

A Hybrid-Statistical Machine Learning Model for Blast Vibration Prediction in Surface Mines

Amresh Kumar¹

¹Department of Mining Engineering, CTAE, Udaipur, Rajasthan, India

ABSTRACT

Blast-induced ground vibration is a major environmental and operational constraint in surface mining, and Peak Particle Velocity (PPV) is the principal response used for vibration assessment. This study develops a physically informed data-driven framework for PPV prediction using controlled-blast records, conventional statistical modelling, and supervised machine learning. The response was the maximum PPV recorded from three orthogonal vibration components. Raw predictors represented blast geometry, explosive loading, initiation timing, monitoring distance, and rock-breakage performance; additional ratios and scaled distance were introduced through feature engineering. After data cleaning and exploratory analysis, the observations were divided into 75% training and 25% independent testing subsets, with cross-validation applied within the training data. Linear Regression, Random Forest, and an Artificial Neural Network (ANN) were compared using R^2 , mean absolute error, and root mean squared error. The ANN produced the strongest testing performance ($R^2 = 0.9337$, MAE = 0.5650 mm/s, RMSE = 0.7328 mm/s), outperforming Linear Regression and Random Forest. Correlation and feature-importance results identified monitoring distance as the dominant vibration-control variable, with total explosive quantity, powder factor, and engineered distance-related features also contributing. The results support ANN-based site-specific PPV estimation while retaining interpretable statistical and ensemble models for engineering checks.

Keywords: blast-induced vibration; peak particle velocity; surface mining; linear regression; random forest; artificial neural network; feature engineering

I. INTRODUCTION

1.1 Background

Blasting remains the dominant rock-fragmentation method in surface mining because it enables efficient excavation, loading, hauling, crushing, and grinding. The same rapid release of explosive energy that fragments rock also generates stress waves that propagate through the surrounding rock mass. These waves produce ground vibration, air overpressure, flyrock, and other effects that must be controlled to protect mine infrastructure, nearby structures, personnel, and surrounding communities.

Peak Particle Velocity (PPV) is widely used to quantify blast-induced ground motion. Its magnitude is governed by explosive energy released per delay, monitoring distance, blast geometry, initiation timing, and the mechanical and structural characteristics of the rock mass. Conventional scaled-distance equations provide a convenient relationship between charge weight, distance, and PPV, but they simplify the nonlinear and site-specific interactions present in operating mines. This motivates the use of multivariable statistical and machine-learning models that can learn complex relationships from field records.

1.2 Objectives and scope

- Investigate the influence of blast-design parameters, explosive characteristics, and site-related variables on PPV.
- Develop and compare supervised models for PPV prediction.
- Evaluate feature sensitivity and relative importance to support physical consistency and engineering interpretation.
- Provide a prediction framework that can assist blast-design optimization and vibration control.

II. RELATED WORK

Research on blast-vibration prediction has progressed from empirical equations to statistical, machine-learning, optimization, and ensemble approaches. Amiri et al. combined Artificial Neural Networks (ANN) with K-Nearest Neighbours and reported improved prediction of PPV and air overpressure relative to a standalone ANN and USBM equations [2]. Hasanipanah et al. applied Particle Swarm Optimization to PPV prediction and obtained improved performance over multiple linear regression and conventional empirical models [6]. Gaussian Process Regression also outperformed several neural-network and empirical predictors in a Ghanaian open-pit mine [3], while a Firefly-optimized ANN achieved the best performance among ANN, CART, KNN, and SVM alternatives in a Vietnamese quarry dataset [12].

More recent work has emphasized model comparison and ensemble learning. Hosseini et al. compared statistical, machine-learning, deep-learning, and optimized hybrid approaches, reporting strong results for an optimized LSTM model [7]. Ensemble ANN and XGBoost frameworks have likewise improved PPV prediction

through stacked generalization [8]. Reviews of mine-blasting applications show increasing use of hybrid and ensemble models [1,5], while broader environmental-effect reviews identify continued need for validated prediction and mitigation frameworks [14]. IoT-based monitoring combined with Random Forest has demonstrated practical real-time potential [4], and a recent open-pit study found Support Vector Regression to outperform ANN for its dataset [13]. Related mining applications include ground-motion modelling [9], flyrock prediction [10], and machine-learning assessment of rockburst potential [11].

2.1 Research gap

The literature confirms that no single algorithm is universally superior: performance depends on data quality, geological variability, predictor selection, and validation design. The present study therefore compares an interpretable statistical baseline, a nonlinear ensemble model, and a nonlinear neural-network model under a common train-test framework. It also retains engineered blasting ratios and scaled distance so that model behaviour can be interpreted against established vibration mechanisms.

