Civil Engineering and Applied Solutions

Research Article

journal homepage: ceas.journals.umz.ac.ir

CEAS Civil Engineering and Applied Solutions

A Hybrid GMDH Neural Network-Genetic Algorithm Approach for Predicting Shear Wave Velocity Profiles

Omolbanin Ataee [a]

[a] *Department of Geography and Urban Planning, University of Mazandaran, Babolsar, Iran*

ARTICLE INFO

Article history:

Received 12 May 2026

Revised 16 June 2026

Accepted 07 August 2026

Available online 01 April 2027

Keywords:

Genetic algorithm

Group method of data handling

Mashhad city

Shear wave velocity

Standard penetration test

ABSTRACT

Shear wave velocity (V_s) is a fundamental parameter in geotechnical earthquake engineering and seismic site characterization; however, direct in-situ measurement remains expensive and spatially limited. This study proposes an interpretable hybrid group method of data handling-genetic algorithm (GMDH-GA) framework for predicting shear wave velocity profiles in heterogeneous alluvial deposits. A comprehensive geotechnical-geophysical database comprising 1991 records from 226 boreholes in Mashhad, Iran, was compiled using downhole measurements and standard penetration test (SPT) data. Depth (Z), SPT blow count (N), and upper-layer shear wave velocity (V_{su}) were selected as predictor variables. Separate soil-specific models were developed for clay, silt, sand, gravel, and combined soil datasets using a 70/30 training-testing strategy. The proposed models generated explicit polynomial equations while maintaining high predictive capability. Statistical evaluation demonstrated strong agreement between measured and predicted V_s values, with testing-stage coefficients of determination reaching 0.98-0.99 and low prediction errors across all soil groups. Comparative analyses showed that the proposed GMDH-GA models substantially outperform widely used empirical correlations in terms of accuracy and generalization capability. Sensitivity analysis indicated that V_{su} is the dominant predictor, followed by SPT resistance and depth. The results demonstrate that the proposed framework provides a reliable and interpretable alternative for estimating shear wave velocity profiles in regions where direct geophysical measurements are unavailable.

How to cite this article: Ataee, O. A Hybrid GMDH Neural Network-Genetic Algorithm Approach for Predicting Shear Wave Velocity Profiles. Civil Engineering and Applied Solutions. 2027; 3(2): 144–160. doi:10.22080/ceas.2026.31713.1099.

1. Introduction

Shear wave velocity (V_s) is one of the most important parameters used to characterize the dynamic behavior and stiffness of soil deposits. In particular, the average shear wave velocity of the upper 30 m of the ground profile (V_{s30}) is widely used for seismic site classification, site-response analysis, and seismic microzonation studies. Although V_{s30} is an important parameter for evaluating local site conditions, it should not be considered a direct measure of seismic hazard, which is controlled by multiple geological and seismological factors.

Direct measurement of V_s can be obtained using geophysical techniques such as downhole, crosshole, and surface-wave methods. However, these methods generally require specialized equipment, experienced personnel, and data interpretation or inversion procedures, which may increase project cost and complexity. Consequently, empirical correlations between V_s and routinely measured geotechnical parameters, particularly standard penetration test (SPT) blow counts, have been widely developed and applied in engineering practice.

* Corresponding author.
E-mail addresses: o.ataee@umz.ac.ir (O. Ataee).

https://doi.org/10.22080/ceas.2026.31713.1099
ISSN: 3092-7749/© 2027 The Author(s). Published by University of Mazandaran.

Despite their practical value, conventional empirical V_s-SPT correlations often exhibit considerable uncertainty because soil behavior is influenced by multiple factors, including soil type, stress conditions, depositional environment, and regional geological characteristics. Therefore, advanced data-driven approaches capable of capturing complex nonlinear relationships may provide more reliable and site-specific predictions of shear wave velocity.

Extensive research has been conducted to establish empirical relationships between SPT blow count (N) and shear wave velocity. Early studies by Ohta and Goto [1] and Imai and Tonouchi [2] introduced power-law and stress-dependent formulations that have formed the basis of most subsequent models. These relationships were later refined by incorporating soil density, confining stress, and depositional effects [3, 4]. However, despite their simplicity and practical utility, such empirical correlations are generally site-specific and exhibit limited transferability due to variations in soil fabric, stress history, and geological conditions [5]. A number of correlations between the V_s and the SPT-N value are collected from the literature and shown in Table 1.

Table 1. Inventory of proposed empirical correlations between V_s and SPT-N value.

Authors	All soils	Gravel	Sand	Silt	Clay
Ohsaki and Iwasaki [6]	$V_s = 81.38N^{0.39}$		$V_s = 59.4N^{0.47}$		
Imai [7]	$V_s = 91N^{0.337}$		$V_s = 80.6N^{0.331}$		$V_s = 80.2N^{0.292}$
Ohta and Goto [1]	$V_s = 85.35N^{0.348}$				
Seed and Idriss [8]	$V_s = 61N^{0.5}$				
Lee Shannon [9]	$V_s = 76.2N^{0.24}$			$V_s = 104(N+1)^{0.334}$	$V_s = 138.4(N+1)^{0.242}$
Iyisan [10]	$V_s = 51.5N^{0.516}$				
Jafari et al. [5]	$V_s = 22N^{0.85}$				
Hasancebi and Ulusay [11]	$V_s = 90N^{0.309}$		$V_s = 90.82N^{0.319}$		$V_s = 97.89N^{0.269}$
Dikmen [12]	$V_s = 58N^{0.39}$		$V_s = 73N^{0.33}$	$V_s = 60N^{0.36}$	$V_s = 44N^{0.48}$
Maheswari et al. [13]	$V_s = 95.64N^{0.301}$		$V_s = 100.53N^{0.265}$		$V_s = 89.31N^{0.358}$
Hafezi et al. [14]	$V_s = 99N^{0.53}$	$V_s = 80N^{0.58}$	$V_s = 80N^{0.58}$	$V_s = 45N^{0.72}$	$V_s = 45N^{0.72}$
Chatterjee and Choudhury [15]	$V_s = 78.21N^{0.38}$		$V_s = 54.82N^{0.53}$		$V_s = 77.11N^{0.39}$
Fatehnia et al. [16]	$V_s = 77.1N^{0.355}$				
Kirar et al. [17]	$V_s = 99.5N^{0.345}$		$V_s = 100.3N^{0.338}$		$V_s = 94.4N^{0.379}$

To overcome these limitations, recent research has increasingly focused on data-driven machine learning approaches. Algorithms such as artificial neural networks (ANN), support vector machines (SVM), random forest (RF), and gradient boosting have demonstrated superior capability in capturing nonlinear relationships between geotechnical parameters and shear wave velocity. Numerous studies report that these models outperform conventional regression-based methods in terms of predictive accuracy and robustness [18-20]. Furthermore, hybrid and ensemble learning frameworks have further improved performance by enhancing generalization capability and reducing model uncertainty [21, 22].

Recent advances in machine learning have substantially expanded the application of intelligent predictive models in civil and geotechnical engineering. Beyond conventional neural networks, researchers have increasingly employed hybrid and ensemble frameworks to improve prediction accuracy, robustness, and generalization capability under complex nonlinear conditions. For example, Tabari et al. [23] developed a supervised committee neural network for estimating aquifer parameters and demonstrated that combining multiple intelligent predictors can significantly improve predictive performance compared with individual models. Their findings highlighted the potential of ensemble learning strategies for modeling highly nonlinear geotechnical and hydrogeological systems.

More recently, machine-learning-driven methodologies have been extended to structural and seismic engineering applications. Farahani [24] proposed a data-driven seismic design framework integrating wavelet transforms and machine learning algorithms for reinforced concrete frame assessment. The study demonstrated the effectiveness of hybrid intelligent systems in extracting complex patterns from engineering datasets and improving prediction reliability. Similarly, Sadeghpour Haji et al. (2026) employed artificial neural networks to model the shear behavior of reinforced concrete columns under constant axial loading and reported high predictive accuracy for nonlinear structural responses. These studies collectively indicate the growing trend toward integrating advanced machine learning techniques with engineering domain knowledge to address complex prediction problems [25].

Despite these developments, several challenges remain unresolved in shear wave velocity prediction. Existing studies predominantly focus on black-box machine learning algorithms that provide limited interpretability for engineering decision-making. Moreover, many published models have been developed using relatively small or geographically restricted datasets, which may not adequately represent the spatial variability and heterogeneity of natural alluvial deposits. In addition, only limited research has explored the integration of interpretable polynomial neural networks with evolutionary optimization algorithms for regional-scale V_s profile prediction. Consequently, there remains a need for predictive frameworks that simultaneously achieve high accuracy, transparency, and robustness while being developed from large-scale geotechnical databases. The present study addresses this research gap through the development of an interpretable hybrid GMDH-GA model calibrated using an extensive geotechnical-geophysical dataset from Mashhad, Iran.

Despite these advances, a key limitation of most machine learning models is their lack of interpretability. Many high-performance algorithms operate as black-box systems, providing accurate predictions without yielding explicit mathematical expressions, which limits their applicability in engineering design. In this context, the group method of data handling (GMDH), originally introduced by Ivakhnenko [26], offers a powerful alternative. GMDH is a self-organizing polynomial neural network that automatically constructs model structure through layer-wise selection of selected neurons. Unlike black-box models, it produces explicit mathematical equations, making it particularly suitable for engineering applications requiring both accuracy and transparency. Previous studies have confirmed its effectiveness in modeling complex geotechnical phenomena, including settlement, shear strength, liquefaction potential, and shear wave velocity [27-29]. To further improve performance, GMDH has been coupled with evolutionary optimization techniques, particularly genetic algorithms (GA). The hybrid GMDH-GA approach enhances model structure selection and parameter optimization, thereby improving generalization and avoiding local minima. Such a combination has demonstrated efficient performance in various geotechnical challenges; for instance, in soil property estimation and site characterization, such as predicting the correlation between unconfined compressive strength and standard penetration test results [30, 31] and in the assessment of soil liquefaction potential [32, 33].

Despite considerable progress in machine learning-based prediction of shear wave velocity, several important limitations remain unresolved. First, many high-performance machine learning models operate as black-box systems and therefore lack transparency and interpretability for practical geotechnical design. Second, most existing studies rely on relatively limited regional datasets that insufficiently capture the heterogeneity of complex alluvial deposits. Third, limited attention has been given to integrating interpretable polynomial neural networks with evolutionary optimization algorithms for regional V_s profile prediction.

