Predicting Shear Wave Velocity of Soil Using Multiple Linear Regression Analysis and Artificial Neural Networks

OMOLBANIN ATAEE[1], NASER HAFEZI MOGHADDAS[2,], GHOLAM REZA LASHKARIPOUR[3], MEHDI JABBARI NOOGHABI[4]*

[1]*Dept. of geology, Ferdowsi University of Mashhad, Mashhad, Iran*

Email: om.ataee@stu.um.ac.ir

[2]*Dept. of geology, Ferdowsi University of Mashhad, Mashhad, Iran*

Email: nhafezi@um.ac.ir

[3]*Dept. of geology, Ferdowsi University of Mashhad, Mashhad, Iran*

Email: lashkaripour@um.ac.ir

[4] *Dept. of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran*

Email: jabbarinm@um.ac.ir

**Corresponding author (Naser Hafezi Moghaddas):*

Email: nhafezi@um.ac.ir

Address: Department of geology, Faculty of science, Ferdowsi Univ. of Mashhad, Mashhad, Iran

Tel: 00989123730591

Omolbanin Ataee is a Ph.D. student of the department of geology at Ferdowsi University of Mashhad, Iran. Her main areas of research interest are soil mechanics, geostatistics and environmental geology.

Naser Hafezi Moghaddas is currently a professor of Engineering Geology at Ferdowsi University of Mashhad. His research interests focus on the landslide, site effect and micro zonation of earthquake, slop stability.

Gholam Reza Lashkaripour is a professor engineering geology at Ferdowsi University of Mashhad. He teaches courses such as environmental geology, hydrology, advanced engineering geology, rock mechanics, groundwater and geotechnical problems, geological hazards, site investigation, and soils improvement at graduate and post graduate levels.

Mehdi Jabbari Nooghabi is an assistant professor in the department of statistics at Ferdowsi University of Mashhad. He taught different courses such as statistical applications in management, statistical quality control, time series, research method and statistical consulting, statistics for managers, business mathematics and statistics, statistical analysis, special topics, advanced applied statistics, statistics meteorology, linear models and their application in genetic improvement of farm animals, advanced methods of statistical inference, industrial statistics, and multivariate discrete and continues analysis at graduate and post graduate levels.

Predicting Shear Wave Velocity of Soil Using Multiple Linear Regression Analysis and Artificial Neural Networks

Abstract

In this paper, the correlation between shear wave velocity and some of the index parameters of soils including standard penetration test blow counts (*SPT*), fine-content (*FC*), soil moisture (*W*), liquid limit (*LL*) and depth (*D*) is investigated. The study attempts to show the application of artificial neural networks and multiple regression analysis in the prediction of the shear wave velocity (V_S) value of soils.

New predicting equations are suggested to correlate V_S and mentioned parameters based on a dataset collected from Mashhad city in the north east of Iran. The results suggest that better and more exact correlations in the estimation of V_S are acquired when *ANN* method is used. The predicted values using *ANN* method are checked against the real values of V_S to evaluate the performance of this method. The minimum correlation coefficient obtained in *ANN* method is higher than the maximum correlation coefficient obtained from the *MLR*. In addition, the value of estimation error in the *ANN* method is much less than the *MLR* method indicating the higher confidence coefficient of the *ANN* in estimating the V_S of soil.

Keywords: Shear wave velocity, *SPT*, depth, Fine-content, Artificial Neural Network, Multiple Linear Regression, and Mashhad.

1. Introduction

Shear wave velocity (V_S) is a fundamental parameter in defining the dynamic properties of soils, dynamic site response evaluation, and dynamic site characterization [1, 2]. The profile of V_S in the ground has been considered as the most reliable predictor of site-dependent properties from a seismic action in stable sites [3]. V_S is often measured by in situ methods in low strain levels, so this measured V_S can be employed to determine the maximum shear modulus (G_{max}) of soil deposit in different depths [1]. G_{max} is an essential input parameter for analyzing the dynamic stability of slopes, dams, embankments, etc. [4].

Although it is preferable to determine V_S directly through field tests, conducting these tests is mostly not feasible due to economic considerations, lack of space in urban areas, lack of specialized personnel, and high noise levels in all situations [4-8]. In the absence of in situ dynamic tests data, it is common worldwide to calculate V_S through reported empirical relationships between V_S and other geotechnical soil properties such as *SPT*, *CPT*, dry density, etc. [4]. *SPT* is one of the most economical and commonly employed in situ tests having very strong relationships with many of geotechnical soil properties. Many studies have proposed statistical relationships between V_S and *SPT* blow counts [1, 3-6, 8-20].

Generally, the relationships between V_S and *SPT* show considerable dispersion, which is probably due to the different methods employed in measuring the V_S and *SPT N* value, as well as the geotechnical and geological conditions specific to any area. Another reason for low accuracy of these relationships is the type of regression analysis employed [21]. Traditionally, statistical methods such as simple and multiple regression analyses are used in geotechnical engineering to create predictive models. All of conventional regression methods have limitations. Besides, empirical methods are not applicable for complicated and non-linear problems [22].

Artificial Neural Networks (*ANN*) is an over-simplified simulation of human brain made up of simple processing units called neurons. This system is able to learn and generalize experimental data, even when the data are noisy, incomplete and with a non-linear nature [23, 24]. The main advantage of *ANN* is that unlike the conventional statistical models, it does not require any prior knowledge related to the kind of relationship between input and output data. *ANN*s are also able to function very well in situations with limited data accessibility [25].

In recent years, artificial intelligent (*AI*) methods have been widely used for predicting purposes [26, 27]. So far, neural networks have been used for estimating and predicting some of the geotechnical properties of soil such as estimation of soil compaction parameters [28, 29], compressive and shear strength of soils [22, 30, 31], prediction of soil permeability [29, 32], prediction of *CBR* in fine-grained soils [33], prediction of

Compressibility indices of saturated clays [34-36], pile bearing capacity [37, 38], prediction of soil settlement [39], soil liquefaction [40-43], and slope stability analysis [44-46]. Researchers have also used neural network models to predict V_S value in oil wells [26, 47-50]. In addition, *ANN*s have been used to estimate and predict V_S values of soils using geotechnical soil properties such as *CPT* [51-53].

The present study aims to develop an optimal model for prediction of V_S of soils in Mashhad city based on the parameters of *SPT*, depth, fine content, liquid limit, and percentage of soil moisture. In this study, a multilayer perceptron with feed forward back propagation is used for modeling V_S in soil. The best neural network model is found and selected through analysis of performance of different models (with different hidden layers and different neurons in each layer). The purpose of this study is to identify properties of soil that have a more efficient role in predicting the V_S of soil; it also attempts to compare the capability of neural networks and the multiple regression technique in predicting the V_S value using the variables mentioned above.

2. Study area

This study was conducted in Mashhad city, the second most populated city in the center of Khorasan Razavi province in Iran. Mashhad is located at latitude 36.10°-36.24°*N* and longitude 59.25°-59.43°*E* in the northeast of Iran. It is situated on Mashhad plain, which is covered with thick Quaternary alluvial sediments. Kashafrood River is the main drainage system of Mashhad plain as well as the streams originating from the southern parts of Mashhad (Figure 1).

From the seismotectonical perspective, Mashhad is situated between two folded-thrusted mountains (Koppe-Dagh in northeast and Binalood in south-west). According to Berberian et al. [54] there were intense earthquake activities in the area in the past centuries, especially in the 18th century. The existence of active faults within a short distance and on the two sides of Mashhad plain indicates that this area has a high potential for earthquake. Shandiz fault, Kashafrood fault, Toos fault, South of Mashhad fault, the north of Neishaboor fault and Kheirabad fault are the main active faults in this area [55].

Mashhad city is an earthquake prone area and according to the Iranian Code of Practice for Seismic Resistance Design of Buildings (Standard No.2800) [56], this city is situated in the high earthquake risk zone, with 0.30-0.35*g* maximum acceleration in return period of 475 years. These issues indicate the seismic vulnerability of Mashhad city, so an accurate estimation of V_S for this city is required.

