Multiple linear regression analysis model and artificial neural network model to calculate and estimate the blast induced area of the tunnel face. A case study Deo Ca tunnel

Thanh Chi Nguyen 1,*, Anh Ngoc Do 1, Vi Van Pham 1, Gospodarikov Alexandr 2

1 Faculty of Civil Engineering, Hanoi University of Mining and Geology, Vietnam
2 Saint Petersburg Mining University, Russia Federation

ARTICLE INFO

Article history:
Received 06th Jan. 2022
Accepted 13th Apr. 2022
Available online 30th June 2022

Keywords:
Area of tunnel face,
Artificial neural network (ANN),
Blasting,
Multiple linear regression,
Predict.

ABSTRACT

The area of the tunnel face after the blasting is a very important factor in underground excavations where the drilling and blasting method is used. The area of the tunnel face, this is a significant factor that has affected the cost and safety of underground constructions in case of using the drilling and blasting method in underground excavations. Because the area of the tunnel after the blasting depends on many different parameters, such as geological conditions in the area where the tunnel is located, the parameters of the explosion, and other parameters of the tunnel, it is very difficult to accurately determine the value of the tunnel face area after blasting. This paper uses the data obtained in the actual blasting of the Deo Ca tunnel (39 datasets) to build the computational and prediction models for the area of the tunnel face after blasting by two methods, the multiple linear regression analysis method and the method of using artificial neural network (ANN). Determination coefficient $R^2$ of multiple linear regression analysis (MLRA) method and ANN method were obtained at 0.9224, and 0.9449, respectively. The applicability of the multiple linear regression analysis method and ANN method in calculating and predicting tunnel face area after blasting were validated based on a comparison with the results of the tunnel face area after blasting in practice.

Copyright © 2022 Hanoi University of Mining and Geology. All rights reserved.

*Corresponding author
E-mail: nguyenthanh.xdctn47@gmail.com
DOI: 10.46326/JMES.2022.63(3).06
1. Introduction

In underground excavation, overbreak and underbreak had long been recognized as the principal cause of hazards and deterioration costs in underground construction management, and as such numerous related research projects have been conducted. Many research papers have been devoted to clarifying the overbreak and underbreak phenomenon, but they are still unable to explain the exact occurrence process (Mahtab et al., 1997; Monjezi and Dehghani, 2008; Jang et al., 2013; Mohammad et al., 2017; Esmaeili et al., 2014; Koopialipoor et al., 2017; Mohammadi et al., 2018; Mottahedi et al., 2018). The area of tunnel face after blasting includes overbreak and underbreak phenomenon, and this value must be accurately determined. If overbreak occurred, undesirable effects will be showed up in the process of underground construction. The area of the tunnel face after blasting is large, which could increase the volume of soil/rock that needs to be transported after blasting, as well as increase the volume of the structure support that needs to be installed for the tunnel. Conversely, if an underbreak occurs, the construction progress of the tunnel is reduced, and the volume of work in underground construction is greatly increased. The safety reduction of working space in underground constructions, and time-consuming due to the creation of unproductive works are such as these negative effects. Based on the foregoing, overbreak and underbreak must be predicted and then controlled.

According to the former research, factors causing overbreak and underbreak can be classified into two groups. In 2008, Mandal et al. had given the different terminology of factors causing overbreak and underbreak phenomenon in underground excavation by blasting method, in these geological and blasting factors were the principal groups influencing the area of tunnel face after blasting.

In this paper, the parameters of geological and parameters of explosive with the tunnel in the blasting design of a tunnel were used and the study focuses on the effects of these parameters on the area of the tunnel face after blasting. The rock mass rating (RMR), specific charge (SC), average boreholes length (L), and design tunnel face area (Sd) were collected through 39 blasting sections, and the area of tunnel face after blasting data was individually investigated. Various methods have been applied in engineering for the area of tunnel face after blasting prediction. In this study, the multiple linear regression analysis (MLRA) method and artificial neural network (ANN) method are used to predict potential the area of the tunnel face after blasting. The geological data sets and the explosive, tunnel datasets are put as input parameters and encountered the area of the tunnel face after blasting results are used as output parameters to ANN models and simultaneously to MLRA. Consequently, the optimum area of tunnel face after blasting predicting model is selected by comparing measured and predicted the area of tunnel face after blasting and the coefficient of determination (R²) of each proposed model. This model can be used for other tunnels that have similar geological conditions.

