Prediction of Ultra-High Performance Concrete Beams. *Bull. Comput. Intell.* **2025**, *1*, 31–52. [CrossRef]
2. Shylaja, G.; Prashanth, R. A systematic survey of hybrid ML techniques for predicting peak particle velocity (PPV) in open-cast mine blasting operations. *Artif. Intell. Rev.* **2025**, *58*, 203. [CrossRef]
3. Gao, F.Q.; Wang, X.Q. Using artificial neural network to predict blast-induced ground vibration. *Appl. Mech. Mater.* **2012**, *170*, 1013–1016. [CrossRef]
4. Fişne, A.; Kuzu, C.; Hüdaverdi, T. Prediction of environmental impacts of quarry blasting operation using fuzzy logic. *Environ. Monit. Assess.* **2011**, *174*, 461–470. [PubMed]
5. Mohamad, E.T.; Noorani, S.A.; Armaghani, D.J.; Saad, R. Simulation of blasting induced ground vibration by using artificial neural network. *Electron. J. Geotech. Eng.* **2012**, *17*, 2571–2584.
6. Mohammadnejad, M.; Gholami, R.; Ramezanzadeh, A.; Jalali, M.E. Prediction of blast-induced vibrations in limestone quarries using support vector machine. *J. Vib. Control* **2012**, *18*, 1322–1329.
7. Iphar, M.; Yavuz, M.; Ak, H. Prediction of ground vibrations resulting from the blasting operations in an open-pit mine by adaptive neuro-fuzzy inference system. *Environ. Geol.* **2008**, *56*, 97–107.
8. Parida, A.; Mishra, M.K. Blast vibration analysis by different predictor approaches-A comparison. *Procedia Earth Planet. Sci.* **2015**, *11*, 337–345. [CrossRef]
9. Rong, K.; Xu, X.; Wang, H.; Yang, J. Predicting the blast-induced ground vibration with support vector regression optimized by five swarm algorithms. *Earth Sci. Inform.* **2025**, *18*, 424. [CrossRef]
10. Ogunsola, N.O.; Shin, C.; Lawal, A.I.; Kim, Y.K.; Cho, S. Blasting-induced ground vibration modeling in tunnel excavation: A comparative study of ANN, hybrid ANNs, and empirical models. *Rock Mech. Rock Eng.* **2025**, *58*, 2711–2737.
11. Khatti, J. Assessment of Blast-Induced Peak Particle Velocity Using Six Hybrid Least Squares Support Vector Machine Models: A Case Study of an Opencast Mine in the Ramagundam Area, Telangana. *J. Inst. Eng. (India) Ser. D* **2025**, 1–10. [CrossRef]
12. Niu, L.; Yin, Q.; Zhu, W.; Chen, Z.; Wang, A.; Jiang, Z.; Chen, T. Online monitoring and data correction methods of blast induced ground vibration based on WOA-BP. *Measurement* **2025**, *258*, 119567. [CrossRef]
13. Zhao, S.; Wang, L.; Cao, M. Chaos Game Optimization-Hybridized Artificial Neural Network for Predicting Blast-Induced Ground Vibration. *Appl. Sci.* **2024**, *14*, 3759. [CrossRef]
14. Bhattacharjee, N.; Sur, A.; Manna, B.; Banerjee, A.; Shahu, J.T. ANN-based approach for the prediction of peak particle velocity of ground induced by underground metro operations. *Transp. Infrastruct. Geotechnol.* **2025**, *12*, 96. [CrossRef]
15. Yuan, H.; Zou, Y.; Li, H.; Ji, S.; Gu, Z.; He, L.; Hu, R. Assessment of peak particle velocity of blast vibration using hybrid soft computing approaches. *J. Comput. Des. Eng.* **2025**, *12*, 154–176. [CrossRef]
16. Ma, T.; Chen, C.; Shen, L.; Luo, K.; Jiang, Z.; Liu, H.; Hu, X.; Lin, Y.; Peng, K. A novel social network search and LightGBM framework for accurate prediction of blast-induced peak particle velocity. *Front. Struct. Civ. Eng.* **2025**, *19*, 645–662. [CrossRef]

17. Gaopale, K.; Sasaoka, T.; Hamanaka, A.; Shimada, H. Hybrid Machine Learning Model for Blast-Induced Peak Particle Velocity Estimation in Surface Mining: Application of Sparrow Search Algorithm in ANN Optimization. *Algorithms* **2025**, *18*, 543. [CrossRef]
18. Ge, L.; He, J.; Zhang, Z.; Yin, L.; Jia, S.; Zhao, Y. Particle peak velocity prediction based on risk-oriented hybrid ensemble learning. *Sci. Rep.* **2026**, *16*, 18391. [CrossRef] [PubMed]
19. Fissha, Y.; Ikeda, H.; Toriya, H.; Owada, N.; Adachi, T.; Kawamura, Y. Evaluation and prediction of blast-induced ground vibrations: A Gaussian process regression (GPR) approach. *Mining* **2023**, *3*, 659–682. [CrossRef]
20. Ogunsola, N.O.; Grobler, H. Predicting Blasting-Induced Ground Vibration in Mines Using Machine Learning and Empirical Models: Advancing Sustainable Mining and Minimizing Environmental Footprint. *Mining* **2026**, *6*, 32. [CrossRef]
21. Hasanipanah, M.; Monjezi, M.; Shahnazar, A.; Armaghani, D.J.; Farazmand, A. Feasibility of indirect determination of blast induced ground vibration based on support vector machine. *Measurement* **2015**, *75*, 289–297. [CrossRef]
22. Armaghani, D.J.; Hajihassani, M.; Mohamad, E.T.; Marto, A.; Noorani, S.A. Blasting-induced flyrock and ground vibration prediction through an expert artificial neural network based on particle swarm optimization. *Arab. J. Geosci.* **2014**, *7*, 5383–5396.