3. Materials and methods

3.1. Regression

Regression analysis is one of the analytical instruments widely applied for investigating relationships between a dependent variable and a set of independent (predictor) variables. Regression analysis is either linear or non-linear. In linear regression, data are modeled using linear independent variables or predictive functions. In non-linear regression, data are modeled using a function which is a non-linear combination of model parameters. This type of regression is dependent on one or more independent variables. Regression analysis is one of the common methods of creating predictive models between V_S and soil geotechnical properties, including *SPT* and *CPT*. In addition to *SPT*, this paper aims to study the influence of fine content, soil moisture, liquid limit, and soil depth on estimating the V_S value. Therefore, multiple linear regression analysis (*MLR*) will be employed.

3.2. Multiple Linear Regression analysis (MLR)

Simple linear regression is a useful technique to predict a response based on a single predictor variable. However, in practice it often happens that there is more than one predictor. *MLR* is used when there are more than one explanatory variable in a model which can help explain or predict the response variable; therefore, all of these explanatory variables are put into the model to do a multiple linear regression analysis. Then, the *MLR* equation is as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_P X_P + \varepsilon \qquad (1)$$

Where X_p represents the p^{th} predictor and β_p quantifies the relation between that variable and the response. We interpret β_p as the mean effect on Y of a one unit increase in X_p, holding all other predictors fixed. The regression coefficients $\beta_0, \beta_1 . . . \beta_p$ in (1) are unknown and must be estimated using the least squares approach as is the case in simple linear regression [57].

3.3. Artificial Neural Network (ANN)

ANN is a massively parallel-distributed information processing system with certain performance characteristics, which simulate the biological neural networks of the human brain [58]. A neuron is the basic component of the neural networks that accepts and sums inputs up, applies a transfer function, which is normally nonlinear, and gives the result that is either as a model prediction or as input to other neurons. An artificial neural network is a combination of many such neurons connected in a systematic way. Neurons with the same properties in an *ANN* are ordered in groups called layers [33]. One-layer neurons are connected to the neurons of the adjacent layers (and not to the neurons of the same layer). The strength of connection between the two neurons in adjacent layers is recognized through the strength of connection or weight.

An *ANN* usually has three layers, one input layer, one hidden layer, and one output layer. As *ANN*s have the ability to work with incomplete data and have an error tolerance, they can easily produce models for complicated problems. Especially, for semi-structural or non-structural problems, neural network models can provide very successful results. Furthermore, they are faster and more reliable than the traditional methods [23].

3.4. Multi-Layer Perceptron (MLP)

*MLP*s, also known as feed-forward neural networks, typically trained with back propagation algorithm are usually used for prediction. Neurons in such networks are arranged in different layers (typically one input layer, one or more hidden layers and one output layer) each of which is interconnected to its preceding and following layers. Figure 2 depicts a feed forward three-layer *ANN* with the description of input and output layers employed in the current study. Connections between neurons have weights associated with them. These weights determine the strength of influence that one neuron can have on another. From the input layer and through the processing layer(s), information flows to the output layer to generate predictions. During training, the network learns to generate better and better predictions through adjusting the connection weights to match predictions to target values for specific records.

Determining the network architecture requires determining the optimum number of hidden layers between input and output layers and the optimum number of neurons in each hidden layer. That is one of the most important and most complicated parts of designing neural networks, as there is no single theory or accepted rule for determining optimal network architecture [59-61]. The number of hidden layers and their neurons is mostly determined by trial and error [62, 63].

3.5. Data Preparation and Normalization

The dataset used in this study was collected from geotechnical and geophysical reports from civil engineering projects done across Mashhad city by consulting engineering companies. Data related to 85 drilled boreholes were employed in data analysis. As a complete dataset of six variables was required for this study, finally, 185 soil samples were used for regression analysis and designing neural network and its accuracy evaluation. The model input variables chosen for the present study were *SPT, LL, W, FC,* and *D* and the target or output variable was the V_S of the soil.

In most of the datasets, there is a lot of variability in the scale of range fields. Therefore, to nullify the effect of scale, range fields are transformed to have the same scale for all of them. In this situation, normalization can speed up the training process and improve the accuracy of the output model. Range fields are rescaled in Clementine to have values between 0 and 1. In this study, before inserting data into *ANN*, the input and output datasets were normalized through the following formula [64]:

$$x_{normalized} = \frac{x_{actual} - x_{min}}{x_{max} - x_{min}} \qquad (2)$$

Where $x_{normalized}$ is the normalized value of the observed variable, x_{actual} is the real value of the observed variable, and x_{max} and x_{min} are the maximum and minimum value of observed variable in the dataset, respectively. When the function of the network is complete, network outputs are post-processed so that data are converted to non-normalized units [28]. For *ANN* modeling, data are divided randomly into three categories of training, testing, and validation [60]. The network is trained by the first category of data. The training set is also used for adjusting the weights of the connections. The validation set is used to test the performance of the network in different stages of training. When the training is successful, the testing set is used to evaluate the performance of the model.

The dataset collected from the Mashhad city region were first analyzed for possible relationships between the parameters, and those variables which seemed likely to be influential in predicting V_S were separated. Finally, five main parameters including *SPT*, *LL*, *W*, *FC*, and *D* were considered as input parameters in *MLR* and *ANN* models. In designing neural network, data were divided into training, testing and validation sets. From a total of 185 data sets used in this study, 80% (137 samples) were employed for training the model, 10% (19 samples) were employed for testing the model, and 10% (29 samples) were used for validation in *ANN* analysis. All data were employed in regression analysis. Table 1 shows descriptive statistics related to the input and output parameters for all sets.

3.6. Performance Evaluation Criteria

For the assessment of methods, the obtained results from each model (*MLR* and *ANN*) were evaluated based on different criteria. There are many criteria for assessing the performance of models. In this section, the efficiency criteria used in this study are presented and evaluated. There are four criteria here: the correlation coefficient (R), coefficient of determination (R^2), root mean square error (*RMSE*), and mean absolute error (*MAE*). Each of the above criteria used in this study was computed through the following equations:

Pearson Correlation Coefficient (R): R can be used to estimate the correlation between model and observations.

$$R = \frac{\sum_{i=1}^{n}(m_i - \overline{m}) \cdot (p_i - \bar{p})}{\sqrt{\sum_{i=1}^{n}(m_i - \overline{m}) \cdot \sum_{i=1}^{n}(p_i - \bar{p})^2}} \qquad (3)$$

Where m_i is the measured value, P_i is the predicted value, $\bar{m}$ is the mean of measured values and $\bar{P}$ is the mean of predicted values.

Coefficient of determination or the square of the Pearson correlation coefficient (*R^2*): *R^2* describes how much of the variance between the two variables is described by the linear fit.

Root mean square error (*RMSE*): The *RMSE* of a model prediction is defined as the square root of the mean squared error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(m_i - p_i)^2}{n}} \qquad (4)$$

Where *n* is the number of data presented in the database.

Mean Absolute Error (*MAE*): The mean absolute error (*MAE*) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The mean absolute error is given by:

$$MAE = \frac{\sum_{i=1}^{n}|m_i - p_i|}{n} \qquad (5)$$

4. Results and Discussion

4.1. Regression Analysis

As previously mentioned, five independent variables for multiple regression analysis were selected. At first, the relationships between the V_S and each of the independent variables were studied. The relationship between V_S and *SPT* as well as V_S and depth has a power-law form [1-20]. So, the most preferred functional form of relation between *SPT* and V_S proposed in literature ($V_S = a \cdot N^b$) has been used as the main format for *MLR* analysis. This equation is given below:

$$V_S = a \cdot N^b \cdot D^c \cdot FC^d \cdot W^e \cdot LL^f \qquad (6)$$

In this equation, *N, D, FC, W,* and *LL* represent *SPT*, depth, fine-content, moisture content, and liquid limit, respectively and *a* to *f* are coefficients that should be determined by regression. The power law form of this model allows us to write it as follows:

$$LogV_S = Loga + bLogN + cLogD + dLogFC + eLogW + fLogLL \quad (7)$$

In this case, the linear regression can be used to determine the constant values. *MLR* analysis was performed on 31 possible combinations of independent variables. After comparing the results, nine combinations whose coefficient of determination exceeded 0.5 were selected to obtain the best model to govern V_S. The combination of input variables in different models in this study is given in Table 2. The *MLR* regression analysis was performed using SPSS software and the criteria for performance evaluation were calculated for each combination as shown in Table 3.