To address these limitations, this study develops a hybrid GMDH-genetic algorithm framework using a comprehensive geotechnical-geophysical database consisting of 1991 records from 226 boreholes in Mashhad, Iran. Unlike conventional black-box approaches, the proposed method generates explicit polynomial equations while maintaining high predictive capability. In addition, separate soil-specific models are developed for clay, silt, sand, and gravel deposits to better represent soil-dependent nonlinear behavior. The main contributions of this study are:

- development of interpretable hybrid GMDH-GA models for V_s prediction;
- incorporation of sequential upper-layer velocity (V_{su}) information to capture stratigraphic continuity;
- establishment of soil-specific predictive equations using a large regional database; and
- comprehensive comparison with widely used empirical correlations.

2. General setting and geological characteristics of the site

Mashhad, the capital of Razavi Khorasan province and the second most populous city in Iran, covers an area exceeding 300 km² in northeastern Iran. The site is geographically located between latitudes 36.10°-36.24° N and longitudes 59.25°-59.43° E (Fig. 1). From a tectonic perspective, the region is situated at the boundary between the Central Iran and Koppe Dagh seismotectonic provinces. The Koppe Dagh structural zone, including the Hezar Masjed mountains, extends from the Caspian sea toward Afghanistan along a northwest-southeast trend. The Mashhad-Ghuchan depression represents a narrow compressional basin separating the folded-thrusted structures of the Koppe Dagh from the Binalood mountains. The northeastern boundary of the basin is controlled by major thrust faults, including the Kashafrood Fault, whereas the southwestern boundary is associated with the South Mashhad and Chenaran faults. The city is predominantly underlain by young alluvial deposits of the Mashhad Plain.

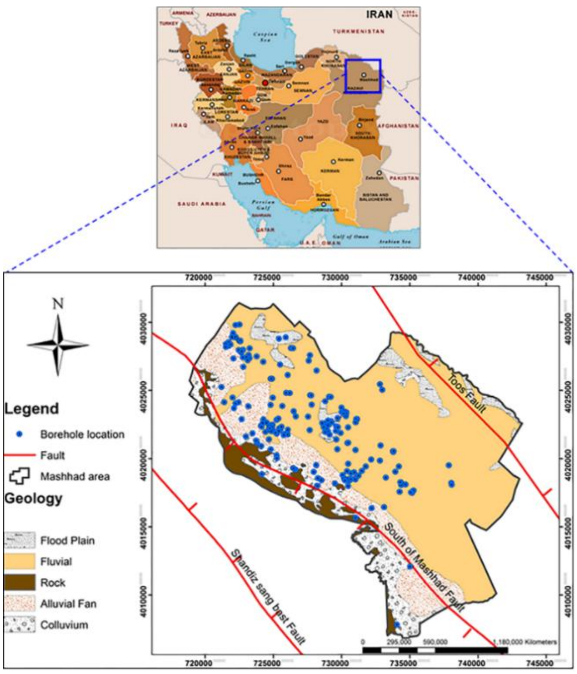


Fig. 1. Geological map of study area with location of boreholes and faults.

Based on regional seismotectonic studies and historical earthquake records, the study area is characterized by high seismic activity. Although the region is currently experiencing a relatively quiescent seismic period, historical evidence indicates the occurrence of numerous destructive earthquakes. According to the Iranian code of practice for seismic resistant design of buildings (Standard No. 2800) [34], the area is classified as a high seismic hazard zone, with a peak ground acceleration (PGA) of approximately 0.30-0.35 g for a return period of 475 years.

3. Methodology

3.1. Data acquisition and preprocessing

A comprehensive geotechnical-geophysical database was developed using site investigation reports provided by several consulting engineering companies in Mashhad, Iran. The final database comprised 1,991 data records obtained from 226 boreholes distributed across 186 investigation sites within the Mashhad Plain. The subsurface deposits primarily consist of clay, silt, sand, and gravelly alluvial materials, representing a wide range of geotechnical conditions. Standard penetration tests (SPT) were conducted at regular depth intervals in accordance with ASTM standards. In addition, downhole geophysical surveys were performed at selected locations to directly measure shear wave velocity (V_s) profiles.

The compiled database included depth (Z), soil type, SPT blow count (N), moisture content, density, Atterberg limits, and measured shear wave velocity (V_s) values. Prior to model development, a rigorous quality-control procedure was implemented to ensure data reliability. Incomplete records, physically inconsistent measurements, and erroneous entries were removed, while potential outliers were identified through statistical analysis and subsequently evaluated using geotechnical engineering judgment. Such preprocessing procedures are essential for improving the robustness and reliability of machine-learning-based geotechnical prediction models [35, 36].

Based on data availability, engineering relevance, and findings reported in previous studies, depth (Z), SPT blow count (N), and shear wave velocity of the upper soil layer (V_{su}) were selected as the input variables, whereas shear wave velocity (V_s) was considered the target output parameter. These variables have been widely recognized as key factors influencing subsurface stiffness characteristics and have been extensively utilized in both empirical and data-driven V_s prediction models [12, 13].

During the preprocessing stage, all input variables were normalized to eliminate scale dependency and improve the numerical stability and convergence behavior of the hybrid GMDH-GA framework. Data normalization is particularly important in evolutionary and neural-network-based algorithms because it facilitates efficient learning and minimizes the influence of variable magnitude disparities [37, 38]. Subsequently, the dataset was partitioned into training and testing subsets to enable an unbiased assessment of model performance.

3.2. Input variable selection

The selection of input variables was guided by geotechnical engineering principles, data availability, and evidence from previous empirical and machine-learning studies on shear wave velocity (V_s) prediction. The objective was to identify predictors that are both physically meaningful and readily obtainable from routine site investigation programs while maintaining an efficient model structure [12, 39, 40].

Accordingly, depth (Z), standard penetration test blow count (N), and shear wave velocity of the upper soil layer (V_{su}) were selected as the input variables, whereas V_s was considered the target output parameter. Depth was incorporated to account for variations in effective stress and soil stiffness with increasing overburden pressure. The SPT-N value was included due to its widespread use as an indicator of soil density, strength, and stiffness, while V_{su} was considered to capture near-surface stratigraphic conditions and wave propagation characteristics that influence subsurface shear wave velocity [11, 38].

To assess the suitability of the selected predictors, preliminary statistical analyses, including correlation and multicollinearity assessments, were performed. The results confirmed that the selected variables exhibited significant relationships with V_s while maintaining acceptable levels of interdependence, thereby minimizing redundancy within the input space [41, 42].

3.3. Model development framework (GMDH)

The group method of data handling (GMDH) is a self-organizing neural network approach capable of modeling complex nonlinear relationships between input and output variables [26]. In a classical GMDH network, neurons are generated by combining different pairs of input variables through polynomial transfer functions at each layer, thereby producing new neurons for the subsequent layer. Through this iterative process, the network gradually identifies the best-performing model structure and establishes functional relationships between the input parameters and the target output [43].

The primary objective of the GMDH algorithm is to determine an approximating function $\hat{f}$ that best represents the actual functional relationship f between the input vector $X = (x_1, x_2, x_3, \dots, x_n)$ and the output variable. Accordingly, the predicted output $\hat{y}$ is estimated such that the error between the predicted and observed values is minimized [26]. For a dataset containing M observations, the relationship between the input and output variables can be expressed as Eq. 1.

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) \quad \text{where} \quad (i = 1, 2, \dots, M) \tag{1}$$

where y_i represents the observed output corresponding to the *i-th* observation, and x_{ij} denotes the *j-th* input variable associated with that observation.

In this step, a GMDH-type NN is trained to predict the output value ($\hat{y}_i$) for the given input vector, $X = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in})$ as shown in Eq. 2.

$$\hat{y}_i = \hat{f}(x_{i1}, x_{i2}, x_{i3}, \ldots, x_{in}) \quad \text{where} \quad (i = 1, 2, \ldots, M) \tag{2}$$

A polynomial function is defined to minimize the squared error between observed and predicted outputs, as shown in Eq. 3.

$$\sum_{i=1}^{M}\left[\hat{f}(x_{i1}, x_{i2}, \ldots, x_{in}) - y_i\right]^2 \rightarrow min \tag{3}$$

In the GMDH algorithm, a discrete form of the Volterra functional series formulates the general connection between input and output variables. This function is the discrete equivalent of the Kolmogorov-Gabor polynomial [26, 44, 45]. Hence:

$$y = a_0 + \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \cdots \tag{4}$$

A system of partial quadratic polynomials including only two input variables can be used for representing this general form of mathematical description as shown in Eq. 5.

$$\hat{y} = G(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{5}$$

The coefficients a_i in Eq. 5 are calculated using regression techniques so that the difference between the real output, y and the predicted one, $\hat{y}$ for each pair of x_i and x_j as input variables is minimized [26, 45]. In this way, for every quadratic function, $\hat{y}_i$, the coefficients a_i are calculated to optimally fit the output in the whole set of input-output data pairs, as shown in Eq. 6.

$$E = \frac{1}{M}\sum_{i=1}^{M}(y_i - \hat{y}_i)^2 \rightarrow min \tag{6}$$

Only the neurons exhibiting the lowest prediction error according to an external validation criterion are retained for the subsequent layer, while the remaining neurons are discarded [42]. This self-organizing mechanism enables automatic determination of the best-performing network architecture and prevents excessive model complexity. In the fundamental form of the GMDH algorithm, all possible combinations of two input variables are considered to construct a regression polynomial that best fits the observed output data ($y_i, i = 1, 2, \ldots, M$) based on the least-squares criterion [26]. The multiple regression analysis together with the least-squares technique is then employed to determine the optimal vector of coefficients for the polynomial model [26, 43, 45]. This can be expressed in the form of Eq. 5 as shown in Eq. 7.

$$a = (A^T A)^{-1} A^T Y \tag{7}$$

where A is the design matrix, and Y is the vector of observed outputs.

In the present study, depth (Z), standard penetration test blow count (N), and shear wave velocity of the upper soil layer (V_{su}) were used as input variables, whereas shear wave velocity (V_s) was considered the target output parameter. Compared with conventional ANN and ensemble learning methods such as random forest and XGBoost, GMDH provides explicit polynomial representations while automatically determining the best-performing network architecture. This capability enhances model interpretability, reduces the risk of overfitting, and improves predictive reliability, making GMDH particularly suitable for heterogeneous geotechnical datasets and nonlinear soil behavior modeling [42, 46].