2. Case study and data in the paper

Deo Ca tunnel is a tunnel connecting the two provinces of Khanh Hoa and Phu Yen (Figure 1). The point of the northern tunnel entrance is at Km 1353+500, National Highway 1A, in Hao Son Bac village, Hoa Xuan Nam commune, Dong Hoa town, Phu Yen province. The point of the southern tunnel entrance is at Km 1371+525, National Highway 1A, in Co Ma village, Van Tho commune, Van Ninh district, Khanh Hoa province. In this paper, there are 39 datasets collected and used to build and check the accuracy of models that be used to predict, and
calculate the area of the tunnel face after blasting, with 80% datasets were used for training models (27 datasets) and 12 datasets were used for testing models (20%). In the datasets above, there are 4 parameters kinds: the rock mass rating (RMR), the design tunnel face area \( S_d \), the average boreholes length \( L \), and the specific charge of explosive \( SC \). The output data is the area of the tunnel face after blasting \( SA \). The values of parameters are shown in Table 1.

3. Calculation and prediction methods of tunnel face area after blasting

3.1. Multiple linear regression analysis model (MLRA)

One of the methods used to describe and analyze the variance of dependent variables considering independent variables is the Multiple linear regression analysis (MLRA) method.

In this paper, the rock mass rating (RMR) of the rock mass on the tunnel face, the design tunnel face area \( S_d \), the specific charge \( SC \), and the length of the average borehole \( L \) are independent variables and the dependent variable, this is the area of the tunnel face after blasting \( SA \).

By using the SPSS software, 39 datasets were obtained in 39 blasting sections, and the area of the tunnel face after blasting data was individually investigated tunnel. In this data, 80% of datasets had been used to train the prediction model for tunnel face area after blasting and 20% datasets for checking the accuracy and performance of the model. The governing relationship between the dependent variable (the area of tunnel face after blasting) and independent variables (factors causing the area of tunnel face, including the rock mass rating RMR; the design tunnel face area \( S_d \), the specific charge \( SC \), and the length of the average borehole \( L \) ) was generated and presented in the formula:

\[
SA = 5.88 + 0.604 \times L + 0.981 \times S_d + 2.047 \times SC - 0.089 \times RMR
\]  

The coefficient of determination \( R^2 \) has been used to check the accuracy of the multiple linear regression analysis models. The value of \( R^2 \) ranges from 0 to 1, in which it is often specified that \( R^2 \) must have a value greater than or equal to 0.5, then the multiple regression equation represents the relationship between the independent variable and the dependent variables' acceptable accuracy (Rodríguez, Benítez-Parejo, 2011; Monjezi et al. 2012). After testing for \( R^2 \), the model must be checked for autocorrelation and multicollinearity problems. In this study, the coefficient of determination of the model \( R^2 = 0.9224 \) and the adjusted coefficient of determination \( R^2 = 0.908 \), satisfying the accuracy of the model.

To ensure the accuracy of the MLRA model. The multicollinearity problem has to be controlled. It occurs when the correlation among the independent variables is strong. Hence, the standard errors of the coefficients are increasing, which leads to an erroneous conclusion of multiple regression analysis. Variance Inflation Factor (VIF) and Tolerance are commonly used to verify the multicollinearity problem. VIF measures how much the variance of the estimated coefficients increases over the case of no correlation among the independent variables. If two independent variables are not correlated, then all the VIFs will be equal to 1. Generally, if the value of VIF is over ten, the model demonstrates a strong multicollinearity problem. Tolerance is an inverse number of VIF, and if the tolerance value is less than 0.1, it is acknowledged that there is a multicollinearity problem associated with the model. The case of the multicollinearity problem that appears in the study model, must eliminate the suspicious variable (Mohammad et al., 2012).

In this paper, with data in Table 2, the parameter values in the MLRA model were presented.

It can be seen that the multicollinearity problem does not occur in the MLRA model built above. The coefficient of determination \( R^2 = 0.9224 \).