By comparing the above results it is found that *C-3* with a higher correlation coefficient and coefficient of determination (0.856 and 0.733, respectively) and smaller values of *RMSE* and *MAE* is the best model for predicting the value of V_S of soil. It can be seen that the combination of the three parameters, including *D*, *SPT*, and *FC* (silt and clay) shows the better correlation with V_S. Figure 3 shows the scatter plot of V_S values predicted by *MLR* and its measured values in the field.

4.2. Artificial Neural Network Development

In this study, a *FFBP-based ANN* solver (Clementine data mining system) was used for designing and testing of *ANN* models. Clementine is a preeminent data mining toolkit widely used in academic research and industrial applications. To apply Clementine *ANN* solver, like other *FFBP*-based softwares, diverse network architectures should be experienced to obtain the best result. The first step in this process is to select input and output variables which for this study were chosen in prior sections.

In the next step, the number of hidden layers and the number of hidden neurons in each layer are defined for the model. There is no obvious solution for determining hidden layers and hidden neurons for the *ANN* network. Although Zhang et al. [65] suggested that the optimum number of hidden layers in *FFBP* architecture is mostly one or two. Some researchers have also suggested that between *n* and *2n* hidden neuron is enough for *FFBP* models [66].

Generally, there are two fundamental approaches for constructing a *FFBP* network [67]. In a method called additive or constructive, the model starts with a minimal network consisting of a single hidden layer and gradually hidden layers and hidden neurons are added and the effectiveness of the obtained model is evaluated using the evaluation instruments. In the second approach, the model starts with a very large network and pruning algorithm is used to reduce the size of the model [68].

Clementine provides both approaches: The dynamic method uses the additive approach. In this state, the network topology changes during the training phase by the Neurons that are added so that the network obtains an optimum performance. This is while the system also monitors lack of improvement and overtraining. This process continues until no improvement can be achieved from the future growing of the model. Conceptually, the prune method is different from the dynamic method. The prune method starts with a large network and then gradually prunes it by removing the unhelpful neurons from the input and hidden layers. There are two stages for pruning, including pruning the hidden neurons and pruning the input neurons. The two-stage process repeats again and again until the overall stopping criteria are met. These two stages are described below. The training process in this approach discards the weakest hidden neurons and selects the optimal size of the hidden layers. When it is clear that one hidden layer is enough, another option called Quick can be used which in a simple mode creates a network with one hidden layer and attempts to determine the number of hidden neurons which give the best results. The stopping rules in Clementine include a measure of desirable accuracy, the number of cycles which the model is run, and the real amount of time in which the model is allowed to run.

In the present study, combinations of these approaches were used to reach the best results. During the construction of these models the Prevent Overtraining parameter was always in selection mode to be avoided the overtraining of the model. Input and output data were normalized before being inserted in the network. To design the neural network, the dataset was randomly divided into three discrete sets called training, testing and validation (80%, 10% and 10%, respectively). Only data concerning the training set are used by neural network to learn the existing patterns in the data and optimize model parameters. While training the network, the optimum numbers of neurons in the hidden layer and the learning rate are calculated. The training phase is stopped when the variation of error became sufficiently small. After building a model using the Training Set, the performance of the trained model must be validated using new data. To get a more realistic estimation of how the model would perform with unseen data, we must allocate a part of the data which are not trained during the training process to this purpose. This set of data is known as the Validation Set. The testing set is the unseen data that is used to evaluate the performance of various candidate model structures and to test the network's generalization.

In this study, data analysis with neural networks was performed using the SPSS Clementine software. The *ANN* analysis was also carried out on nine selected combinations of independent variables in the previous sections. Different creating network methods were employed for each combination and the method which could estimate the value of the target parameter with a higher accuracy and the least number of hidden layer and hidden neurons was selected.

The performance criteria required for evaluate the accuracy of the designed models were computed for the training, testing and validation stages (Table 4). Based on the model performance in validation stage, the best *ANN* model was determined. By comparing the evaluation criteria in Table 4, the combination (*C-3)* with a structure of 3-2-3-1 which has the highest value of correlation and determination coefficient and lowest value of *RMSE* and *MAE* in comparison with other combinations was selected as the optimal model in neural network analysis. It is observed that for this combination, the values of *RMSE* and *MAE* were obtained 63.42 and 52.64 m/s for testing set and 66.92 and 57.34 m/s for validation set, respectively. Therefore, as it was also found in regression analysis it can be concluded that the V_S correlates well with *SPT*, *FC* and *D*. Results showed that high correlation coefficient and low *RMSE* were also obtained for *C-2* and *C-5* in both *ANN* and *MLR* methods. This implies that the composition of two parameters *SPT* and *Depth* of soil could be good estimators for predicting V_S but joining parameters such as *FC* and *W* would improve the prediction accuracy.

Furthermore, the coefficient of determination for both the testing and validation data shows that the forecasted values of V_S using *ANN* method show reasonably good correlations with actual values. Figure 4 shows the relationship between the actual values of V_S and the predicted values of V_S using the *ANN* in testing and validation phase for the optimal model.

4.3. Comparison between ANN Predictions and Results of MLR

Combinations 1 to 9 were analyzed using both *ANN* and *MLR* methods. The predicted V_S by the *ANN* models for the testing and validation sets were compared with the estimated V_S by multiple regression analysis models. *The MAE*, *RMSE*, and R^2 values extracted from *ANN* and *MLR* methods for different combinations were depicted in Figures 5, 6, and 7. Figures 5 and 6 show that the values of *RMSE* and *MAE* obtained from regression analysis are greater than the *ANN* method for all of the above combinations in both testing and validation sets. It is also obvious according to Figure 7 that the correlation coefficient obtained from *ANN* models is more than the *MLR* models and all of which reflects the higher ability of *ANN* models for accurate prediction of the V_S value.

Finally, for comparing the *ANN* and *MLR* methods and performance evaluation of each of them, the V_S values predicted by these two methods for the optimal model (*C-3*) and for 20 random soil samples were presented in Figure 8. As the figure shows, the neural networks predict V_S values closer to the actual (measured V_S) values for most of the samples.

5. Conclusions

The goal of this study was to evaluate the feed forward neural networks as a possible tool for predicting V_S in Mashhad city. Five important input variables were used for predicting the V_S value. Different combinations of these inputs have been studied. Nine combinations of these variables which achieved the highest correlation coefficient in regression analysis were selected for comparing the *ANN* and *MLR* methods. The neural networks were trained for nine mentioned combinations by the feed forward back propagation algorithm and it seems that it well correlated the static properties of soil with V_S. To validate the neural network models, new observation data were introduced to the networks and the forecasted V_S were compared with actual values of V_S in the study area. There was a good fit between real and calculated data.

Both of the methods showed that the combinations of the three parameters including *D*, *SPT N* value, and *FC* give the best estimation of V_S of soil. The value of correlation coefficient and coefficient of determination obtained from the *ANN* method was higher than that of the *MLR* method. It should be mentioned that the error values computed through *RMSE* and *MAE* obtained from the *MLR* method were more than those extracted from the *ANN* method for all combinations under study. Therefore, it can be concluded that in comparison with *MLR* models, *ANN* models give more reliable prediction of V_S. In other words, *ANN* models have a better performance and can be used with a higher confidence coefficient to predict the V_S value of soil.