3.4. Genetic algorithm (GA) optimization procedure

To enhance the predictive performance of the GMDH model and improve its generalization capability, a genetic algorithm (GA) was employed as a global optimization tool. GA is an evolutionary search technique that has been widely applied for optimizing machine-learning models in complex nonlinear engineering problems due to its ability to efficiently explore large solution spaces and avoid local optima [47]. Within the proposed hybrid framework, each chromosome represents a candidate configuration of the GMDH model, including parameters related to network architecture and neuron selection. An initial population of candidate solutions was randomly generated to ensure sufficient diversity and facilitate effective exploration of the search space. The fitness of each chromosome was evaluated based on the root mean square error (RMSE) between observed and predicted shear wave velocity (V_s) values, with lower RMSE values indicating superior model performance [42].

The optimization process was performed using the basic genetic operators of selection, crossover, and mutation. In each generation, candidate solutions with better fitness values were more likely to be selected for reproduction, while crossover and mutation were applied to generate new solutions and maintain diversity within the population. The process was repeated until the stopping criteria were satisfied, such as reaching the maximum number of generations or observing no significant improvement in the fitness function. The best solution obtained by the genetic algorithm (GA) was then incorporated into the GMDH framework to develop the final hybrid GMDH-GA model. By automatically optimizing the network structure and model parameters, the GA improved the prediction capability of the model and reduced the possibility of overfitting. Similar applications of evolutionary optimization techniques have demonstrated their effectiveness in enhancing the performance of GMDH-based and neural-network-based predictive models in engineering problems [38, 46]. Similar studies have shown that evolutionary optimization algorithms can substantially improve the predictive capability of GMDH-based and neural-network models by optimizing model structure, enhancing convergence behavior, and reducing the likelihood of premature convergence and overfitting [38, 46].

3.5. Training and testing strategy (70/30 Split)

To evaluate the predictive capability of the proposed GMDH-GA model, the available dataset was partitioned into two mutually exclusive subsets using a holdout validation strategy. Specifically, 70% of the data were randomly assigned to the training set for model development, while the remaining 30% were reserved as an independent testing set for performance evaluation. This partitioning ratio is commonly adopted in machine-learning applications because it provides an effective balance between model calibration and unbiased performance assessment [38, 48]. The training subset was used to establish the GMDH network structure and optimize its parameters through the GA procedure. To prevent data leakage, no information from the testing dataset was utilized during model development or parameter tuning. The testing subset was subsequently employed to evaluate the predictive performance of the developed model on previously unseen observations, thereby providing an objective assessment of its practical applicability and generalization capability [39, 40, 47].

Prior to data partitioning, the dataset was randomly shuffled to preserve the statistical characteristics of the original database and minimize sampling bias. Furthermore, a 5-fold cross-validation procedure was incorporated within the training phase to enhance model robustness and reduce the risk of overfitting [49]. This additional validation step improves the stability and reliability of the optimization process, particularly when dealing with heterogeneous geotechnical datasets [35].

3.6. Model validation and performance evaluation metrics

The performance of the proposed GMDH-GA model was evaluated using a comprehensive validation framework designed to assess both predictive accuracy and generalization capability. To ensure an unbiased assessment, model performance was examined using an independent testing dataset that was not involved in either model training or parameter optimization. Such an approach is widely recognized as an effective means of evaluating the robustness and practical applicability of machine-learning models on previously unseen data [38, 48]. To quantitatively assess prediction performance, four widely used statistical indicators were employed: the coefficient of determination (R^2), root mean square error (RMSE), mean absolute deviation (MAD), mean absolute percentage error (MAPE), and mean squared error (MSE). These metrics provide complementary information regarding prediction accuracy, error magnitude, and model reliability.

The performance indicators were calculated as follows:

$$\mathrm{R}^2 = 1 - \left[\frac{\sum_{i=0}^{n}\left(Y_{i(Predicted)} - Y_{i(Measured)}\right)^2}{\sum_{i=1}^{n}\left(Y_{i(Actual)}\right)^2}\right] \tag{8}$$

$$\mathrm{RMSE} = \left[\frac{\sum_{i=0}^{n}\left(Y_{i(Predicted)} - Y_{i(Measured)}\right)^2}{n}\right]^{1/2} \tag{9}$$

$$\mathrm{MAD} = \frac{\sum_{i=1}^{n}\left|Y_{i(Predicted)} - Y_{i(Measured)}\right|}{n} \tag{10}$$

$$\mathrm{MAPE} = \frac{1}{n}\left[\frac{\sum_{i=1}^{n}\left|Y_{i(Predicted)} - Y_{i(Measured)}\right|}{\sum_{i=1}^{n} Y_{i(Measured)}}\right] \times 100 \tag{11}$$

$$\mathrm{MSE} = \frac{\sum_{i=0}^{n}\left(Y_{i(Predicted)} - Y_{i(Measured)}\right)^2}{n} \tag{12}$$

where n is the number of observations. The coefficient of determination (R^2) measures the proportion of variance in the observed data explained by the model, with values approaching unity indicating superior predictive capability. RMSE quantifies the overall magnitude of prediction errors and is particularly sensitive to large deviations. MAE provides a direct measure of the average absolute prediction error, while MAPE expresses prediction errors in percentage terms, enabling relative performance comparisons across different datasets and modeling approaches [50, 51].

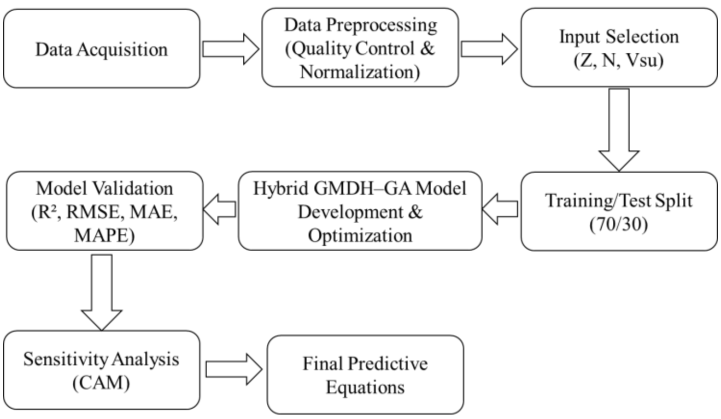


Fig. 2. Overall methodological framework of the proposed hybrid GMDH-GA model for shear wave velocity (V_s) prediction.

The overall methodological framework of the proposed GMDH-GA hybrid model for shear wave velocity (V_s) prediction, including data collection, preprocessing, input variable selection, data normalization, training/testing partitioning, GMDH model development, GA-based optimization, model validation, comparison with existing empirical correlations, and sensitivity analysis, is shown in Fig. 2.

4. Modelling shear wave velocity using GMDH neural networks

The feed-forward GMDH-type neural network for predicting shear wave velocity (V_s) was developed using the experimental dataset described in the previous sections. The input variables included depth (Z), SPT blow count (N), and shear wave velocity of the upper soil layer (V_{su}), while the target output was the corresponding shear wave velocity of the soil (V_s). To evaluate the predictive performance of the model, the dataset was randomly divided into training and testing subsets with similar statistical characteristics. For each soil group (all soils, clay, silt, sand, and gravel), 70% of the data were allocated for model training and 30% for testing and validation (Table 2) [48, 49].

Table 2. Descriptive statistics of input and output parameters used in the testing set.

Variable	Minimum	Maximum	Mean	Groups
Z	1	49	16.65	(1) All soils
N	8	86	45.19	
V_{su}	110	975	493.19	
V_s	141	990	527.68	
Z	1.50	56	20.75	(2) Clay
N	12.00	84	41.76	
V_{su}	111.30	858	513.16	
V_s	166.00	886.8	548.76	
Z	1.5	45	20.29	(3) Silt
N	6	76	34.44	
V_{su}	161	810	495.39	
V_s	187	875	521.50	
Z	1.5	41.5	15.47	(4) Sand
N	7	86	47.40	
V_{su}	110	859.9	508.47	
V_s	141	945	538.08	
Z	1	38	15.07	(5) Gravel
N	10	86	48.83	
V_{su}	134.57	947	504.32	
V_s	236.5	975	541.99	

The GMDH-type neural networks were employed to construct explicit polynomial relationships between input variables and V_s. In this study, a genetic algorithm (GA) was integrated into the GMDH framework to optimize network architecture and model parameters. The GA controls key evolutionary parameters, including population size, number of generations, crossover probability, and mutation probability. After preliminary sensitivity analysis, a population size of 100, crossover probability of 0.9, and mutation probability of 0.1 were adopted over 300 generations, as no significant improvement was observed beyond this configuration, which is consistent with commonly reported GA settings in engineering optimization studies [47].

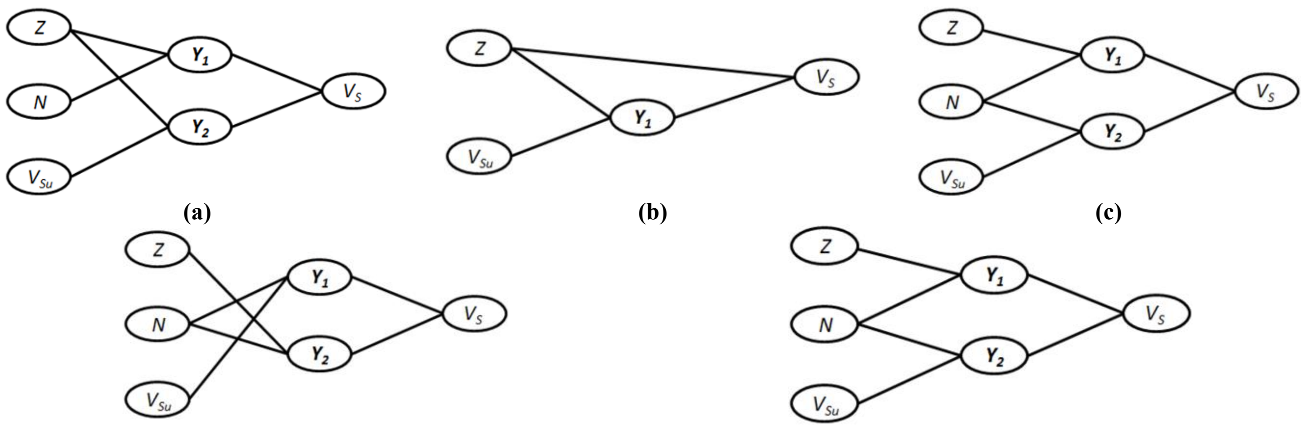


Fig. 3. Optimized single-hidden-layer GMDH network structures for the V_s: (a) all soils; (b) clay; (c) silt; (d) sand; and (e) gravel.