23. Saubi, O.; Jamisola, R.S., Jr.; Suglo, R.S.; Matsebe, O. Simultaneous prediction and design minimizing of blast impacts at Jwaneng mine using a GA-ANN framework. *Discov. Appl. Sci.* **2026**, *8*, 547. [CrossRef]
24. Zhou, J.; Li, C.; Koopialipoor, M.; Jahed Armaghani, D.; Thai Pham, B. Development of a new methodology for estimating the amount of PPV in surface mines based on prediction and probabilistic models (GEP-MC). *Int. J. Min. Reclam. Environ.* **2021**, *35*, 48–68.
25. Bayat, P.; Monjezi, M.; Rezakhah, M.; Armaghani, D.J. Artificial neural network and firefly algorithm for estimation and minimization of ground vibration induced by blasting in a mine. *Nat. Resour. Res.* **2020**, *29*, 4121–4132. [CrossRef]
26. Behzadafshar, K.; Mohebbi, F.; Soltani Tehrani, M.; Hasanipanah, M.; Tabrizi, O. Predicting the ground vibration induced by mine blasting using imperialist competitive algorithm. *Eng. Comput.* **2018**, *35*, 1774–1787. [CrossRef]
27. Ghasemi, M.; Samadi, M.; Soleimanian, E.; Chau, K.W. A comparative study of black-box and white-box data-driven methods to predict landfill leachate permeability. *Environ. Monit. Assess.* **2023**, *195*, 862. [CrossRef] [PubMed]
28. Shishegaran, A.; Saeedi, M.; Mirvalad, S.; Korayem, A.H. Computational predictions for estimating the performance of flexural and compressive strength of epoxy resin-based artificial stones. *Eng. Comput.* **2023**, *39*, 347–372.
29. Shishegaran, A.; Varaee, H.; Rabczuk, T.; Shishegaran, G. High correlated variables creator machine: Prediction of the compressive strength of concrete. *Comput. Struct.* **2021**, *247*, 106479. [CrossRef]
30. Shishegaran, A.; Samadi, M.; Torabi, M. Novel computational methods to predict the compressive strength of hydrothermally solidified clay. *Front. Struct. Civ. Eng.* **2025**, *19*, 2054–2072. [CrossRef]
31. Azma, A.; Borthwick, A.G.; Liu, Y.; Zhang, D.; Afaridegan, E. Comparative assessment of white-box and black-box machine learning models for predicting free-falling jet scour depth. *Eng. Appl. Comput. Fluid Mech.* **2026**, *20*, 2638090. [CrossRef]
32. Samadi, M.; Shishegaran, A.; Torabi, M.; Sheikh Khozani, Z. Prediction of Scour Hole Geometry Downstream of Ski-Jump Spillways Using Novel Intelligent Computational Machine Learning Models. *Forecasting* **2026**, *8*, 49. [CrossRef]
33. Shishegaran, A.; Varaee, H. Comparison among creator variable machine methods: Compressive strength prediction of ultra-high-performance concrete. *Case Stud. Constr. Mater.* **2026**, *35*, e06253. [CrossRef]
34. Sheikh Khozani, Z.; Safari, M.J.S.; Danandeh Mehr, A.; Melini Wan Mohtar, W.H. An ensemble genetic programming approach to develop incipient sediment motion models in rectangular channels. *J. Hydrol.* **2020**, *584*, 124753. [CrossRef]
35. Samadi, M.; Torabi, M.; Ghorbani, M.K.; Hamidifar, H.; Nikoo, M.R. The application of data-driven models for estimating the discharge coefficients of labyrinth gates. *Eng. Appl. Artif. Intell.* **2026**, *171*, 114277. [CrossRef]
36. Samadi, M.; Jabbari, E.; Azamathulla, H.M.; Mojallal, M. Estimation of scour depth below free overfall spillways using multivariate adaptive regression splines and artificial neural networks. *Eng. Appl. Comput. Fluid Mech.* **2015**, *9*, 291–300. [CrossRef]
37. Samadi, M.; Sarkardeh, H.; Jabbari, E. Explicit data-driven models for prediction of pressure fluctuations occur during turbulent flows on sloping channels. *Stoch. Environ. Res. Risk Assess.* **2020**, *34*, 691–707. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.