3.2. ANN model

An artificial neural network (ANN) can be identified as a simplified mathematical model of reasoning based on the human brain. ANN is able to determine the complex relationship among variables for the simulation of one (or more)
Table 1. The limit and average values of parameters in models.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Unit</th>
<th>Category</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The average boreholes length</td>
<td>L</td>
<td>m</td>
<td>Input</td>
<td>1.0</td>
<td>3.0</td>
<td>2.0815</td>
<td>0.65518</td>
</tr>
<tr>
<td>The design tunnel face area</td>
<td>Sd</td>
<td>m²</td>
<td>Input</td>
<td>49.26</td>
<td>64.855</td>
<td>54.2289</td>
<td>6.52140</td>
</tr>
<tr>
<td>Specific charge</td>
<td>SC</td>
<td>kg/m³</td>
<td>Input</td>
<td>0.45</td>
<td>2.07</td>
<td>1.3556</td>
<td>0.43075</td>
</tr>
</tbody>
</table>

Table 2a. The coefficients values in the MLRA model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Unit</th>
<th>Category</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The rock mass rating</td>
<td>RMR</td>
<td>-</td>
<td>Input</td>
<td>8.0</td>
<td>72.0</td>
<td>49.9630</td>
<td>16.7825</td>
</tr>
<tr>
<td>The actual area of the tunnel face after blasting</td>
<td>SA</td>
<td>m²</td>
<td>Output</td>
<td>51.221</td>
<td>70.131</td>
<td>58.6592</td>
<td>6.76450</td>
</tr>
</tbody>
</table>

Table 2b. The coefficients values in the MLRA model.

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>Std.Error</th>
<th>Beta</th>
<th>t</th>
<th>Sig</th>
<th>Collinearity statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.88</td>
<td>3.427</td>
<td>-</td>
<td>1.716</td>
<td>0.100</td>
<td></td>
</tr>
<tr>
<td>The average boreholes length (L)</td>
<td>0.604</td>
<td>1.272</td>
<td>0.059</td>
<td>0.475</td>
<td>0.639</td>
<td>0.233</td>
</tr>
<tr>
<td>Area of the tunnel face (Sd)</td>
<td>0.981</td>
<td>0.068</td>
<td>0.946</td>
<td>14.405</td>
<td>0.000</td>
<td>0.819</td>
</tr>
<tr>
<td>Specific charge (SC)</td>
<td>2.047</td>
<td>1.920</td>
<td>0.130</td>
<td>1.066</td>
<td>0.298</td>
<td>0.236</td>
</tr>
<tr>
<td>Rock Mass Rating (RMR)</td>
<td>-0.089</td>
<td>0.065</td>
<td>-0.220</td>
<td>-1.364</td>
<td>0.135</td>
<td>0.205</td>
</tr>
</tbody>
</table>

output(s) (Holland, 1975; Hecht-Nielsen, 1987; Specht 1991; Zorlu et al., 2008).

An ANN model is defined by the following important parameters: the transfer function, learning rule, and network architecture (Simpson P.K 1990;). One of the most commonly used artificial neural models is the multilayer perceptron (MLP), which is the feedforward neural network model and typically contains an input layer of source neurons, at least one hidden layer of computational neurons, and one output layer (Figure 2). Each of these layers has its specific function. The input layer accepts inputs from the outside world and distributes them to the subsequent layers. Features hidden in the input patterns are detected by the neurons in the hidden layer. The output layer exploits these features to determine the output pattern (Whitley, 1993a; Bounds et al., 1998; Danial et al., 2015; HasaniPanah et al., 2016a, 2017a; Liu and Hou, 2019; Nguyen et al., 2019; Moayedi et al., 2019).

According to the author (Basheer and Hajmeer 2000; Poli et al., 2007; Hajihassani et al., 2014; Gordan et al., 2016; Alsarraf et al., 2019; Longqi et al., 2019; Lawal and Kwon, 2022), the BP is the most popular learning method among a vast number of MLP learning algorithms. In the BP method, the input data are presented to the input layer to be propagated through the network until an output is generated:

\[ X = \sum_{i=1}^{m} x_i \times w_i - a \]  

(2)

Where: \( X \) - output data at the output layer, \( x_i \) - value of the \( i \)th input, \( w_i \) - weight of the \( i \)th input, respectively, \( m \) - number of input data, \( a \) - the threshold applied to the neuron that is processing the data.