Acknowledgment

The authors appreciate cooperation from consulting engineering companies and soil mechanics laboratories in Mashhad city for providing their geotechnical reporting for use in this paper. They also acknowledge the kind cooperation from the Zamin Physics Pooya Consulting Engineering Company for giving permission to use measured shear wave velocity.

Geolocation Information: Mashhad is the second most populated city in the center of the Khorasan Razavi province. It is located 850 km North East of Tehran the capital of Iran at 36.20º north latitude and 59.35 º east longitudes in the valley of the Kashafrood River near Turkmenistan, between the two mountain ranges of Binalood and Hezar-Masjid, which are close to the borders of Afghanistan, and Turkmenistan.

Role of funding sources: the work was not financially supported by any funding sources.

References

1. Akin, M. K., Kramer, S. L., and Topal, T. "Empirical correlations of shear wave velocity (V_S) and penetration resistance (*SPT-N*) for different soils in an earthquake-prone area (Erbaa-Turkey)", *EngGeol.*, 119, PP. 1–17 (2011).

2. Fabbrocino, S., Lanzano, G., Forte, G., Santucci de Magistris, F., and Fabbroccini, G. *"SPT* blow count vs. Shear wave velocity relationship in the structurally complex formations of the Molise Region (Italy)", *Engineering Geology*., 187, PP. 84-97 (2015).

3. Sil, A., and Sitharam, T. G. "Dynamic site characterization and correlation of shear wave velocity with standard penetration test '*N'* values for the city of Agartala, Tripura state, India", *Pure and Applied Geophysics.*, 171(8), PP. 1859–1876 (2014).

4. Chatterjee, K., and Choudhury, D. "Variations in shear wave velocity and soil site class in Kolkata city using regression and sensitivity analysis", *Nat Hazards*., 69, PP. 2057–2082 (2013).

5. Hasancebi, N., and Ulusay, R. "Empirical correlations between shear wave velocity and penetration resistance for ground shaking assessments", Bulletin of Engineering Geology and the Environment., 66, PP. 203–213 (2007).

6. Dikmen, U. "Statistical correlations of shear wave velocity and penetration resistance for soils", *Journal of Geophysics and Engineering.*, 6, PP. 61–72 (2009).

7. Maheswari, R. U., Boominathan, A., and Dodagoudar, G. R. "Use of surface waves in statistical correlations of shear wave velocity and penetration resistance of Chennai soils", *Geotechnical and Geological Engineering.*, 28, PP. 119–137 (2010).

8. Ghazi, A., Hafezi Moghadas, N., Sadeghi, H., Ghafoori, M., and Lashkaripour, G. L. "Empirical relationships of shear wave velocity, *SPT-N* value and vertical effective stress for different soils in Mashhad, Iran", *Annals of geophysics.*, 58(3), S0325, (2015).

9. Brandenberg, S.J., Bellana, N., and Shantz, T. "Shear wave velocity as a statistical function of standard penetration test resistance and vertical effective stress at Caltrans bridge sites", *Soil Dynamics and Earthquake Engineering.*, 30, PP. 1026-1035 (2010).

10. Shooshpasha, I., Mola-Abasi, H., Jamalian, A., Dikmen, U., and Salahi, M. "Validation and application of empirical shear wave velocity models based on standard penetration test", *Computational Methods in Civil Engineering*., 4 (1), PP. 25-41 (2013).

11. Imai, T. "*P*- and *S*-wave velocities of the ground in Japan", In Proceedings of the IX, *international conference on soil mechanics and foundation engineering*, PP. 127–132 (1977).

12. Imai, T., and Yoshimura, Y. "Elastic wave velocity and soil properties in soft soil (in Japanese)", *Tsuchito- Kiso.,* 18(1), PP. 17–22 (1970).

13. Jafari, M. K., Shafiee, A., and Razmkhah, A. "Dynamic properties of fine grained soils in south of Tehran", *Soil Dynamics and Earthquake Engineering*., 4, PP. 25–35 (2002).

14. Kiku, H., Yoshida, N., Yasuda, S., Irisawa, T., Nakazawa. H., Shimizu. Y., Ansal, A., and Erkan, A. "In situ penetration tests and soil profiling in Adapazari, Turkey", In Proceedings of the *ICSMGE/TC4 satellite conference on lessons learned from recent strong earthquakes*, PP.259–265 (2001).

15. Lee, SHH. "Regression models of shear wave velocities", *Journal of the Chinese Institute of Engineers.,* 13, PP. 519–532 (1990).

16. Ohsaki, Y., and Iwasaki, R. "On dynamic shear moduli and Poisson’s ratio of soil deposits", *Soils and Foundations*., 13(4), PP. 61–73 (1973).

17. Ohta, Y., and Goto, N. "Empirical shear wave velocity equations in terms of characteristics soil indexes", *Earthquake Engineering and Structural Dynamics*., 6, PP. 167–187 (1978).

18. Pitilakis, K. D., Anastasiadis, A., and Raptakis, D. "Field and laboratory determination of dynamic properties of natural soil deposits", In Proceedings of *the 10th world conference on earthquake engineering, Rotherdam*, PP. 1275–1280 (1992).

19. Seed, H. B., and Idriss, I. M. "Evaluation of liquefaction potential sand deposits based on observation of performance in previous earthquakes", *Preprint 81–544, in situ testing to evaluate liquefaction susceptibility, ASCE National Convention*, Missouri, PP. 81–544 (1981).

20. Sykora, D.E., and Stokoe, K.H. "Correlations of in-situ measurements in sands of shear wave velocity", *Soil Dynamics and Earthquake Engineering*., 20, PP. 125–36 (1983).

21. Dehghan Nayeri, G., Dehghan Nayeri, D., and Barkhordari, K. "A New Statistical Correlation between Shear Wave Velocity and Penetration Resistance of Soils Using Genetic Programming", *Electronic Journal of Geotechnical Engineering*., 18, PP. 2071-2078 (2013).

22. Sivrikaya, O. "Comparison of Artificial Neural Networks Models with Correlative Works on Undrained Shear Strength", *Eurasian Soil Science*., 42(13), PP. 1487–1496 Pleiades Publishing, Ltd (2009).

23. Dehghan, S., Sattari, Gh. Chehreh chelghani, S., and Aliabadi, M.A. "Prediction of uniaxial compressive strength and modulus of elasticity for Travertine samples using regression and artificial neural networks", *Mining Science and Technology*., 20, PP. 0041-0046 (2010).

24. Sarmadian, F., Keshavarzi, A. "Developing pedotransfer functions for estimating some soil properties using artificial neural network and multivariate regression approaches", *International Journal of Environmental and Earth Sciences*., 1(1), PP. 31-37 (2010).

25. Schaap M.G., Leij, F.J., and Van Genuchten, M.Th. "Neural network analysis for hierarchical prediction of soil hydraulic properties", *Soil Science Society of America Journal.*, 62, PP. 847 – 855 (1998).

26. Maleki, S., Moradzadeh, A., Ghavami Riabi, R., Gholami, R., and Sadeghzadeh, F. "Prediction of shear wave velocity using empirical correlations and artificial intelligence methods", *NRIAG Journal of Astronomy and Geophysics*, 3, PP. 70-81 (2014).

27. Mohammadi, H., Rahmannejad, R. "The estimation of rock mass deformation modulus using regression and artificial neural network analysis", *Arabian Journal for Science and Engineering.*, 35 (1A), PP. 67–77 (2009).

28. Gunaydm, O. "Estimation of soil compaction parameters by using statistical analyses and artificial neural networks", *Environ. Geol*., 57, PP. 203–215 (2009).

29. Sudha Rani, Ch. "Artificial Neural Networks (*ANNs*) For Prediction of Engineering Properties of Soils", *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*., 3 (1), PP. 123-130 (2013).