III. MATERIALS AND METHODS

3.1 Data source and response definition

The dataset was obtained from controlled blasting operations in a mining environment. Ground vibration was monitored using standard seismographs at specified distances from the blast site, and the maximum PPV from three orthogonal components was used as the response variable. The supplied thesis does not state the final number of usable blast records or identify the mine site; these details should be added from the original dataset before journal submission.

Table 1. Study design and modelling framework.

Element	Description
Study type	Quantitative, site-specific PPV prediction using field blast records
Response	Maximum recorded Peak Particle Velocity (PPV)
Training/testing	75% training and 25% independent testing
Internal validation	Cross-validation within the training subset
Models	Linear Regression, Random Forest Regressor, Artificial Neural Network
Reported test metrics	R^2 , MAE, and RMSE

3.2 Predictor set

The predictor set represented blast geometry, explosive loading, timing, monitoring geometry, and rock breakage. Feature engineering introduced ratios that describe confinement and geometric proportionality, together with scaled distance to connect data-driven learning with conventional vibration attenuation concepts.

Table 2. Raw and engineered variables retained for modelling.

Feature group	Variables
Blast geometry	Burden; spacing; depth; height; stemming; charge-column length (CCL)
Blast scale and energy	Number of rows; number of holes; total explosive quantity; charge per delay
Timing and monitoring	Hole delay; row delay; distance
Breakage response	Volume of rock broken; powder factor (PF)
Engineered variables	Spacing/burden (S/B); height/burden (H/B); stemming/burden (St/B); depth/burden (D/B); scaled distance

III. MATERIALS AND METHODS (CONTINUED)

3.3 Blast variables and preprocessing

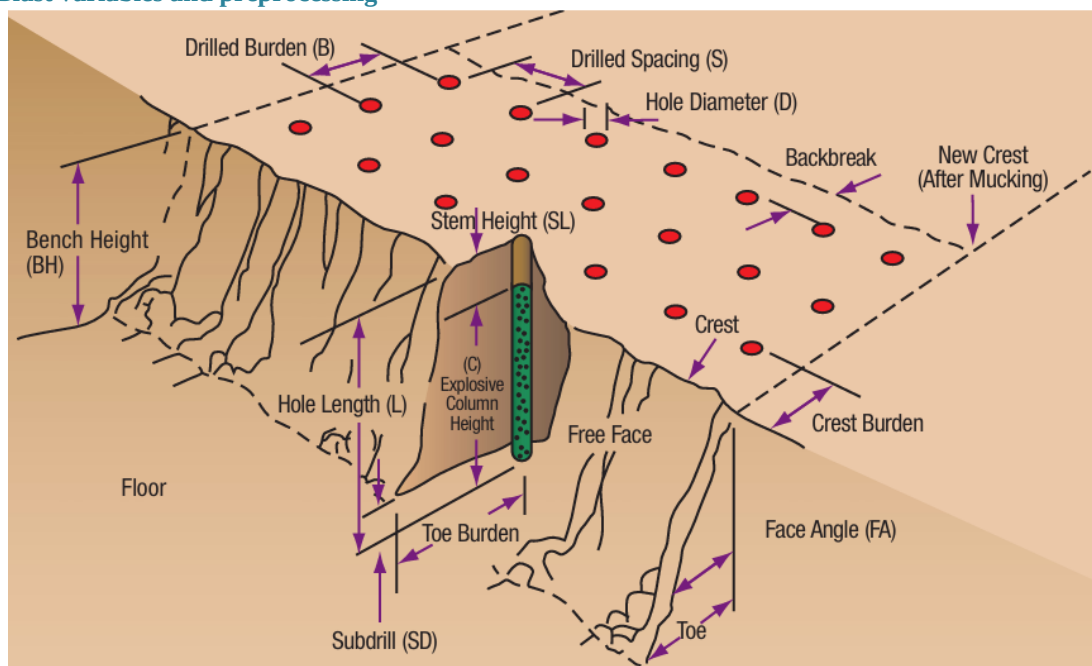


Figure 1. Blast-design parameters considered in the study (retained from the thesis).

Burden and spacing govern explosive confinement and stress-wave interaction; depth, bench height, stemming, and charge-column length describe the vertical charge geometry; and the number of holes and rows represent blast scale. Total explosive quantity and charge per delay characterize energy input, while hole and row delays describe the temporal separation of detonations. Distance controls wave attenuation, and volume of rock broken and powder factor link explosive energy to fragmentation performance.

Data cleaning addressed missing values, inconsistent records, and non-physical values. Outliers were examined using distribution plots and engineering judgement rather than removed automatically, because extreme PPV observations may correspond to genuine high-vibration blasting conditions. Variables were converted to consistent numerical formats, and categorical information was encoded where present.