In this study, the transfer function of the form TANSIG function is used because of the advantages of this form of transfer function such as fast model convergence, reflecting variable values in the range \([-1, 1]\). To be able to use the TANSIG transfer function as mentioned above in the ANN model, it is necessary to normalize the data of the input variables and output variables in the ANN model according to the following formula:

\[ X_n = \frac{(X - X_{min})}{(X_{max} - X_{min})} \]  

(3)
Where: $X_n$ - the normalized value of the variable, $X$ - the initial value of the variable, $X_{\text{max}}$ - the initial maximum value of the unnormalized variable, $X_{\text{min}}$ - the initial minimum value of the unnormalized variable.

In the ANN, the most difficult thing is to determine the best network architecture with the number of hidden layers and the most reasonable number of neurons in the hidden layer, from there will get the most accurate results for the ANN (Simpson, 1990). There have been many publications successfully using Levenberg-Marquardt back-propagation training algorithm for the ANN. Some authors in their publications have presented that, with a hidden layer, an ANN can approximate any continuum function (Basheer et al., 2001; Dey, et al., 2012; Armaghani, et al., 2014; 2017; Mohsen et al., 2015; Shahnazar, et al., 2017). Due to the advantages of having only one hidden layer of neurons in the ANN model, such as: reducing the complexity of the model, and reducing the processing time of the model's results, etc., therefore, the ANN model built in this study, the number of hidden layers in the model is chosen equal 1 (Figure 3). According to some researchers, the number of neurons in the hidden layer of the ANN model is determined through the number of neurons in the input layer and the number of neurons in the output layer of the model. Hecht-Nielsen et al., 1987 mentioned the number of neurons in the hidden layer determined by the formula: $N \leq \frac{(N_i + N_o)}{2}$ where $N_i$ is the number of neurons in the input layer. According to Ripley, 1993, the number of neurons in the hidden layer of ANN model satisfies $N \leq \frac{(N_i + N_o)}{3}$ with $N_i$ is the number of neurons in the input layer, $N_o$ is the number of neurons in the output layer, According to Wang, $N \leq \frac{2*N_i}{3}$.

In this paper, in order to determine the reasonable number of neurons for the hidden layer of the ANN model, the authors had been built a series of ANN models corresponding to the same data.

Each model corresponds to a specific number of neurons in the hidden layer. After building the model, determine the accuracy of the ANN model in calculating and predicting the tunnel face area after blasting SA based on two indicators, the coefficient of determination $R^2$ and mean squared error MSE. The more accurate the model, the larger the $R^2$ and the smaller the MSE value. Each model will be repeated 5 times to ensure the representativeness of the results obtained.
Where the mean square error MSE is determined through the equation:

\[ MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - y'_i)^2 \]  \hspace{1cm} (4)

The coefficient of determination \( R^2 \) is determined through the equation:

\[ R^2 = \left[ \frac{\sum_{i=1}^{N}(y - \bar{y})(y' - \bar{y}')} {(\sum_{i=1}^{N}(y - \bar{y})^2)(\sum_{i=1}^{N}(y' - \bar{y}')^2)} \right]^2 \]  \hspace{1cm} (5)

Where: \( N \) - the number of data at the input layer; \( y_i \) - the \( i \)th actual measured value; \( y'_i \) - the \( i \)th predicted value; \( \bar{y} \) - the actual measured mean and \( \bar{y}' \) - the value average prediction.

Based on the results obtained in Tables 3, 4 and Figures 4, 5, conclusions can be drawn: the appropriate number of neurons in the hidden layer is \( n=5 \). Thus, the ANN model used to predict and calculate the area of tunnel face after blasting SA has a structure of 4x5x1 with 4 neurons in the input layer, 5 neurons in the hidden layer and 1 neuron in the output layer.

On the base of the values of the coefficient of determination \( R^2 \) as well as the prediction results of the area of tunnel face after blasting SA in the MLRA model and ANN model (Figures 6 to 11), the following comments can be made: the predicted value of the area of tunnel face after blasting obtained in the artificial neural network ANN model has a deviation from the actual value of the tunnel face after blasting, which is smaller.

Table 3. The correlation coefficients \( R \) for several ANN models with different hidden nodes.