30. Khanlari, GR., Heidari, M., Momeni, AA., and Abdilor, Y. "Prediction of shear strength parameters of soils using artificial neural networks and multivariate regression methods", *Engineering Geology.*, 131– 132, PP. 11–18 (2012).

31. Mola-Abasi, H., and Shooshpasha, I. "Prediction of zeolite-cement-sand unconfined compressive strength using polynomial neural network", *The European Physical Journal Plus.*, 131 (4), PP. 1-12 (2016).

32. Park, H.I. "Development of neural network model to estimate the permeability coefficient of soils", *Marine Geosources and Geotechnology*., 29(4), PP. 267-278 (2011).

33. Harini, H.N., and Sureka Naagesh. "Predicting *CBR* of fine grained soils by artificial neural network and multiple linear regression", *International journal of civil engineering and technology (IJCIET)*., 5 (2), PP.119-126 (2014).

34. Moayed, RZ., Kordnaeij, A., and Mola-Abasi, H. "Compressibility indices of saturated clays by group method of data handling and genetic algorithms", *Neural Computing and Applications.*, PP. 1-14 (2016).

35. Mola-Abasi, H., and Shooshpasha, I. "Prediction of Compression Index of Saturated Clays (*Cc*) Using polynomial models", *scientia iranica*., 23 (2), PP. 500-507 (2016).

36. Kordnaeij, A., Kalantary, F., Kordtabar, B., and Mola-Abasi, H. "Prediction of recompression index using *GMDH*-type neural network based on geotechnical soil properties", *Soils and Foundations*., 55 (6), PP. 1335-1345 (2015).

37. Das, S.K. and Basudhar, P.K. "Undrained lateral load capacity of piles in clay using artificial neural network", *Computers and Geotechnics.*, 33, PP. 454–459 (2006).

38. Teh, C. I., Wong, K. S., Goh, A. T. C., and Jaritngam, S. "Prediction of pile capacity using neural networks", *J. Computing in Civil Engineering*, *ASCE*., 11(2), PP. 129-138 (1997).

39. Sivakugan, N., Eckersley, J. D., and Li, H. "Settlement predictions using neural networks", *Australian Civil Engineering Transactions*., CE40, PP. 49-52 (1998).

40. Mola-Abasi, H., Shooshpasha, I., and Amiri, I. "Prediction of Liquefaction Induced Lateral Displacements Using *GMDH* type Neural Networks", *Global Journal of Scientific Researches*., 2(1), PP. 21-26 (2014).

41. Goh, A.T.C. "Probabilistic neural network for evaluating seismic liquefaction potential", *Canadian Geotechnical Journal*., 39, PP. 219-232 (2002).

42. Kim, Y.S. and Kim, B.K. "Use of artificial neural networks in the prediction of liquefaction resistance of sands", *Journal of Geotechnical and Geo environmental Engineering.*, 132(11), PP. 1502-1504 (2006).

43. Ural, D. N., and Saka, H. "Liquefaction assessment by neural networks", *Electronic Journal of Geotechnical Engineering*., (http://www.ejge.com/1998/JourTOC3.htm), 3, (1998).

44. Cho, S.E. "Probabilistic stability analyses of slopes using the *ANN*-based response surface", *Computers and Geotechnics.*, 36, PP. 787–797 (2009).

45. Ferentinou, M.D. and Sakellariou, M.G. "Computational intelligence tools for the prediction of slope performance", *Computers and Geotechnics*., 34(5), PP. 362- 384 (2007).

46. Wong, F. S. "Time series forecasting using back-propagation neural networks", *Neurocomputing.*, 2, PP. 147–259 (1991).

47. Eskandari, H., Rezaee, M.R. and Mohammadnia, M. "Application of Multiple Regression and Artificial Neural Network Techniques to Predict Shear Wave Velocity from Well Log Data for a Carbonate Reservoir, South-West Iran", *Cseg Recorder*., PP. 42-48 (2004).

48. Moatazedian, I., Rahimpour-Bonab, H., Kadkhodaie-Ilkhchi, A. and Rajoli, M.R. "Prediction of Shear and Compressional Wave Velocities from Petrophysical Data Utilizing Genetic Algorithms Technique: A Case Study in Hendijan and Abuzar Fields Located in Persian Gulf", *Geopersia.*, 1, PP. 1-17 (2011).

49. Akhundi, H., Ghafoori, M., and Lashkaripour, G.R. "Prediction of shear wave velocity using artificial neural network technique, multiple regression and petrophysical data: A case study in Asmari reservoir (SW Iran)", *Open Journal of Geology*., 4, PP. 303–313 (2014).

50. Rezaee, M.R., Kadkhodaie Ilkhchi, A., and Barabadi, A. "Prediction of shear wave velocity from petrophysical data utilizing intelligent systems: An example from a sandstone reservoir of Carnarvon Basin, Australia", *Journal of Petroleum Science and Engineering.*, 55, PP. 201–212 (2007).

51. Mola-Abasi, H., Dikmen, U., and Shooshpasha, I. "Prediction of shear-wave velocity from *CPT* data at Eskisehir (Turkey) using a polynomial model", *Near Surface Geophysics*., 13 (2), PP. 155-167 (2015).

52. Mola-Abasi, H., Eslami, A., and Tabatabaie Shourijeh, P. "Shear Wave Velocity by Polynomial Neural Networks and Genetic Algorithms Based on Geotechnical Soil Properties", *Arabian Journal for Science and Engineering*., 38(4), PP. 829–838 (2013).

53. Shooshpasha, I., Kordnaeij, A., Dikmen, U., MolaAbasi, H, and Amir, I. "Shear wave velocity by support vector machine based on geotechnical soil properties", *Natural Hazards and Earth System Sciences Discussions*., 2 (4), PP. 2443-2461 (2014).

54. Berberian, M. and Ghoreshi, M. "Seismic - fault hazard and project engineering of thermal power plant of Nishapur, Seismotectonical Survey (In Persian)", Ministry of Energy, Power Engineering Corporation (Moshanir), Tehran, (1989).

55. Azadi, A., Javan-Doloei, G. h., Hafezi Moghadas, N. and Hessami-Azar, K. "Geological, geotechnical and geophysical characteristics of the Toos Fault located North of Mashhad, North-eastern Iran (In Persian) ", *Journal of the Earth and Space Physics*., 35(4), PP. 17-34 (2010).

56. Building and housing research center. "Iranian code of practice for seismic resistant design of buildings, standard No. 2800", 3rd edition, Tehran, Iran (2007).

57. James, G., Witten, D., Hastie, T., Tibshirani, R. "An Introduction to Statistical Learning with Applications in R", Springer, New York, Heidelberg Dordrecht, London (2013).

58. Haykin, S. "Neural Networks: A Comprehensive Foundation", Second ed. Prentice-Hall, Upper Saddle River, New Jersey, PP. 26-32 (1999).

59. Shahin, M. A., Jaksa, M. B., and Maier, H. R. "Artificial neural network applications in geotechnical engineering", *Australian Geomechanics.*, 36 (1), PP. 49–62 (2001).

60. Isik, F., and Ozden, G. "Estimating compaction parameters of fine and coarse grained soils by means of artificial neural networks", *Environmental Earth Sciences*., 69, PP. 2287-2297 (2013).

61. Kisi, O. "Stream flow forecasting using different artificial neural network algorithms", *Journal of Hydrologic Engineering*. *ASCE*., 12 (5), PP. 532-539 (2007).

62. Kanungo, D.P., Arora, M.K., Sarkar, S., and Gupta, R.P. "A comparative study of conventional, *ANN* black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas", *Engineering Geology*., 85, PP. 347 -366 (2006).

63. Kartam, N., Flood, I., and Garrett, J. H. *"Artificial Neural Networks for Civil Engineers", Fundamentals and Applications.*, ASCE, New York (1997).