3.4 Exploratory analysis and feature engineering

Exploratory Data Analysis (EDA) used histograms, boxplots, scatterplots, and correlation heatmaps to examine distributions, skewness, outliers, pairwise relationships, and multicollinearity. The resulting patterns guided retention of physically meaningful raw variables and development of the S/B, H/B, St/B, D/B, and scaled-distance features. These derived variables were not treated as substitutes for raw measurements; they were included to represent interactions that are difficult to express through individual predictors alone.

3.5 Model development and validation

The observations were partitioned into 75% training and 25% testing subsets. The testing subset remained isolated until final evaluation, while cross-validation within the training data supported model selection and hyperparameter tuning. Scaling parameters were estimated from the training subset and applied to the testing subset to avoid information leakage. Scaling was particularly important for the ANN, whereas the tree-based Random Forest was comparatively insensitive to feature magnitude.

Linear Regression was used as the transparent baseline. Random Forest represented nonlinear ensemble learning and provided feature-importance estimates. The feed-forward ANN represented a flexible nonlinear model trained with feature scaling, regularization, and early stopping. Predictive performance was assessed through the coefficient of determination (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). Lower MAE and RMSE indicate smaller average and large-error deviations, respectively, while higher R^2 indicates greater explained variance.

IV. RESULTS

4.1 Distributions of blasting variables and PPV

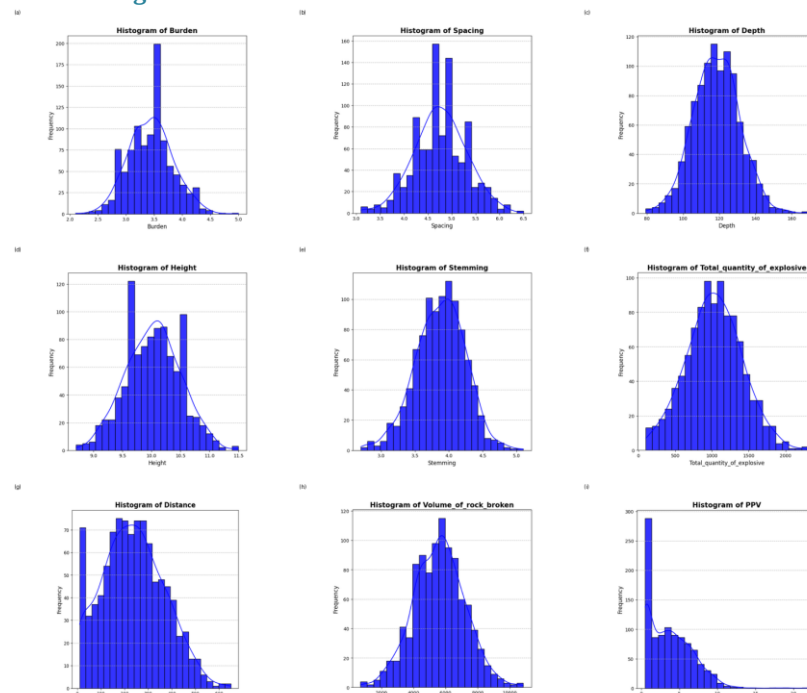


Figure 2. Histograms of (a) burden, (b) spacing, (c) depth, (d) height, (e) stemming, (f) total explosive quantity, (g) distance, (h) volume of rock broken, and (i) PPV.

Most geometric and production variables show concentrated central ranges with varying degrees of skewness. PPV is strongly right-skewed: a large proportion of observations occur at low vibration levels, while comparatively few observations extend to high PPV. This imbalance is operationally important because the high-PPV tail contains the cases most relevant to structural safety and regulatory compliance.