<table>
<thead>
<tr>
<th>The model number</th>
<th>The results of the models</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td>1</td>
<td>0.97094</td>
<td>0.8726</td>
</tr>
<tr>
<td>2</td>
<td>0.9675</td>
<td>0.97951</td>
</tr>
<tr>
<td>3</td>
<td>0.95341</td>
<td>0.96047</td>
</tr>
<tr>
<td>4</td>
<td>0.98502</td>
<td>0.95128</td>
</tr>
<tr>
<td>5</td>
<td>0.91831</td>
<td>0.94033</td>
</tr>
<tr>
<td>6</td>
<td>0.93647</td>
<td>0.93976</td>
</tr>
<tr>
<td>7</td>
<td>0.93821</td>
<td>0.94948</td>
</tr>
<tr>
<td>8</td>
<td>0.92508</td>
<td>0.91615</td>
</tr>
</tbody>
</table>

Table 4. The mean squared error (MSE) for several ANN models with different hidden nodes.

<table>
<thead>
<tr>
<th>The model number</th>
<th>The results of the models</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train</td>
<td>Test</td>
</tr>
<tr>
<td>1</td>
<td>0.02248</td>
<td>0.11723</td>
</tr>
<tr>
<td>2</td>
<td>0.0277</td>
<td>0.02194</td>
</tr>
<tr>
<td>3</td>
<td>0.0613</td>
<td>0.04784</td>
</tr>
<tr>
<td>4</td>
<td>0.01107</td>
<td>0.02066</td>
</tr>
<tr>
<td>5</td>
<td>0.07226</td>
<td>0.05733</td>
</tr>
<tr>
<td>6</td>
<td>0.06715</td>
<td>0.09177</td>
</tr>
<tr>
<td>7</td>
<td>0.03311</td>
<td>0.03099</td>
</tr>
<tr>
<td>8</td>
<td>0.06337</td>
<td>0.1126</td>
</tr>
</tbody>
</table>
Figure 4. Relationship between the average coefficient of determination $R^2$ and the number of neurons present in hidden layer of the ANN model.

Figure 5. Relationship between the average mean square error MSE and the number of neurons in hidden layer of the ANN model.

Figure 6. Comparison between measured and predicted SA for training datasets using MLRA model and ANN model.

Figure 7. Comparison between measured and predicted SA for testing datasets using MLRA model and ANN model.

Figure 8. Measured and predicted the area of tunnel face SA obtained via the MLRA model for training datasets.

Figure 9. Measured and predicted the area of tunnel face SA obtained via the MLRA model for testing datasets.
than the accuracy deviation of values of the tunnel face area after blasting, respectively obtained in the MLRA model. Based on the coefficients of determination values $R^2$ of the MLRA model and ANN model for training and testing data, respectively, a conclusion can give as follows: with the ANN model, the coefficient of determination $R^2$ of the ANN model for the training dataset has the highest value.

4. Conclusion

In this paper, the multiple linear regression analysis MLRA model combined with an ANN model was built to calculate and predict the value of the tunnel area after blasting. By using SPSS software to build MLRA models as well as using Matlab 2019b software to build ANN models based on actual data sets obtained during the construction Deo Ca tunnel (including 39 datasets), the paper evaluated the results obtained in each method and the following conclusions can be drawn:

- With the associated of significance and coefficients related to the independent variables, it is found that the parameters (independent variables) of the models, including specific charge (SC), the design tunnel face area ($S_d$), and the average hole length (L), the rock mass rating (RMR) play an important role in the generation of the area of tunnel face after blasting;
- The ANN model had an optimum architecture that was 4x5x1, with 4 neurons in the input layer, one hidden layer with 5 neurons, and one neuron in the output layer;
- Both the MLRA model and ANN model can calculate and predict the tunnel area after blasting with high accuracy, the coefficients of determination of both models are greater than 0.9 (their values are 0.9224, and 0.9449, respectively). The result of the ANN model and the MLRA model had an acceptable prediction performance;
- Based on the results comparing the coefficient of determination $R^2$ in the multiple linear regression MLRA model and ANN model, it is possible to realize the ANN model could predict, calculate the area of the tunnel face after blasting with higher accuracy than the MLRA model, and can use the ANN model to predict and calculate tunnel face area after blasting with high accuracy.

Acknowledgments

This research is supported by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 17/2020/STS02, the Vietnamese Ministry of Education and Training under grant number B2022-MDA-09, and the Hanoi University of Mining and Geology under grant number T21-30.

Author contributions

Thanh Chi Nguyen - data curation, formal analysis, funding acquisition, investigation, methodology, writing original draft; Anh Ngoc Do - conceptualization, supervision, writing review & editing; Vi Van Pham - software, validation; Gospodarikov Alexandr - writing review.

References