The hybrid GMDH-GA framework was applied separately to five soil groups in order to capture soil-specific behavior. Although multiple network architectures with different numbers of hidden layers were tested, single-hidden-layer structures were selected as they provided an optimal balance between predictive accuracy and model simplicity. The final optimized network structures for each soil group are illustrated in Fig. 3. Although most models retained all input variables during construction, the GMDH algorithm inherently performs automated feature selection by retaining only the most influential terms during polynomial formation. This capability allows the generation of explicit mathematical expressions while maintaining high predictive performance [26, 46].

The resulting polynomial equations for shear wave velocity prediction are presented in Eqs. 13 to 17. Similar explicit polynomial formulations have been successfully reported in previous GMDH-based geotechnical prediction studies. For the case of all soil types, intermediate variables Y_1 and Y_2 are defined as Eqs. 13a and 13b.

for all soils:

$$Y_1 = 203.0504 + 20.8002Z + 3.6121N - 0.2484Z^2 - 0.02666N^2 - 0.02974ZN \tag{13a}$$

$$Y_2 = 124.1691 + 1.6931Z + 0.8119V_{su} - 0.05386Z^2 - 0.000089{V_{su}}^2 + 0.007775ZV_{su} \tag{13b}$$

$$V_s = -88.1546 + 0.5184Y_1 + 0.8366Y_2 - 0.000642{Y_1}^2 - 0.0000214{Y_2}^2 + 0.000322Y_1Y_2 \tag{13c}$$

Similar formulations were obtained for clayey, silty, sandy, and gravelly soils, as shown in Eqs. 14 to 17. In all cases, depth (Z) and shear wave velocity of the upper layer (V_{su}) are expressed in meters and m/s, respectively. The coefficients of the polynomial equations were automatically determined through the combined GMDH-GA optimization process.

For clayey soils:

$$Y_1 = 101.2230 - 3.1997Z + 0.9595V_{su} - 0.077789Z^2 - 0.0003378{V_{su}}^2 + 0.0130567ZV_{su} \tag{14a}$$

$$V_s = 8.1441 + 0.8703Z + 0.9329Y_1 + 0.01513Z^2 + 0.000119{Y_1}^2 - 0.002897ZY_1 \tag{14b}$$

For silty soils:

$$Y_1 = 219.3582 + 19.3295Z + 1.4077N - 0.223459Z^2 - 0.021702N^2 + 0.033419ZN \tag{15a}$$

$$Y_2 = 101.36088 - 0.64436N + 0.824197V_{su} + 0.0065467N^2 + 0.0000761{V_{su}}^2 + 0.0000631NV_{su} \tag{15b}$$

$$V_s = 48.47287 - 0.27349Y_1 + 1.071459Y_2 - 0.0000827{Y_1}^2 - 0.000437{Y_2}^2 + 0.000735Y_1Y_2 \tag{15c}$$

For sandy soils:

$$Y_1 = 127.4764 + 0.400969N + 0.68966V_{su} - 0.0019389N^2 + 0.0001691{V_{su}}^2 - 0.0001319NV_{su} \tag{16a}$$

$$Y_2 = 195.4441 + 23.76196Z + 3.32919N - 0.379395Z^2 - 0.0270525N^2 - 0.007853967ZN \tag{16b}$$

$$V_s = -127.93006 + 0.55972Y_1 + 0.961276Y_2 - 0.000461927{Y_1}^2 - 0.00176247{Y_2}^2 + 0.00172848Y_1Y_2 \tag{16c}$$

For gravely soils:

$$Y_1 = 244.4176 + 24.5336Z + 1.93860N - 0.312301Z^2 - 0.0136057N^2 - 0.038851ZN \tag{17a}$$

$$Y_2 = 226.2265 - 1.516388N + 0.51278V_{su} + 0.00534585N^2 + 0.000228235{V_{su}}^2 + 0.00197189NV_{su} \tag{17b}$$

$$V_s = -160.70616 + 0.877126Y_1 + 0.688160Y_2 - 0.000411841{Y_1}^2 + 0.000452851{Y_2}^2 - 0.000531079Y_1Y_2 \tag{17c}$$

To evaluate model performance, predicted and measured values of V_s were compared for both training and testing datasets. Scatter plots (Figs. 4 and 5) indicate that the predicted values are closely distributed around the 1:1 reference line, confirming strong agreement between measured and modeled data.

The predictive performance of the developed models was quantitatively evaluated using standard statistical metrics, including the coefficient of determination (R^2), root mean square error (RMSE), mean absolute deviation (MAD), mean absolute percentage error (MAPE), and mean squared error (MSE), defined in Eqs. 8 to 12. The corresponding performance results for all soil groups are summarized in Table 3.

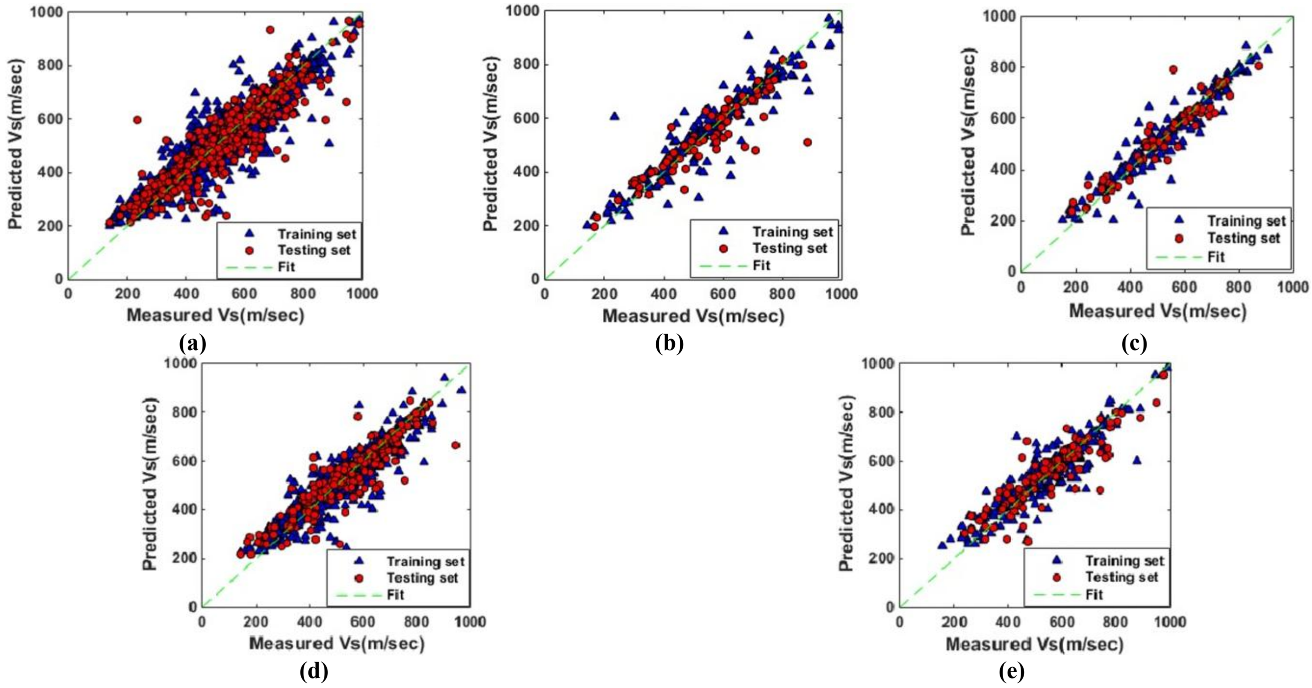


Fig. 4. The relationships between measured and predicted V_s by GMDH through the training and testing process: (a) all soils; (b) clay; (c) silt; (d) sand; and (e) gravel.