IV. RESULTS (CONTINUED)

4.2 Outlier structure

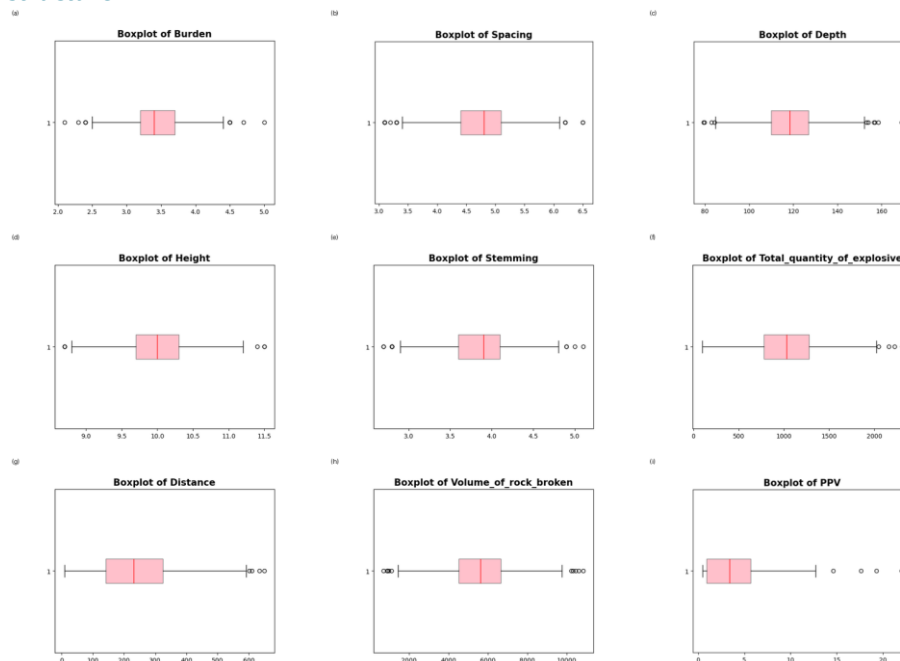


Figure 3. Boxplots of (a) burden, (b) spacing, (c) depth, (d) height, (e) stemming, (f) total explosive quantity, (g) distance, (h) volume of rock broken, and (i) PPV.

The boxplots identify extreme values in several variables, including burden, spacing, depth, height, stemming, total explosive quantity, distance, volume of rock broken, and PPV. The thesis retained physically plausible extremes rather than applying automatic deletion. This decision preserves rare but important operating conditions and reduces the risk that the models will be evaluated only on routine low-vibration observations.

IV. RESULTS (CONTINUED)

4.3 Pairwise relationships with PPV

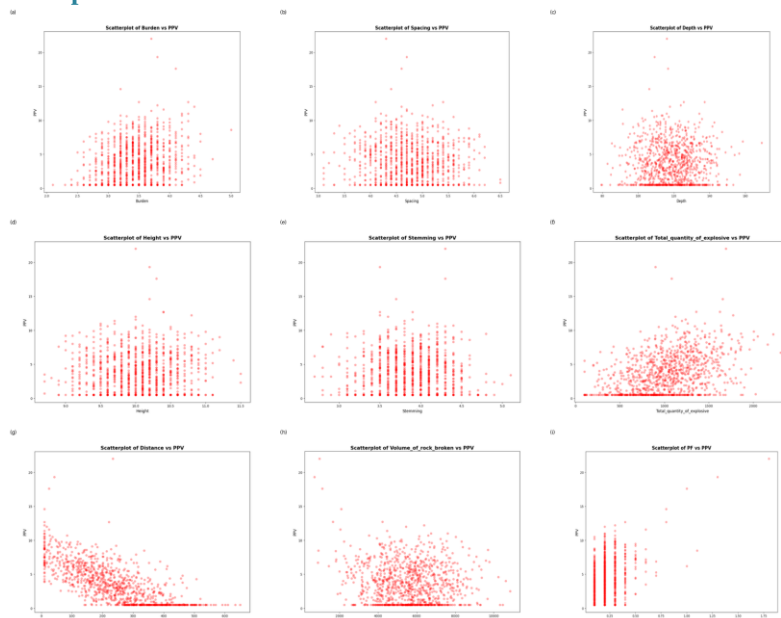


Figure 4. PPV relationships with (a) burden, (b) spacing, (c) depth, (d) height, (e) stemming, (f) total explosive quantity, (g) distance, (h) volume of rock broken, and (i) powder factor.

The clearest pairwise trend is the reduction of PPV with increasing distance. Powder factor and total explosive quantity show positive tendencies, while the geometric variables exhibit broad dispersion and weaker marginal trends. The scatter confirms that PPV cannot be represented adequately by a single input: observations with similar values of one blast parameter can produce substantially different vibration levels because of interactions with energy, timing, geometry, and site conditions.

IV. RESULTS (CONTINUED)

4.4 Correlation before feature engineering

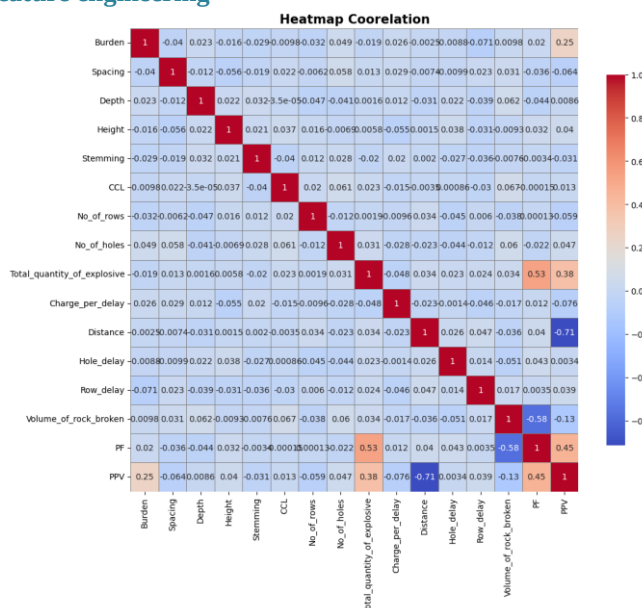


Figure 5. Correlation matrix before feature engineering.