64. Kayadelen, C. "Estimation of effective stress parameter of unsaturated soils by using artificial neural networks", *International Journal for Numerical and Analytical Methods in Geomechanics.,* 32 (9), PP. 1087–1106 (2008).

65. Zhang, G., Patuwo, E. B., and Hu, M. Y. "Forecasting with artificial neural network: The state of the art", *International Journal of Forecasting.*, 14, PP. 35–62 (1998).

66. Wang, H.B.; Xu, W.Y. and Xu, R.C. "Slope stability evaluation using Back Propagation Neural Networks", *Engineering Geology.*, 80, PP. 302– 315 (2005).

67. Rivals, I., and Personnaz, L. "Neural-network construction and selection in nonlinear modeling", *IEEE Transaction on Neural Networks.*, 14(4), PP. 804–819 (2003).

68. Ghiassi, M., and Nangoy, S. "A dynamic artificial neural network model for forecasting nonlinear processes", *Computers & Industrial Engineering.*, 57 (1), PP. 287–297 (2009).

Table 1. Descriptive statistics of input and output data

Table 2. Input and Output for the different combinations

Table 3. Performance evaluation criteria for the different combinations obtained by *MLR* analysis

Table 4. Performance criteria for different models in testing and validation stage by *ANN* method

Figure 1. Geological map of study area with location of boreholes and faults

Figure 2. Architecture of Multilayer Neural Network for this study

Figure 3. Scatter plot of measured V_S versus predicted V_S for the best model by *MLR* analysis

Figure 4. Relationship between measured and predicted shear wave velocity by *ANN* analysis in (a) testing and (b) validation phase

Figure 5. Comparison of *RMSE* values obtained by *ANN* and *MLR*: (a) testing set, (b) validation set

Figure 6. Comparison of *MAE* values obtained by *ANN* and *MLR*: (a) testing set, (b) validation set

Figure 7. Coefficient of determination (R^2) obtained by *ANN* and *MLR*: (a) testing set, (b) validation set

Figure 8. Measured V_S versus predicted V_S by *ANN* and *MLR* methods for the best model(*C-3*): (a) testing set, and (b) validation set

Table 1. Descriptive statistics of input and output data

Partition	Statistics	*D*	*SPT*	*FC*	*W*	*LL*	V_S *(m/s)*
All dataset	Mean	16.20	43	31.41	6.56	21.46	515
	Std. Deviation	10.40	21	28.21	5.18	5.97	157
	Minimum	0.50	10	1.80	1.60	2.00	202
	Maximum	49.00	97	99.00	26.10	55.00	850
Training set	Mean	15.94	44	30.50	6.18	21.48	516
	Std. Deviation	10.11	21	26.77	4.48	6.38	148
	Minimum	0.50	10	1.80	1.60	2.00	202
	Maximum	49.00	97	97.00	21.95	55.00	850
Testing set	Mean	19.00	42	37.97	9.22	21.74	524
	Std. Deviation	12.25	20	36.34	8.54	5.60	175
	Minimum	1.50	16	6.00	2.10	15.00	204
	Maximum	41.00	91	99.00	26.10	35.00	773
Validation set	Mean	15.59	40	31.42	6.63	21.18	504
	Std. Deviation	10.53	22	29.45	5.12	4.06	186
	Minimum	1.50	11	7.30	2.00	15.00	210
	Maximum	41.50	91	98.50	26.00	32.70	790

Table 2. Input and Output for the different combinations

Combination No.	Input	Output
C-1	*D, SPT, LL*	
C-2	*D, SPT, W*	
C-3	*D, SPT, FC*	
C-4	*D, FC, W*	
C-5	*D, SPT*	V_S
C-6	*D, LL*	
C-7	*D, W*	
C-8	*D, FC*	
C-9	*SPT*	

Table 3. Performance evaluation criteria for the different combinations obtained by *MLR* analysis

Combination No.	R^2	R	*RMSE (m/s)*	*MAE (m/s)*
C-1	0.719	0.848	84.28	70.84
C-2	0.729	0.854	83.34	69.89
C- 3	0.733	0.856	82.29	68.81
C-4	0.588	0.767	100.74	82.02
C-5	0.711	0.843	85.7	71.34
C-6	0.506	0.711	111.37	90.27
C-7	0.560	0.748	106.45	87.73
C-8	0.573	0.757	102.09	83.77
C-9	0.599	0.774	99.52	85.44

Table 4. Performance criteria for different models in testing and validation stage by *ANN* method

Combination No.	R^2		R		*RMSE*		*MAE*	
	Test	Validation	Test	Validation	Test	Validation	Test	Validation
C-1	0.856	0.869	0.926	0.932	64.27	69.41	53.7	61.84
C-2	0.854	0.885	0.924	0.941	70.95	67.32	58.4	57.37
C-3	0.878	0.887	0.931	0.942	63.42	66.92	52.64	57.34
C-4	0.681	0.817	0.825	0.904	101.07	86.29	77.96	73.52
C-5	0.863	0.870	0.929	0.933	76.86	69.32	60.71	60.2
C-6	0.748	0.839	0.865	0.916	88.97	82.93	70.52	67.58
C-7	0.760	0.759	0.872	0.871	90.46	96.65	75.5	83.1
C-8	0.741	0.812	0.861	0.901	90.39	87.53	76.19	71.47
C-9	0.669	0.803	0.818	0.896	104.05	89.99	88.07	74.19