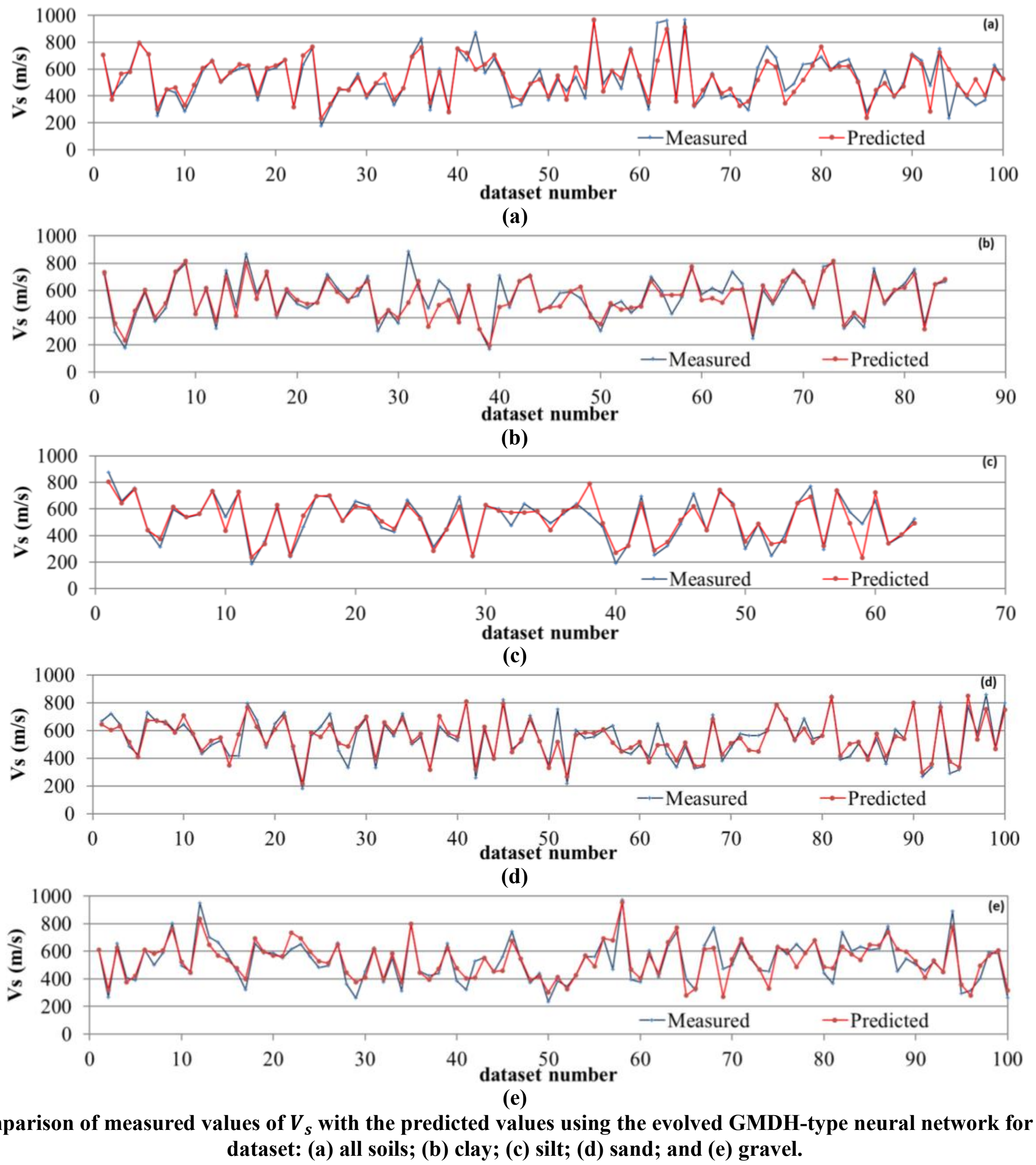


Fig. 5. Comparison of measured values of V_s with the predicted values using the evolved GMDH-type neural network for the testing dataset: (a) all soils; (b) clay; (c) silt; (d) sand; and (e) gravel.

Table 3. The amount of performance indexes for the GMDH models.

Groups	Set	Number of data	MAD (m/s)	MAPE (%)	MSE $(m/s)^2$	RMSE (m/s)	R^2	Correlation coefficient
All soils	Testing set	597	44.5	8.43	4275.91	65.39	0.99	0.91
	Training set	1394	44.22	8.3	3862.85	62.15	0.99	0.92
Clay	Testing set	125	44.15	8.05	4748.68	68.91	0.99	0.90
	Training set	293	42.95	7.69	4276.98	65.4	0.99	0.93
Silt	Testing set	51	39.22	7.52	3781.47	61.49	0.99	0.93
	Training set	118	39.79	7.69	2909.26	53.94	0.99	0.95
Sand	Testing set	281	44.34	8.24	4003.52	63.27	0.99	0.91
	Training set	657	43.15	8.39	3643.2	60.36	0.99	0.91
Gravel	Testing set	140	49.89	9.21	4794.01	69.24	0.98	0.89
	Training set	326	44.95	8.32	3947.26	62.83	0.99	0.90

These performance indicators are widely employed for evaluating machine-learning models in geotechnical engineering and provide complementary information regarding prediction accuracy, error magnitude, and model robustness.

The obtained results demonstrate high predictive accuracy and low error levels for both training and testing datasets across all soil categories. The consistency between training and testing performance indicates that the proposed GMDH-GA models exhibit strong generalization capability and indicate no significant evidence of overfitting. Moreover, the ability of the model to generate explicit polynomial equations enhances its practical applicability in geotechnical engineering design. Overall, the results confirm that the hybrid GMDH-GA framework provides a reliable, accurate, and interpretable tool for predicting shear wave velocity profiles in heterogeneous alluvial deposits.

A comparison between measured and predicted V_S values revealed that the developed GMDH-GA models successfully captured the nonlinear relationships among depth, SPT-N value, and upper-layer shear wave velocity across different soil categories. The resulting polynomial expressions provide transparent predictive formulations that can be directly implemented in engineering practice without requiring retraining procedures or specialized computational platforms.

The accuracy and predictive performance of the proposed GMDH model for group 1 (all soils) were statistically evaluated and compared with previously published empirical correlations listed in Table 1. The analysis was carried out using the full dataset of 1,991 samples, and the results presented in Table 4 clearly indicate that the GMDH-based approach outperforms all existing empirical correlations. Similar improvements achieved by hybrid machine-learning models over conventional empirical relationships have been reported in recent geotechnical studies. Accordingly, the developed model demonstrates higher predictive accuracy and enhanced reliability in estimating shear wave velocity (V_S).

Table 4. Performance indexes for GMDH model compared with other empirical correlations.

Equation of this author for all soils	No. of data	MAD (m/s)	MAPE (%)	MSE $(m/s)^2$	RMSE (m/s)	R^2	Correlation coefficient
Ohsaki and Iwasaki [6]	1991	194.25	36.57	54930.87	234.37	0.82	0.28
Imai [7]	1991	218.24	41.08	66361.26	257.61	0.78	0.28
Ohta and Goto [1]	1991	224.21	42.21	69410.08	263.46	0.77	0.28
Seed and Idriss [8]	1991	161.96	30.49	41250.88	203.10	0.87	0.28
Lee Shannon [9]	1991	344.42	64.84	141689.82	376.42	0.54	0.28
Iyisan [10]	1991	188.93	35.57	53192.95	230.64	0.83	0.28
Jafari et al. [5]	1991	171.82	32.35	44857.26	211.80	0.85	0.27
Hasancebi and Ulusay [11]	1991	249.30	46.93	82512.25	287.25	0.73	0.28
Dikmen [12]	1991	282.22	53.13	101196.46	318.11	0.67	0.28
Maheswari et al. [13]	1991	241.33	45.43	78236.13	279.71	0.74	0.28
Hafezi et al. [14]	1991	231.26	43.54	75992.59	275.67	0.75	0.28
Chatterjee and Choudhury [15]	1991	215.07	40.49	64911.56	254.78	0.79	0.28
Fatehnia et al. [16]	1991	244.19	45.97	79786.08	282.46	0.74	0.28
This study (GMDH model)	1991	186.21	35.06	51013.35	225.86	0.83	0.28
This study (GMDH model)	1991	44.30	8.34	3986.74	63.14	0.99	0.92

The scatter plots presented in Fig. 6 further illustrate the comparison between measured and predicted V_S values. It is evident that most of the existing empirical correlations systematically underestimate V_S and exhibit considerable data dispersion around the 1:1 reference line. Among the evaluated relationships, the correlation proposed by Jafari et al. [5] exhibits comparatively better performance; however, noticeable scatter remains, indicating limitations in its predictive capability. This observation is consistent with previous findings suggesting that empirical V_S-SPT relationships often struggle to capture the inherent heterogeneity and nonlinear behavior of soil deposits.