Before feature engineering, PPV has its strongest linear association with distance ($r = -0.71$). Positive associations are observed with powder factor ($r = 0.45$), total explosive quantity ($r = 0.38$), and burden ($r = 0.25$), while the remaining raw variables show comparatively weak marginal correlations. These coefficients describe linear pairwise association only; weak correlation does not exclude nonlinear or interaction effects captured by the machine-learning models.

IV. RESULTS (CONTINUED)

4.5 Correlation after feature engineering

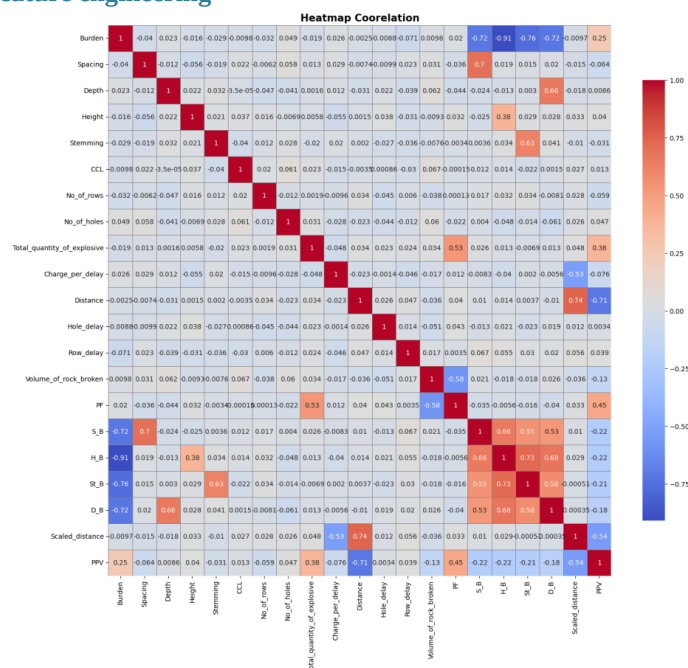


Figure 6. Correlation matrix after feature engineering.

After feature engineering, scaled distance shows a substantial negative association with PPV ($r = -0.54$). The geometric ratios S/B, H/B, St/B, and D/B have negative PPV correlations of approximately -0.22, -0.22, -0.21, and -0.18, respectively. Distance remains the strongest individual raw correlate, while powder factor and total explosive quantity retain positive associations. The matrix also reveals strong relationships between several engineered ratios and their component variables, which must be considered when interpreting model coefficients and importance scores.

IV. RESULTS (CONTINUED)

4.6 Linear Regression and Random Forest diagnostics

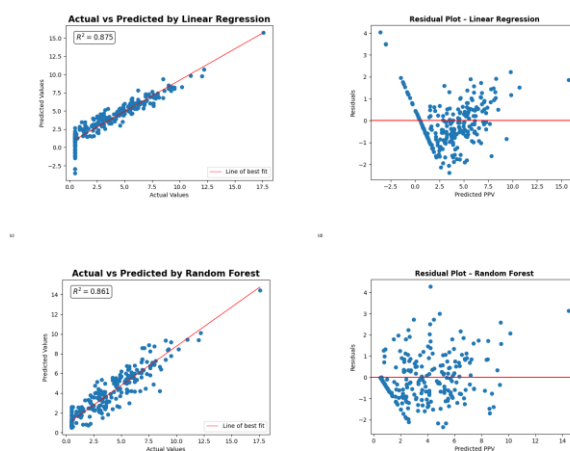


Figure 7. Model diagnostics: (a) Linear Regression actual versus predicted, (b) Linear Regression residuals, (c) Random Forest actual versus predicted, and (d) Random Forest residuals.

Linear Regression follows the overall PPV trend but shows structured residual patterns and larger deviations at the extremes, indicating that a linear additive form does not fully represent the response. Random Forest

captures nonlinear variation, but its test-set performance in this dataset is slightly weaker than Linear Regression. Its residual plot also shows dispersion across the prediction range, and the difference between training and testing R^2 indicates a greater risk of overfitting than for the other two models.