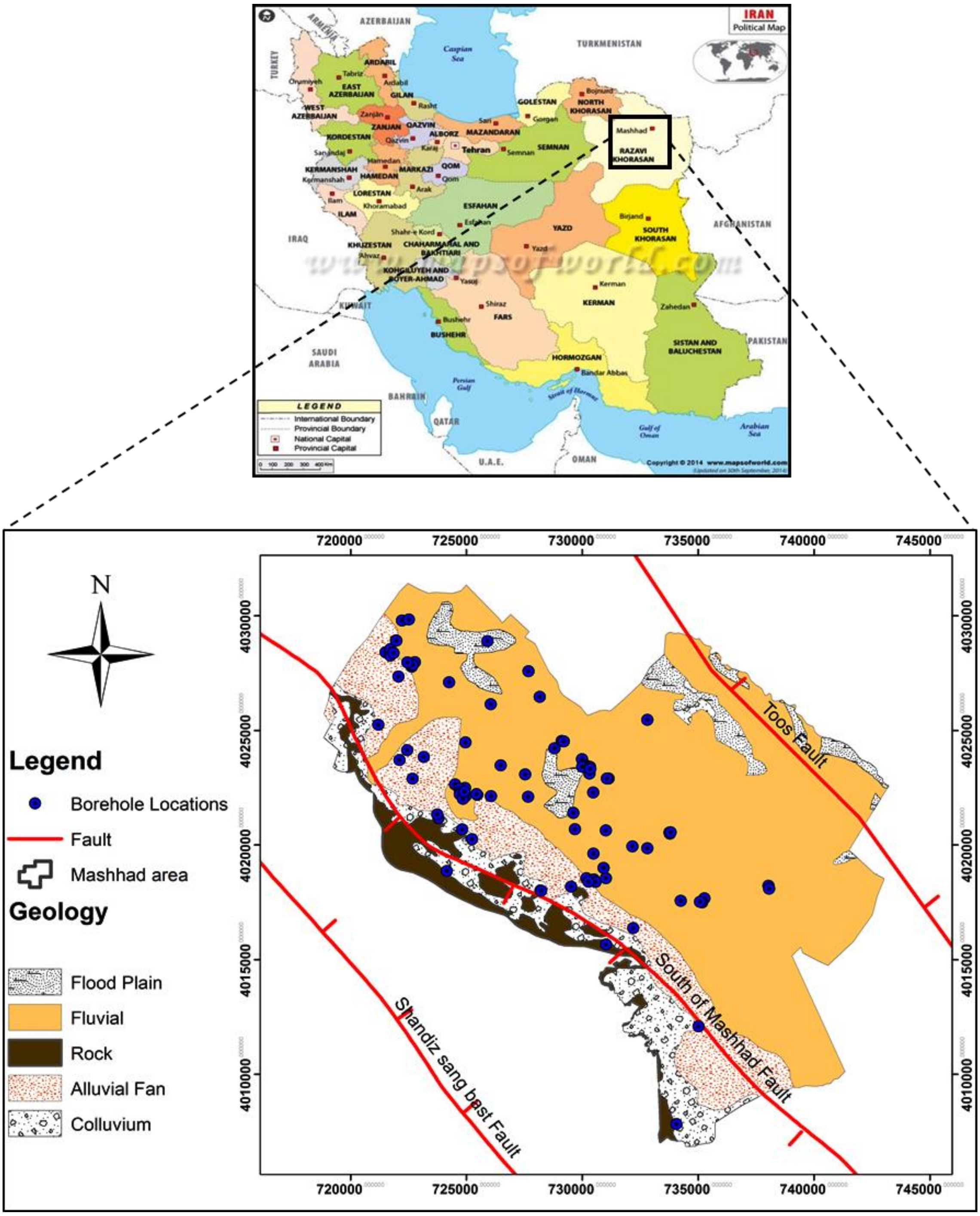

Figure 1. Geological map of study area with location of boreholes and faults

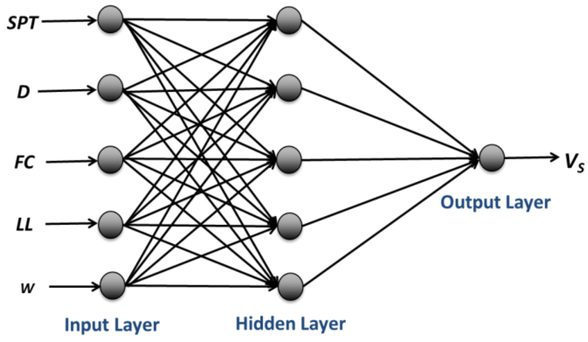

Figure 2. Architecture of Multilayer Neural Network for this study

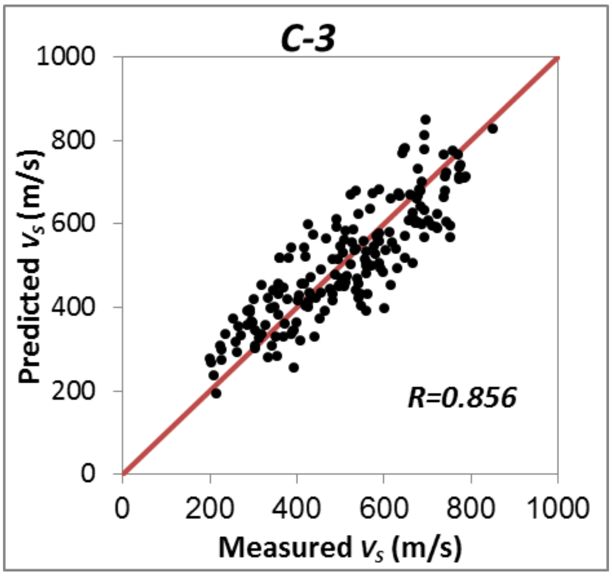

Figure 3. Scatter plot of measured V_S versus predicted V_S for the best model by *MLR* analysis

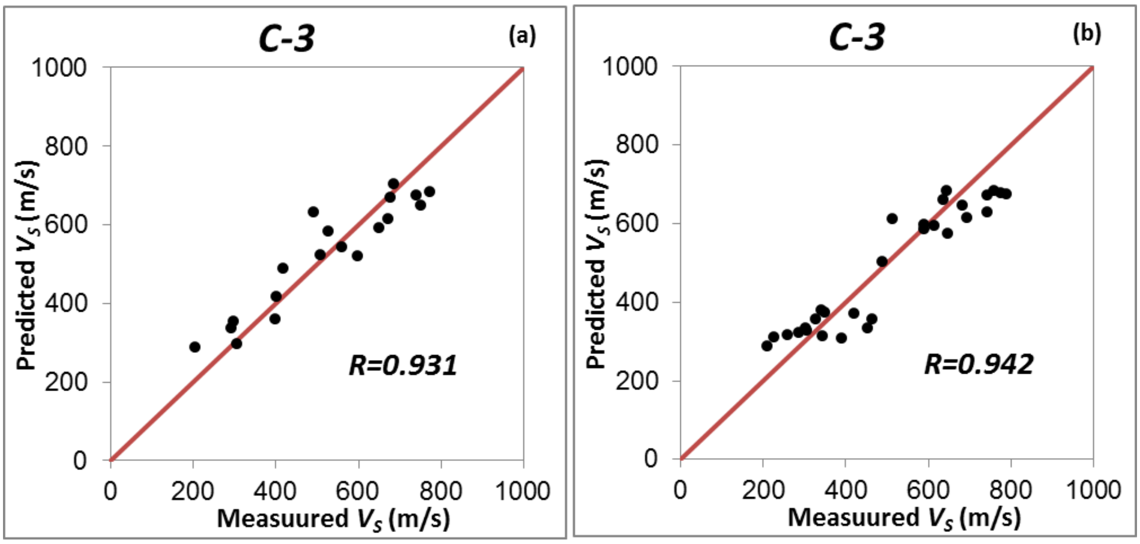

Figure 4. Relationship between measured and predicted shear wave velocity by *ANN* analysis in (a) testing and (b) validation phase

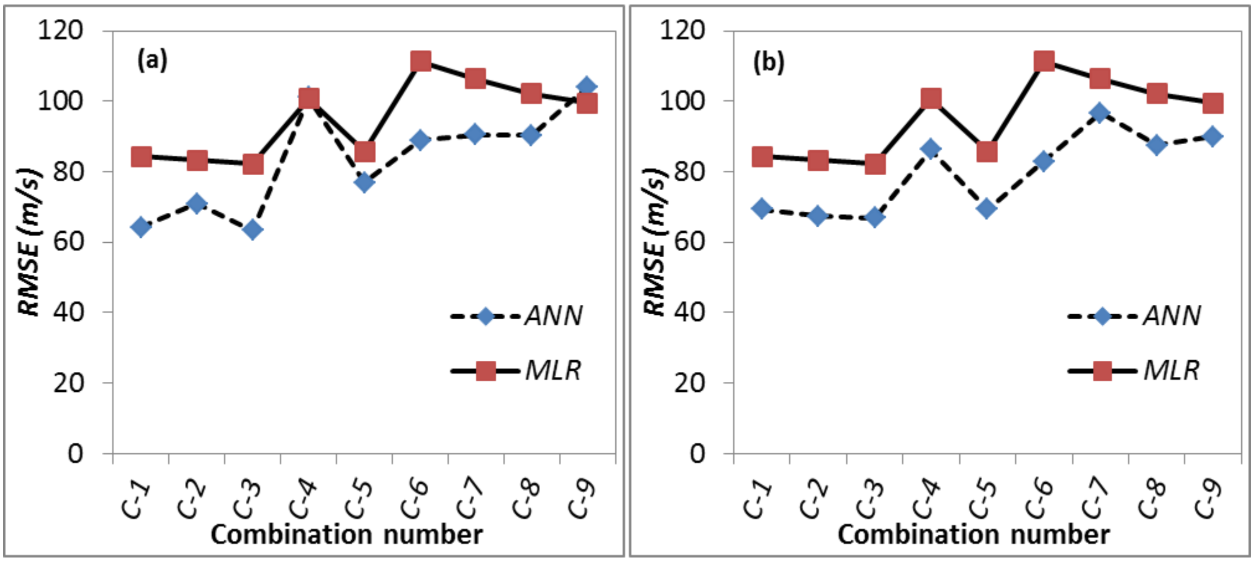

Figure 5. Comparison of *RMSE* values obtained by *ANN* and *MLR*: (a) testing set, (b) validation set