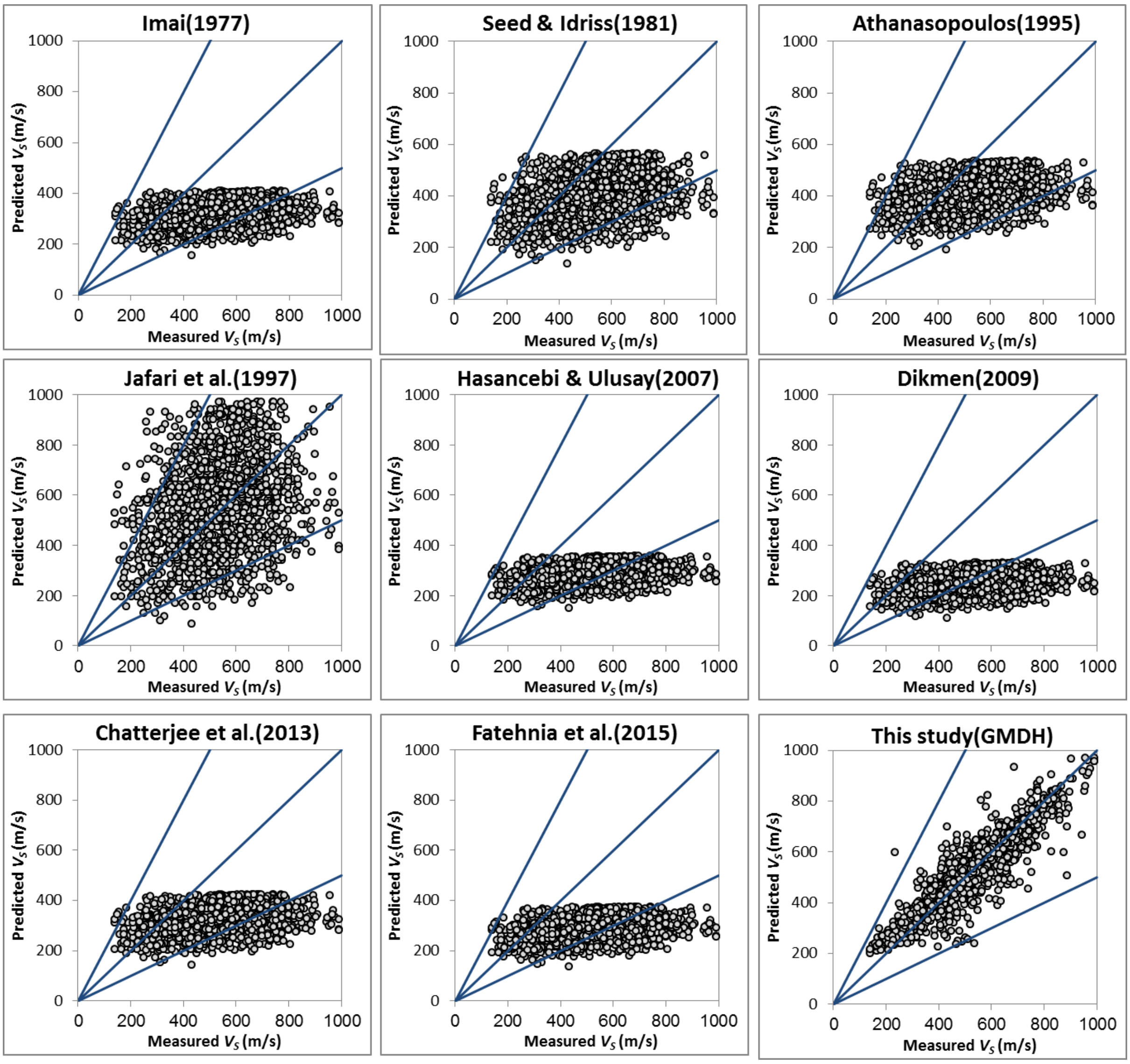


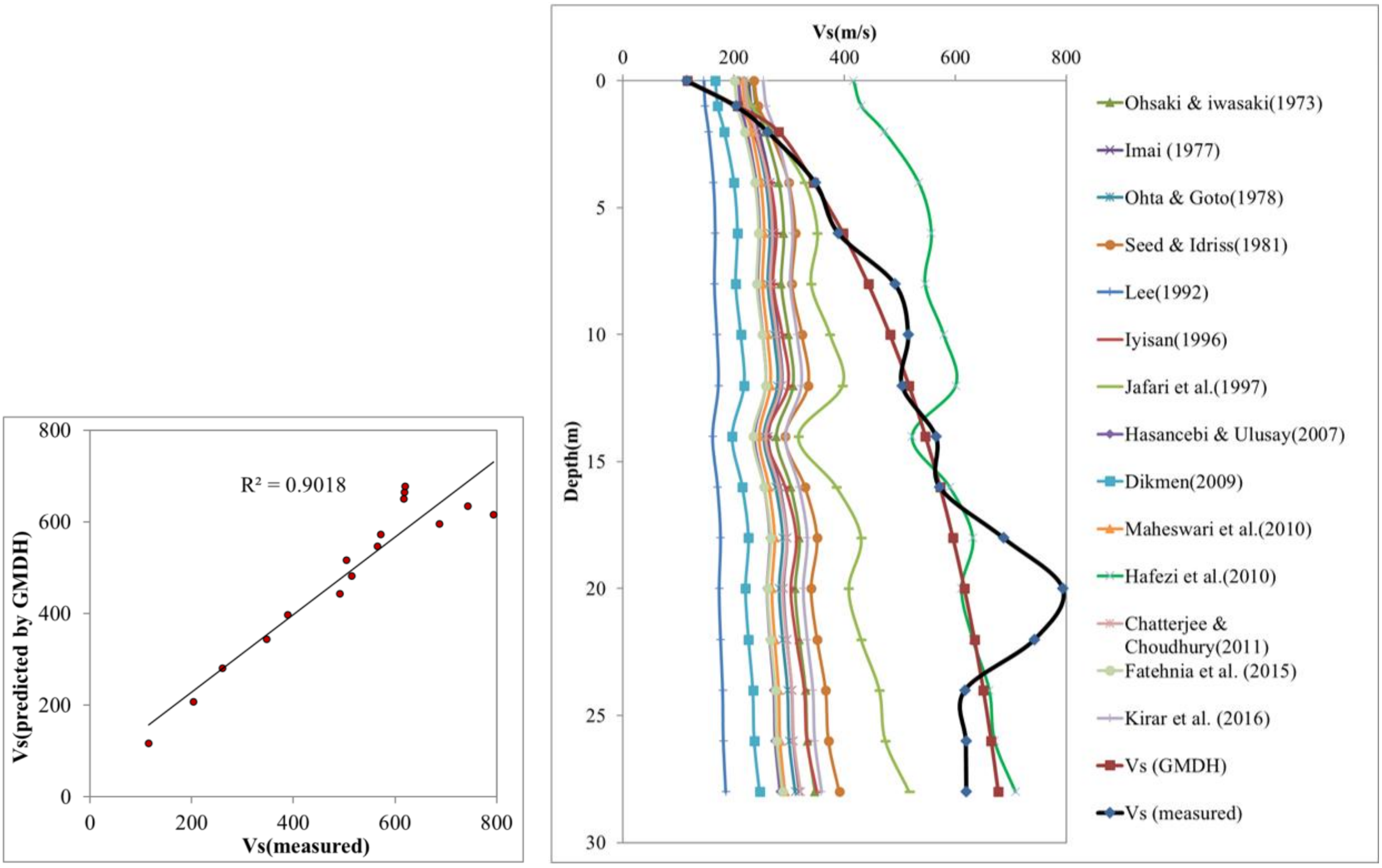
Fig. 6. Comparison of measured values of V_s with predicted values using different methods.

To further evaluate the capability of the proposed model in reproducing shear wave velocity variations with depth, a representative 30-m-deep borehole located in central Mashhad was analyzed using Eqs. 13a to 13c (Table 5). The coefficient of determination shown in Fig. 7a confirms that the GMDH model provides a close approximation to the measured V_s profile. In the stepwise implementation procedure, the assumption of $V_s = 0$ at the ground surface was adopted to initiate the calculations, and the predicted upper-layer velocity (V_{su}) was subsequently utilized as an input for estimating deeper soil layers.

Table 5. A sample of bore log profile and calculating procedure using Eqs. 13a to 13c.

Borehole ID	X	Y	Soil type	Depth	SPT	V_{su}	V_s (measured)	Y_1	Y_2	V_s (GMDH)
I6-3-1	726055	4026141	SM	0.5	15	0	115	261.4	123.3	116.69
I6-3-1	726055	4026141	SP-SM	1.5	16	116.7	204	284	216.4	207.10
I6-3-1	726055	4026141	SW-SM	3	19	207.1	261	320.5	287.8	280.72
I6-3-1	726055	4026141	SP	5	24	280.7	347	368.6	346.2	343.83
I6-3-1	726055	4026141	GP	7	26	343.8	389	407	397	397.29
I6-3-1	726055	4026141	GW	9	25	397.3	491	437.1	440.9	442.51
I6-3-1	726055	4026141	SP-SM	11	28	442.5	515	472.9	478.7	481.91
I6-3-1	726055	4026141	SC	13	30	481.9	504	504.2	512.4	516.22
I6-3-1	726055	4026141	SW-SM	15	23	516.2	565	517.9	542.3	545.91
I6-3-1	726055	4026141	SW-SM	17	29	545.9	572	552.5	568.7	572.29
I6-3-1	726055	4026141	SP-SM	19	33	572.3	687	580.1	592.6	595.46
I6-3-1	726055	4026141	SP-SM	21	31	595.5	794	597.3	614	616.11
I6-3-1	726055	4026141	GP-GM	23	33	616.1	743	617.6	633.3	634.35
I6-3-1	726055	4026141	GP	25	36	634.3	617	636.5	650.7	650.37
I6-3-1	726055	4026141	SM	27	37	650.4	619	651	666.1	664.64

Figure 7-b compares the measured downhole V_s profile with predictions obtained from the proposed GMDH model and previously developed empirical correlations. The results demonstrate that none of the existing relationships accurately reproduces the complete velocity profile throughout the investigated depth range. Although the model proposed by Hafezi et al. [14] exhibits partial agreement at greater depths, it tends to overestimate V_s values in the upper layers. In contrast, the proposed GMDH-based model consistently follows the measured trend and provides a more reliable representation of the full shear wave velocity profile. The improved performance can be attributed to the capability of GMDH to capture nonlinear interactions among depth, SPT-N values, and upper-layer shear wave velocity while simultaneously controlling model complexity through its self-organizing architecture.



(a) (b)

Fig. 7. (a) The relationship between the measured and calculated V_s using Eq. 8 for borehole I6-3-1; and (b) the measured and predicted V_s profile for borehole I6-3-1.

5. Sensitivity analysis

Since the polynomial equations generated by the evolved GMDH-type neural network are highly nonlinear and do not explicitly quantify the relative contribution of each input variable to the predicted shear wave velocity (V_s), a sensitivity analysis was performed to evaluate the influence of predictor variables on model output. Sensitivity analysis is widely used in machine learning-based geotechnical and geophysical studies to improve model interpretability and quantify the relative importance of input parameters [52, 53]. In this study, the cosine amplitude method (CAM) was employed to determine the relative strength of association between input variables and the output variable. The CAM is a simple yet effective statistical technique commonly used in artificial intelligence and data-driven modeling because of its ability to quantify variable dependency in multidimensional datasets [27, 54]. The relationship coefficient is defined as Eq. 18.

$$R_{ij} = \frac{\sum_{k=1}^{n}(x_{ik} \times x_{jk})}{\sqrt{\sum_{k=1}^{n} x_{ik}^2 \sum_{k=1}^{n} x_{jk}^2}} \tag{18}$$

where x_i and x_j represent the input and output variables, respectively, and n is the number of data samples. The coefficient R_{ij} ranges between 0 and 1, where values closer to zero indicate weak influence, while values approaching unity indicate strong correlation with the predicted shear wave velocity (V_s).

The sensitivity analysis results for the developed one-hidden-layer GMDH models are presented in Fig. 8. The results indicate that the shear wave velocity of the upper soil layer (V_{su}) is the most influential parameter in predicting V_s. Furthermore, the SPT blow count (N) exhibits higher sensitivity in coarse-grained soils, whereas its influence decreases in fine-grained deposits. In contrast, depth (Z) shows relatively lower sensitivity compared to the other input variables. Similar findings regarding the dominant role of near-surface shear wave velocity and SPT-N in V_s prediction have been reported in previous studies, confirming the physical consistency of the proposed model [12].

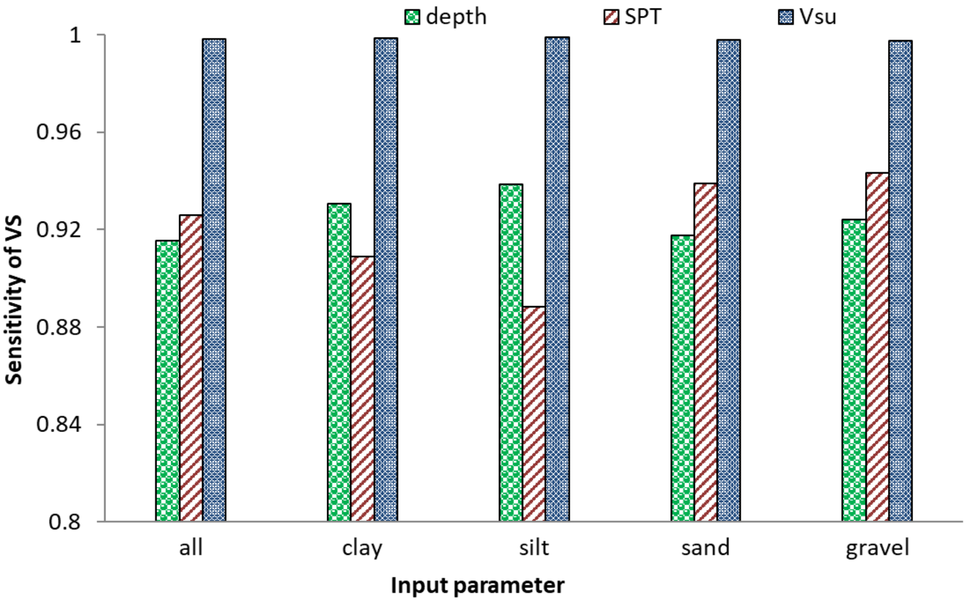


Fig. 8. Contributions of the predictor variables in the best-selected GMDH models for five soil groups under study.