IV. RESULTS (CONTINUED)

4.7 Random Forest feature importance

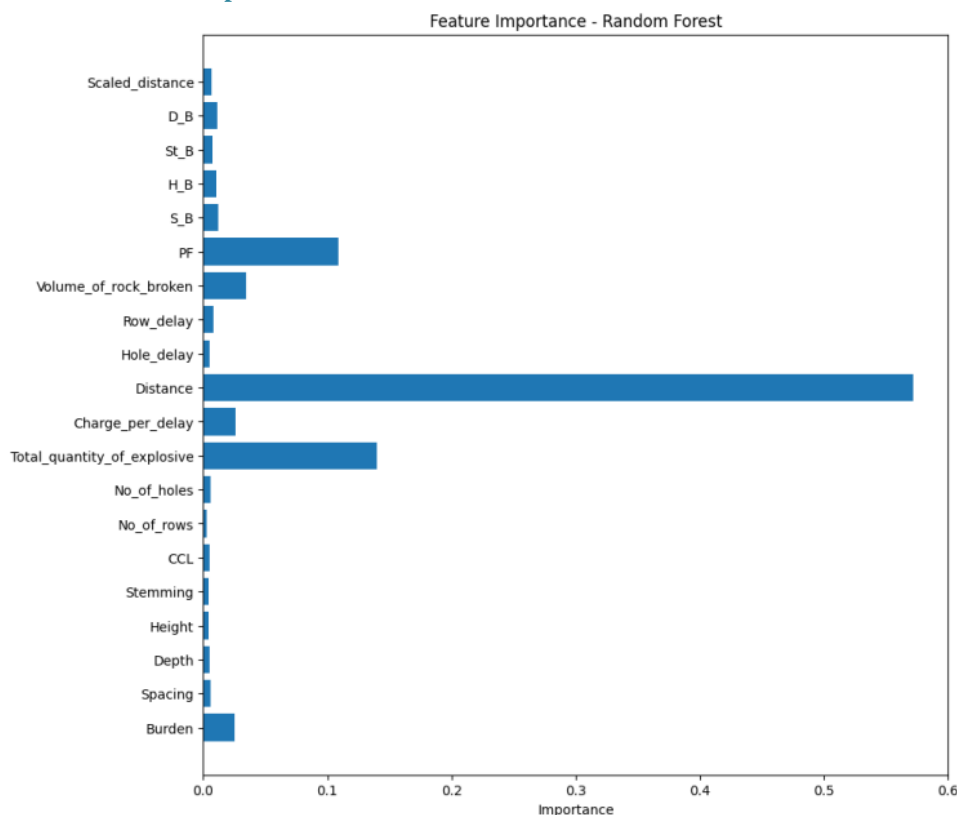


Figure 8. Random Forest feature importance.

Monitoring distance dominates the Random Forest importance profile. Total explosive quantity and powder factor form the next most influential group, followed by volume of rock broken, charge per delay, burden, and the engineered ratios. The ranking is physically credible because distance governs attenuation, while explosive quantity and powder factor represent energy release and energy intensity. Importance values should nevertheless be interpreted as model-specific predictive contributions rather than causal effects.

4.8 Comparative model performance

Table 3. Reported performance of the three PPV prediction models.

Model	Train R^2	Test R^2	MAE (mm/s)	RMSE (mm/s)
Linear Regression	0.8744	0.8747	0.7966	1.0072
Random Forest	0.9748	0.8609	0.8211	1.0612
Artificial Neural Network	0.9729	0.9337	0.5650	0.7328

The ANN achieved the highest test R^2 and the lowest MAE and RMSE. Linear Regression marginally outperformed Random Forest on the independent test set, despite the much higher Random Forest training R^2 . This result demonstrates why model ranking must be based on unseen data rather than training fit. Relative to Linear Regression, the ANN reduced MAE by approximately 29%; relative to Random Forest, the reduction was approximately 31%, as reported in the thesis.

IV. RESULTS (CONTINUED)

4.9 Artificial Neural Network diagnostics

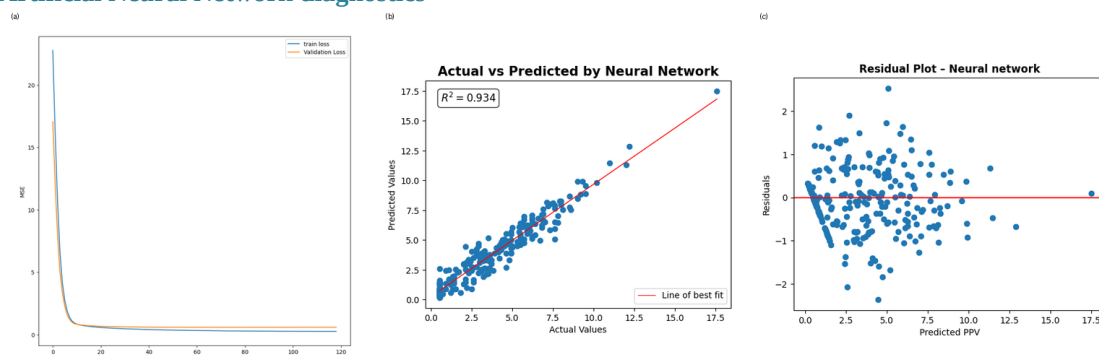


Figure 9. ANN diagnostics: (a) training and validation loss, (b) actual versus predicted PPV, and (c) residuals.