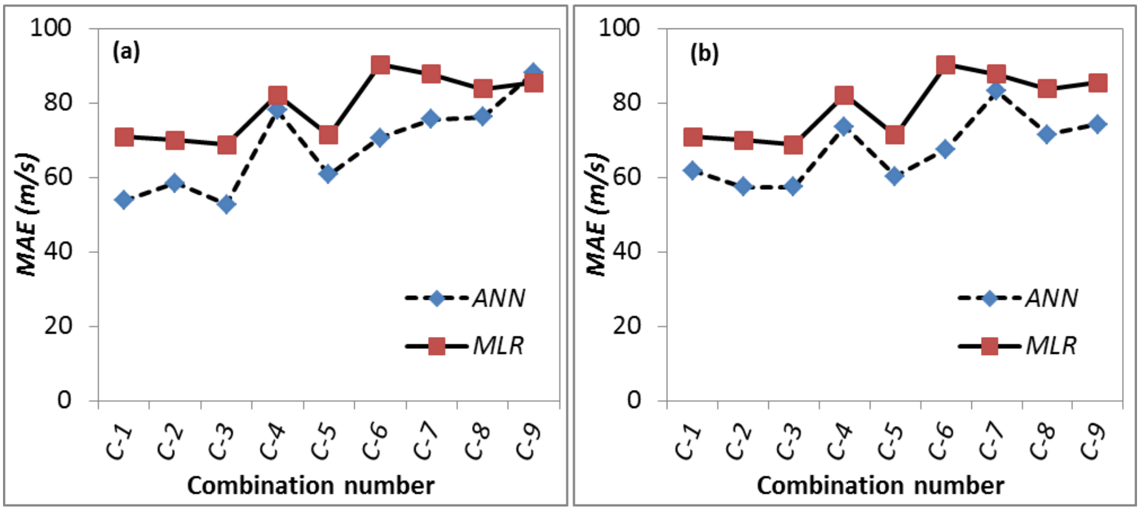

Figure 6. Comparison of *MAE* values obtained by *ANN* and *MLR*: (a) testing set, (b) validation set

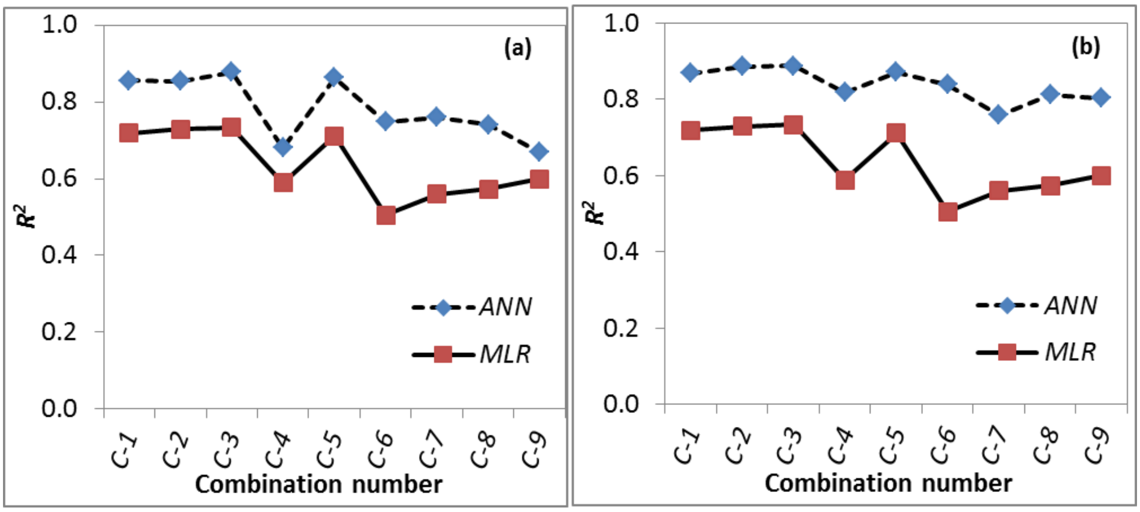

Figure 7. Coefficient of determination (R^2) obtained by *ANN* and *MLR*: (a) testing set, (b) validation set

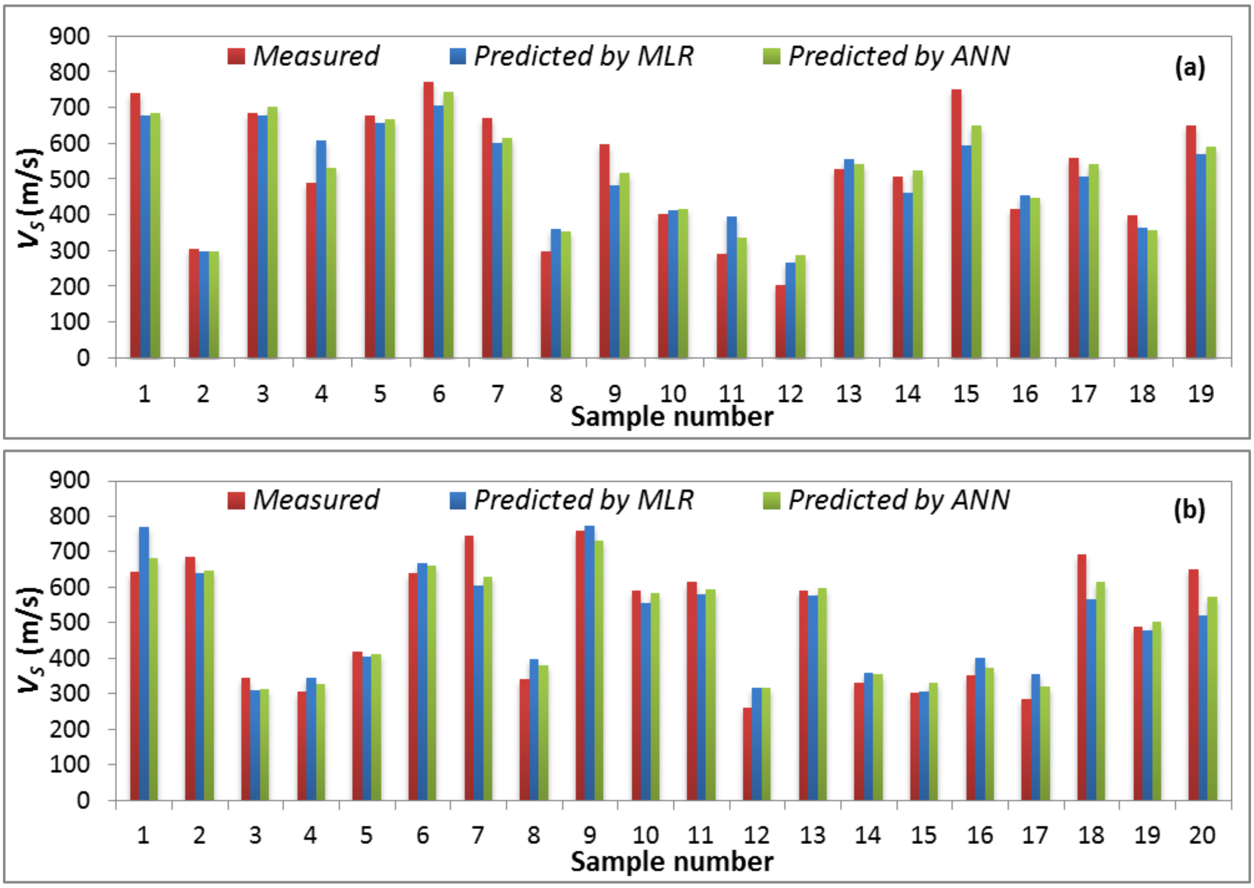

Figure 8. Measured V_S versus predicted V_S by *ANN* and *MLR* methods for the best model(*C-3*): (a) testing set, and (b) validation set