6. Results and discussion

High-performance predictive models were developed for estimating shear wave velocity (V_s) using a hybrid GMDH-genetic algorithm (GMDH-GA) framework. The main objective was to derive explicit and reliable formulations for estimating the shear wave velocity of the upper 30 m soil profile (V_{s30}) in the absence of in-situ geophysical measurements. The V_{s30} parameter is widely recognized as a key index in seismic site classification and ground response analysis. Accordingly, V_s was modeled as a function of depth (Z), SPT blow count (N), and shear wave velocity of the upper soil layer (V_{su}), which were identified as dominant predictors in the sensitivity analysis.

The dataset was compiled from geotechnical and geophysical investigations conducted in Mashhad, comprising 226 boreholes and 1,991 data records representing clay, silt, sand, and gravel deposits. Separate GMDH-based polynomial models were developed for each soil group as well as for the combined dataset. For each case, 70% of the data were used for training and 30% for testing. Model performance was evaluated against previously published empirical correlations, demonstrating that the proposed GMDH-GA approach consistently provides higher accuracy and improved predictive capability across all soil types. Although deeper network structures with additional hidden layers showed marginal improvements in accuracy, single-hidden-layer models were selected as the final formulations due to their superior generalization capability and higher interpretability, which is a key requirement in engineering design applications.

Sensitivity analysis confirmed that V_{su} is the most influential predictor, and its exclusion significantly reduces model accuracy. This finding is consistent with previous studies highlighting the dominant role of near-surface shear wave velocity in soil stiffness characterization and wave propagation behavior. Additionally, SPT blow count exhibits a stronger influence in coarse-grained soils, whereas its effect decreases in fine-grained materials due to differences in fabric and compressibility characteristics. Depth (Z) shows a relatively lower but consistent contribution across all soil groups.

The inclusion of V_{su} is physically justified by the layered continuity of alluvial deposits and the progressive increase in soil stiffness with depth. In sedimentary environments, mechanical properties of adjacent layers are strongly interdependent due to stress history and depositional processes. Therefore, incorporating V_{su} captures the sequential nature of shear wave propagation in stratified media rather than introducing redundant information. However, this variable may limit direct applicability in sites lacking preliminary V_s measurements; hence, the proposed framework is particularly suitable for profile reconstruction, interpolation, and data-enhanced geotechnical site characterization.

Overall, the results demonstrate that the proposed GMDH-GA models provide reliable and robust predictions of shear wave velocity profiles and can serve as a complementary or alternative tool to conventional in-situ testing methods such as downhole and crosshole measurements. Since V_{su} profiles play a critical role in seismic microzonation, site response analysis, liquefaction potential assessment, and seismic site classification, the proposed models offer an efficient and practical tool for geotechnical earthquake engineering applications. Comparison with existing empirical correlations further confirms the superiority of region-specific modeling approaches, as global empirical relationships show significant deviations from measured data, whereas the proposed hybrid V_{su} framework exhibits strong agreement with downhole observations.

7. Conclusions

This study developed a hybrid GMDH-genetic algorithm (GMDH-GA) framework for predicting shear wave velocity (V_s) profiles in heterogeneous alluvial deposits using a comprehensive geotechnical database from Mashhad, Iran. The proposed models were constructed using depth (Z), standard penetration test blow count (SPT-N), and upper-layer shear wave velocity (V_{su}) as input parameters and were evaluated across different soil groups, including clay, silt, sand, and gravel.

The results demonstrated that the proposed GMDH-GA models achieve high predictive accuracy and consistently outperform conventional empirical correlations in terms of error reduction and generalization capability. These findings are consistent with recent studies highlighting the superiority of hybrid machine learning approaches over traditional empirical relationships in geotechnical parameter estimation. Sensitivity analysis indicated that V_{su} is the most influential predictor in V_s estimation, followed by SPT blow count, particularly in coarse-grained soils. In contrast, depth exhibited a comparatively lower but stable contribution across all soil types, reflecting the gradual variation of stress-dependent soil stiffness with depth.

Unlike conventional black-box machine learning models, the proposed GMDH-GA framework provides explicit polynomial expressions, thereby enhancing model transparency, interpretability, and practical usability in geotechnical engineering design. Overall, the developed models offer a reliable and efficient tool for seismic site characterization, seismic microzonation, ground response analysis, and preliminary site classification in regions where direct geophysical measurements are limited. The proposed approach bridges the gap between empirical correlations and advanced artificial intelligence techniques by combining predictive accuracy with physical interpretability.

8. Limitations and future work

Although the proposed GMDH-GA framework demonstrates strong predictive performance, several limitations should be acknowledged. First, the developed models are based on data from a specific geological setting (Mashhad alluvial deposits), which may limit their direct applicability to regions with different geological and stratigraphic conditions. Therefore, external validation in other geological environments is necessary before broader application. Second, the inclusion of V_{su} improves predictive accuracy but may limit the applicability of the model in sites where prior shear wave velocity information is not available. This reflects a trade-off between model accuracy and field data dependency. Third, as with most data-driven approaches, model performance is influenced by the quality and uncertainty of input data, particularly SPT measurements and in-situ geophysical observations.

Furthermore, future research should focus on validating the proposed framework across diverse geological conditions and incorporating additional geotechnical and geophysical parameters such as CPT data and soil index properties. A direct comparative analysis with modern machine-learning algorithms such as ANN, random forest, and XGBoost was beyond the scope of the present study and should be addressed in future investigations.

Statements & declarations

Author contributions

The author contributed to all stages of the research, including conceptualization, methodology, analysis, interpretation of results, writing the original draft, and revising and editing the manuscript. The author has read and approved the final version of the manuscript.

Acknowledgments

The author expresses sincere appreciation to all individuals who provided scientific, technical, or logistical assistance during the course of this research.

Funding

This research received no financial support from any governmental, commercial, or non-profit funding agency.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

The author declares no conflict of interest.

References

[1] Ohta, Y., Goto, N. Empirical shear wave velocity equations in terms of characteristic soil indexes. Earthquake Engineering & Structural Dynamics, 1978; 6: 167–187. doi:10.1002/eqe.4290060205.

[2] Imai, T., Tonouchi, K. Correlation of N-value with S-wave velocity and shear modulus. In: Proceedings of the 2nd European Symposium on Penetration Testing (ESOPT II); 1982 May 24–27; Amsterdam, Netherlands. p. 67–72.

[3] Seed, H. B., Idriss, I. M., Arango, I. Evaluation of Liquefaction Potential Using Field Performance Data. Journal of Geotechnical Engineering, 1983; 109: 458–482. doi:10.1061/(ASCE)0733-9410(1983)109:3(458).

[4] Sykora, D. E., H., S. K. Correlations of in-situ measurements in sands of shear wave velocity with standard penetration test resistance. Geotechnical Testing Journal, 1983; 6: 107–114. doi:10.1520/GTJ10700J.

[5] Jafari, M. K., Asghari, A., Rahmani, I. Empirical correlation between shear wave velocity (Vs) and SPT-N value for south of Tehran soils. In: Proceedings of 4th International Conference on Civil Engineering; 1997 May 04–06; Tehran, Iran.

[6] Ohsaki, Y., Iwasaki, R. On Dynamic Shear Moduli and Poisson's Ratio of Soil Deposits. Soils and Foundations. Soils and Foundations, 1973; 13: 61–73. doi:10.3208/sandf1972.13.4_61.

[7] Imai, T. P- and S-Wave Velocities of the Ground in Japan. In: 9th International Conference on Soil Mechanics and Foundation Engineering; 1977 Jul 10–15; Tokyo, Japan. p. 257–260.

[8] Seed, H. B., Idriss, I. M. Evaluation of Liquefaction Potential Sand Deposits Based on Observation of Performance in Previous Earthquakes. ASCE National Convention (MO), 481-544.

[9] Lee Shannon, H.-H. Analysis of the multicollinearity of regression equations of shear wave velocities. Soils and Foundations, 1992; 32: 205–214. doi:10.3208/sandf1972.32.205.

[10] Iyisan, R. Correlations between shear wave velocity and in-situ penetration test results. Teknik Dergi-tmmob Insaat Muhendisleri Odas, 1996; 7: 371–374.

[11] Hasancebi, N., Ulusay, R. Empirical correlations between shear wave velocity and penetration resistance for ground shaking assessments. Bulletin of Engineering Geology and the Environment, 2007; 66: 203–213. doi:10.1007/s10064-006-0063-0.

[12] Dikmen, Ü. Statistical correlations of shear wave velocity and penetration resistance for soils. Journal of Geophysics and Engineering, 2009; 6: 61. doi:10.1088/1742-2132/6/1/007.

[13] Maheswari, R. U., Boominathan, A., Dodagoudar, G. R. Use of Surface Waves in Statistical Correlations of Shear Wave Velocity and Penetration Resistance of Chennai Soils. Geotechnical and Geological Engineering, 2010; 28: 119–137. doi:10.1007/s10706-009-9285-9.

[14] Hafezi, M. N., Azadi, A., Amanian, M. Assessment of shear wave velocity based on standard penetration and accuracy of the results in the range of Mashhad City. In: Fourteenth Conference of Iran Geophysics, Geophysical Institute (in Persian); 2010 May 13; Tehran, Iran.

[15] Chatterjee, K., Choudhury, D. Variations in shear wave velocity and soil site class in Kolkata city using regression and sensitivity analysis. Natural Hazards, 2013; 69: 2057–2082. doi:10.1007/s11069-013-0795-7.