The ANN training and validation loss curves decrease rapidly and then stabilize without marked divergence, consistent with the regularization and early-stopping strategy described in the thesis. The actual-versus-predicted plot shows the closest alignment to the fitted line among the three models, with a plotted R^2 of approximately 0.934. Residuals are concentrated more closely around zero than in the baseline models, although some structured bands and isolated deviations remain.

The ANN test R^2 of 0.9337 indicates that the model explains more than 93% of the observed PPV variance in the held-out subset. Its MAE of 0.5650 mm/s and RMSE of 0.7328 mm/s show improved average accuracy and better control of large prediction errors. The smaller training-testing gap relative to Random Forest suggests stronger generalization for the reported split, but the result remains site-specific and should be confirmed using additional mines and geological conditions.

V. DISCUSSION

5.1 Engineering interpretation

The results demonstrate that increasing algorithmic flexibility is beneficial only when it improves independent predictive performance. Linear Regression provides transparent coefficients and a stable baseline, but the residual structure confirms under-representation of nonlinear blast interactions. Random Forest offers useful feature importance and nonlinear modelling, yet its stronger training fit did not translate into superior testing performance. The ANN achieved the best balance of explained variance and error reduction for the reported dataset.

The dominance of distance in both correlation and Random Forest importance is consistent with geometric spreading and material damping of blast-generated waves. The contributions of total explosive quantity and powder factor reflect the role of energy input and energy concentration. Engineered ratios and scaled distance improve representation of geometry and attenuation, but their correlations with component variables require cautious interpretation to avoid attributing duplicated information to separate physical mechanisms.

V. DISCUSSION (CONTINUED)

5.2 Practical implications

For operational blast design, the models can be used to compare prospective parameter combinations and identify designs likely to remain within site-specific vibration limits. The ANN provides the most accurate prediction in the reported comparison, while Linear Regression and Random Forest remain valuable for checking directionality, plausibility, and parameter influence. A practical implementation should therefore combine the ANN prediction with engineering review rather than treat the model as an autonomous substitute for blast design expertise.

5.3 Limitations

The study is limited by the information available in the thesis source. The total number of usable blast records, the mine location, detailed geological classifications, instrumentation settings, and the underlying raw dataset are not reported in the supplied document. Model performance is based on one 75:25 split with cross-validation inside the training data; external validation at other mines is not presented. The ANN is also less interpretable than the statistical and tree-based alternatives, and the observed relationships should not be transferred to new sites without recalibration.

VI. CONCLUSION AND FUTURE SCOPE

A comparative framework integrating statistical modelling, feature engineering, Random Forest, and an Artificial Neural Network was developed for blast-induced PPV prediction. The ANN produced the strongest independent test results, with $R^2 = 0.9337$, MAE = 0.5650 mm/s, and RMSE = 0.7328 mm/s. Linear Regression

remained a useful transparent baseline, while Random Forest contributed interpretable feature rankings despite lower test accuracy. Distance was the dominant predictor, followed by variables representing explosive quantity and energy intensity.

The findings support site-specific machine learning as a complement to blasting theory and conventional scaled-distance analysis. Future work should expand the dataset across multiple mines and geological conditions; incorporate Rock Mass Rating, Geological Strength Index, joint characteristics, and in-situ stress; quantify predictive uncertainty; and evaluate explainable-AI methods such as SHAP and sensitivity analysis. Integration with real-time monitoring and blast-planning systems could extend the framework to vibration control, air overpressure, flyrock, fragmentation, and blast-induced damage.

Data and reproducibility statement

The article was converted exclusively from the supplied thesis. The underlying data file and executable modelling code were not included in the supplied materials. The figures and numerical values reproduced here are therefore traceable to the thesis document, while independent reproduction will require the original dataset and code.

APPENDIX A. LINEAR REGRESSION COEFFICIENTS

The thesis reports the following fitted Linear Regression intercept and coefficients. They are reformatted as a table to preserve the complete model while improving readability; coefficient values are unchanged.

Table A1. Linear Regression intercept and coefficients reported in the thesis.