[16] Fatehnia, M., Hayden, M., Landschoot, M. Correlation between Shear Wave Velocity and SPT-N Values for North Florida Soils. Electronic Journal of Geotechnical Engineering, 2015; 20: 12421–12430.

[17] Kirar, B., Maheshwari, B. K., Muley, P. Correlation Between Shear Wave Velocity (Vs) and SPT Resistance (N) for Roorkee Region. International Journal of Geosynthetics and Ground Engineering, 2016; 2: 9. doi:10.1007/s40891-016-0047-5.

[18] Sarkar, A., Mazumdar, A., Bhowmik, D. Application of Soft Computing Techniques for Prediction of Seismic Wave Velocity. In: 9th International Conference on Control, Decision and Information Technologies (CoDIT); 2023 Jul 03–06; Rome, Italy. p. 2532–2537. doi:10.1109/CoDIT58514.2023.10284120.

[19] Mousavi, Z., Bayat, M., Yang, J., Feng, W.-Q. Evaluation of empirical and machine learning models for predicting shear wave velocity of granular soils based on laboratory element tests. Soil Dynamics and Earthquake Engineering, 2024; 183: 108805. doi:10.1016/j.soildyn.2024.108805.

[20] Jin, L., Fuggle, A., Roberts, H., Armstrong, C., Pua, L.-M. Estimating In Situ Shear Wave Velocity Using Machine Learning Techniques. In: Geo-Congress 2024; 2024 Feb 25–28; Vancouver, Canada. p. 436–444. doi:10.1061/9780784485347.044.

[21] Joshi, A., Raman, B., Mohan, C. K., Cenkeramaddi, L. R. A new machine learning approach for estimating shear wave velocity profile using borelog data. Soil Dynamics and Earthquake Engineering, 2024; 177: 108424. doi:10.1016/j.soildyn.2023.108424.

[22] Zhong, K., Tan, X., Liu, S., Lu, Z., Hou, X., Wang, Q. Prediction of slope failure probability based on machine learning with genetic-ADASYN algorithm. Engineering Geology, 2025; 346: 107885. doi:10.1016/j.enggeo.2024.107885.

[23] Tabari, M. M. R., Azari, T., Dehghan, V. A supervised committee neural network for the determination of aquifer parameters: a case study of Katasbes aquifer in Shiraz plain, Iran. Soft Computing, 2021; 25: 4785–4798. doi:10.1007/s00500-020-05487-2.

[24] Farahani, S. A Data-Driven Seismic Design Framework for RC Frames Using Wavelet Transforms and Machine Learning. Civil Engineering and Applied Solutions, 2027; 3: 1–24. doi:10.22080/ceas.2026.30950.1073.

[25] Sadeghpour Haji, M., Niknam, R., Shayanfar, J. ANN-Based Modeling of Shear Behavior of Reinforced Concrete Columns under Constant Axial Loads. Civil Engineering and Applied Solutions, 2026; 2: 28–48. doi:10.22080/ceas.2025.30228.1050.

[26] Ivakhnenko, A. G. Polynomial Theory of Complex Systems. IEEE Transactions on Systems, Man, and Cybernetics, 1971; SMC-1: 364–378. doi:10.1109/TSMC.1971.4308320.

[27] Gandomi, A. H., Roke, D. A. Assessment of artificial neural network and genetic programming as predictive tools. Advances in Engineering Software, 2015; 88: 63–72. doi:10.1016/j.advengsoft.2015.05.007.

[28] MolaAbasi, H., Khajeh, A., Jamshidi Chenari, R. Use of GMDH-type neural network to model the mechanical behavior of a cement-treated sand. Neural Computing and Applications, 2021; 33: 15305–15318. doi:10.1007/s00521-021-06157-6.

[29] Kordnaeij, A., Kalantary, F., Kordtabar, B., Mola-Abasi, H. Prediction of recompression index using GMDH-type neural network based on geotechnical soil properties. Soils and Foundations, 2015; 55: 1335–1345. doi:10.1016/j.sandf.2015.10.001.

[30] Kalantary, F., Ardalan, H., Nariman-Zadeh, N. An investigation on the Su–NSPT correlation using GMDH type neural networks and genetic algorithms. Engineering Geology, 2009; 104: 144–155. doi:10.1016/j.enggeo.2008.09.006.

[31] MolaAbasi, H., Saberian, M., Kordnaeij, A., Omer, J., Li, J., Kharazmi, P. Predicting the stress-strain behaviour of zeolite-cemented sand based on the unconfined compression test using GMDH type neural network. Journal of Adhesion Science and Technology, 2019; 33: 945–962. doi:10.1080/01694243.2019.1571659.

[32] Farrokhi, F., Firoozfar, A., Maghsoudi, M. S. Evaluation of liquefaction-induced lateral displacement using a GMDH-type neural network optimized by genetic algorithm. Arabian Journal of Geosciences, 2019; 13: 4. doi:10.1007/s12517-019-4980-1.

[33] Javdanian, H., Heidari, A., Kamgar, R. Energy-Based Estimation of Soil Liquefaction Potential Using GMDH Algorithm. Iranian Journal of Science and Technology, Transactions of Civil Engineering, 2017; 41: 283–295. doi:10.1007/s40996-017-0061-4.

[34] Building & housing research center (BHRC). standard No. 2800: Iranian code of practice for seismic resistant design of buildings. Tehran (IR): BHRC; 2007 (in Persian).

[35] Koopialipoor, M., Armaghani, D. J., Hedayat, A., Marto, A., Gordan, B. Applying various hybrid intelligent systems to evaluate and predict slope stability under static and dynamic conditions. Soft Computing, 2019; 23: 5913–5929. doi:10.1007/s00500-018-3253-3.

[36] Nguyen, M. D., Costache, R., Sy, A. H., Ahmadzadeh, H., Van Le, H., Prakash, I., Pham, B. T. Novel approach for soil classification using machine learning methods. Bulletin of Engineering Geology and the Environment, 2022; 81: 468. doi:10.1007/s10064-022-02967-7.

[37] Jiawei, H., Micheline, K., Jian, P. Data mining: Concepts and techniques. Waltham: Morgan Kaufmann; 2006. p. 559–569.

[38] Lee, S., Kim, J., Kang, H., Kang, D.-Y., Park, J. Genetic algorithm based deep learning neural network structure and hyperparameter optimization. Applied Sciences, 2021; 11: 744. doi:10.3390/app11020744.

[39] Kumar, A., Satyannarayana, R., Rajesh, B. G. Correlation between SPT-N and shear wave velocity (VS) and seismic site classification for Amaravati city, India. Journal of Applied Geophysics, 2022; 205: 104757. doi:10.1016/j.jappgeo.2022.104757.

[40] Daag, S., Halasan, O. P., Magnaye, A., Grutas, R., Solidum, R. Empirical Correlation between Standard Penetration Resistance (SPT-N) and Shear Wave Velocity (Vs) for Soils in Metro Manila, Philippines. Applied Sciences, 2022; 12: 8067. doi:10.3390/app12168067.

[41] Cheng, J., Sun, J., Yao, K., Xu, M., Cao, Y. A variable selection method based on mutual information and variance inflation factor. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2022; 268: 120652. doi:10.1016/j.saa.2021.120652.

[42] Sugiyama, M. Introduction to Statistical Machine Learning. 1st ed. Waltham (MA): Morgan Kaufmann; 2016. doi:10.1016/B978-0-12-802121-7.00012-1.

[43] Ivakhnenko, A., Müller, J. A. Self-organization of nets of active neurons. Systems Analysis Modelling Simulation, 1995; 20: 93–106.

[44] Madala, H. R. Inductive Learning Algorithms for Complex Systems Modeling. 1st ed. Boca Raton (FL): CRC Press; 1993. doi:10.1201/9781351073493.

[45] Lemke, F., Müller, J. A. Self-Organising Data Mining. Systems Analysis Modelling Simulation, 2003; 43: 231–240. doi:10.1080/0232929031000136135.

[46] Taheri, A., Makarian, E., Manaman, N. S., Ju, H., Kim, T.-H., Geem, Z. W., RahimiZadeh, K. A Fully-Self-Adaptive Harmony Search GMDH-Type Neural Network Algorithm to Estimate Shear-Wave Velocity in Porous Media. Applied Sciences, 2022; 12: 6339. doi:10.3390/app12136339.

[47] Katoch, S., Chauhan, S. S., Kumar, V. A review on genetic algorithm: past, present, and future. Multimedia Tools and Applications, 2021; 80: 8091–8126. doi:10.1007/s11042-020-10139-6.

[48] Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning. 2nd ed. New York (NY): Springer; 2009. doi:10.1007/978-0-387-84858-7.

[49] Arlot, S., Celisse, A. A survey of cross-validation procedures for model selection. Statistics Surveys, 2010; 4: 40–79. doi:10.1214/09-SS054.

[50] Chai, T., Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)?– Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 2014; 7: 1247–1250. doi:10.5194/gmd-7-1247-2014.

[51] Botchkarev, A. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology. Interdisciplinary Journal of Information, Knowledge, and Management, 2019; 14: 45–76. doi:10.28945/4184.

[52] Asteris, P. G., Rizal, F. I., Koopialipoor, M., Roussis, P. C., Ferentinou, M., Armaghani, D. J., Gordan, B. Slope Stability Classification under Seismic Conditions Using Several Tree-Based Intelligent Techniques. Applied Sciences, 2022; 12: 1753. doi:10.3390/app12031753.

[53] Onyelowe, K. C., Mojtahedi, F. F., Ebid, A. M., Rezaei, A., Osinubi, K. J., Eberemu, A. O., Salahudeen, B., Gadzama, E. W., Rezazadeh, D., Jahangir, H., Yohanna, P., Onyia, M. E., Jalal, F. E., Iqbal, M., Ikpa, C., Obianyo, I. I., Rehman, Z. U. Selected AI optimization techniques and applications in geotechnical engineering. Cogent Engineering, 2023; 10: 2153419. doi:10.1080/23311916.2022.2153419.

[54] Jolfaei, S., Lakirouhani, A. Sensitivity analysis of effective parameters in borehole failure, using neural network. Advances in Civil Engineering, 2022; 2022: 4958004. doi:10.1155/2022/4958004.