Term	Coefficient	Term	Coefficient
Intercept	-2.4283027005844913	Distance	-1.11134753e-02
Burden	1.73261041e+00	Hole delay	-1.56489932e-03
Spacing	1.63452844e-01	Row delay	1.48061791e-02
Depth	-1.85555710e-02	Volume of rock broken	1.52405606e-04
Height	-3.80699617e-02	Powder factor (PF)	1.00417144e+01
Stemming	4.77291998e-01	Spacing/burden (S/B)	-1.34458288e+00
Charge-column length (CCL)	8.17915035e-02	Height/burden (H/B)	6.41876307e-01
Number of rows	-9.95898084e-02	Stemming/burden (St/B)	-2.04932158e+00
Number of holes	1.10427118e-02	Depth/burden (D/B)	6.82170213e-02
Total quantity of explosive	1.20673480e-03	Scaled distance	-4.58929904e-02
Charge per delay	-1.09216901e-02		

Model form: $PPV = \text{intercept} + \sum(\text{coefficient} \times \text{predictor})$. The coefficient signs and magnitudes should be interpreted with caution because several raw and engineered predictors are correlated.

Editorial items required before submission

- Insert the final number of usable blast records and the mine/site description.
- Confirm measurement units for every predictor and provide the original dataset or a data-access statement.
- Add the author's correspondence email and any supervisor/co-author details, if applicable.
- Confirm funding, conflict-of-interest, and institutional declarations required by the target journal.

REFERENCES

1. Abd Elwahab, A., Topal, E., & Jang, H. D. (2023). Review of machine learning application in mine blasting. *Arabian Journal of Geosciences*, 16(2), 133.
2. Amiri, M., Bakhshandeh Amnieh, H., Hasanipanah, M., & Mohammad Khanli, L. (2016). A new combination of artificial neural network and K-nearest neighbors models to predict blast-induced ground vibration and air-overpressure. *Engineering with Computers*, 32(4), 631-644.
3. Arthur, C. K., Temeng, V. A., & Ziggah, Y. Y. (2020). Novel approach to predicting blast-induced ground vibration using Gaussian process regression. *Engineering with Computers*, 36(1), 29-42.
4. Aruna, M., Vardhan, H., Tripathi, A. K., Parida, S., Raja Sekhar Reddy, N. V., Sivalingam, K. M., et al. (2025). Enhancing safety in surface mine blasting operations with IoT based ground vibration monitoring and prediction system integrated with machine learning. *Scientific Reports*, 15(1), 3999.

5. Dumakor-Dupey, N. K., Arya, S., & Jha, A. (2021). Advances in blast-induced impact prediction-A review of machine learning applications. *Minerals*, 11(6), 601.
6. Hasanipanah, M., Naderi, R., Kashir, J., Noorani, S. A., & Zeinali Aaq Qaleh, A. (2017). Prediction of blast-produced ground vibration using particle swarm optimization. *Engineering with Computers*, 33(2), 173-179.
7. Hosseini, S., Khatti, J., Taiwo, B. O., Fissaha, Y., Grover, K. S., Ikeda, H., et al. (2023). Assessment of the ground vibration during blasting in mining projects using different computational approaches. *Scientific Reports*, 13(1), 18582.
8. Hosseini, S., Pourmirzaee, R., Armaghani, D. J., & Sabri Sabri, M. M. (2023). Prediction of ground vibration due to mine blasting in a surface lead-zinc mine using machine learning ensemble techniques. *Scientific Reports*, 13(1), 6591.
9. Lapčević, R., Kostić, S., Pantović, R., & Vasović, N. (2014). Prediction of blast-induced ground motion in a copper mine. *International Journal of Rock Mechanics and Mining Sciences*, 69, 19-25.
10. Manoj, K., & Monjezi, M. (2013). Prediction of flyrock in open pit blasting operation using machine learning method. *International Journal of Mining Science and Technology*, 23(3), 313-316.
11. Pu, Y., Apel, D. B., Liu, V., & Mitri, H. (2019). Machine learning methods for rockburst prediction-state-of-the-art review. *International Journal of Mining Science and Technology*, 29(4), 565-570.
12. Shang, Y., Nguyen, H., Bui, X. N., Tran, Q. H., & Moayedi, H. (2020). A novel artificial intelligence approach to predict blast-induced ground vibration in open-pit mines based on the firefly algorithm and artificial neural network. *Natural Resources Research*, 29(2), 723-737.
13. Ullah, S., Ren, G., Ge, Y., Fissaha, Y., Kinyua, E. M., & Zhang, L. (2025). Machine Learning-Based Prediction of Blast-Induced Ground Vibration in Open-Pit Mining. *Journal of Vibration Engineering & Technologies*, 13(5), 293.
14. Zhou, J., Zhang, Y., & Qiu, Y. (2024). State-of-the-art review of machine learning and optimization algorithms applications in environmental effects of blasting. *Artificial Intelligence Review*, 57(1), 5.