Journal of Structural Engineering and Geotechnics,
11(2), 13- 26 , Summer & Autumn 2021

Modeling of Accumulated Energy Ratio (AER) for Estimating Liquefaction Potential using Artificial Neural Network (ANN) and Gene Expression Programming (GEP) (using Data from Tabriz)

Armin Sahebkaram Alamdari [a], Rouzbeh Dabiri [b,]*, Rasoul Jani [b],
Fariba Behrouz Sarand [b]

[a] *Department of Civil Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran*

[b] *Department of Civil Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran*

Received 2 October 2021, Accepted 20 January 2022

Abstract

Presenting a model specific to the city of Tabriz to estimate the liquefaction potential due to the region's seismicity and the high groundwater level can be effective in dealing with and predicting solutions to deal with this phenomenon. In recent years, the accumulation energy ratio (AER) as a parameter for estimating the liquefaction potential in the energy-based method proposed by Kokusho (2013) has been considered by many researchers. In this research, using perceptron multilayer (MLP) and radial base function (RBF) methods in artificial neural network (ANN) and genetic expression programming (GEP), the accumulation energy ratio using seismic and geotechnical data is modeled for the city of Tabriz. These modeling's performed by all three methods are well consistent with the outputs. Still, the modeling performed using the Perceptron Multilayer (MLP) method is very compatible with the outputs and can estimate the results with an acceptable percentage. The relationship presented by genetic expression programming (GEP), which is trained with local data, can also yield satisfactory results from estimating the rate of accumulated energy in the study area and provided an independent and accessible relationship trained. With data specific to the study area, there is another advantage.

Keywords: Liquefaction potential, Accumulated energy ratio, Artificial neural network, Genetic expression programming, Tabriz city.

1. Introduction

Soil liquefaction during an earthquake is one of the hazards that has always caused severe damage. For this reason, this issue has always been considered by many researchers. They have sought to understand better, study the effective parameters, and provide methods to estimate its potential in different areas to offer a solution to prevent this natural phenomenon. Various methods have been proposed to estimate the liquefaction potential, one of the most important of which is the energy-based method. These methods have always been considered due to the high importance of seismic parameters that lead to liquefaction. In recent years the results of some researchers based on seismic parameters have been presented. One of these methods is estimating liquefaction potential using the accumulation energy ratio presented by Kokusho (2013). Compared with other methods based on energy, stress, and strain, this method provides acceptable results in estimating the liquefaction potential. The key to the proposed method is to compare the upward wave energy with the energy capacity for liquefaction in each layer. Studies have shown that estimating the liquefaction potential by the energy-based method provides similar results to the stress-based method for several ground motions of recent earthquakes but is easier to apply without considering the relevant parameters [1, 2, and 3].

Others have suggested various methods for prediction in engineering applications, one of which is the use of artificial intelligence-based methods [4]. In their previous studies, the authors have attempted to zoning Tabriz's seismic and liquefaction potential (Figure 1 and 2) using data collected from the range.

*Corresponding Author: Email Address: rouzbeh_dabiri@iaut.ac.ir

In this research, based on the data and results of previous studies and using artificial neural network and genetic expression planning, a model has been tried to estimate the accumulation energy ratio caused by earthquakes using data collected from the desired area should be specific to the city of Tabriz and the parameters affecting energy-based methods should be presented [5, 6]. The data used in this research has been collected from reputable consulting companies active in this field, such as Tabriz Metro, the provincial soil laboratory, and consulting companies. In some places where we lacked data density in those areas or the need, the data used were validated. The data were obtained by drilling and field and laboratory tests.

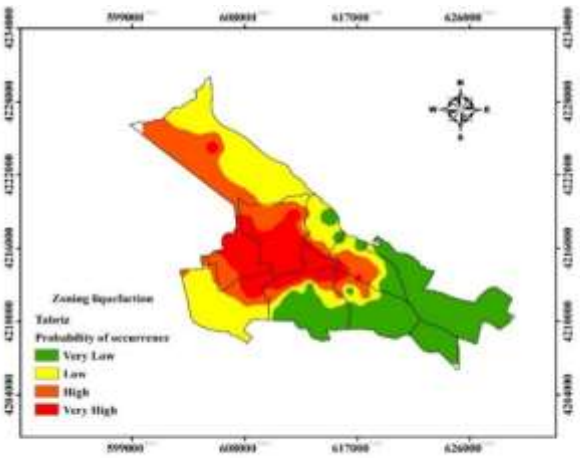

Fig.1. Zoning of the probability of liquefaction in Tabriz using the relation of accumulation energy ratio [6].

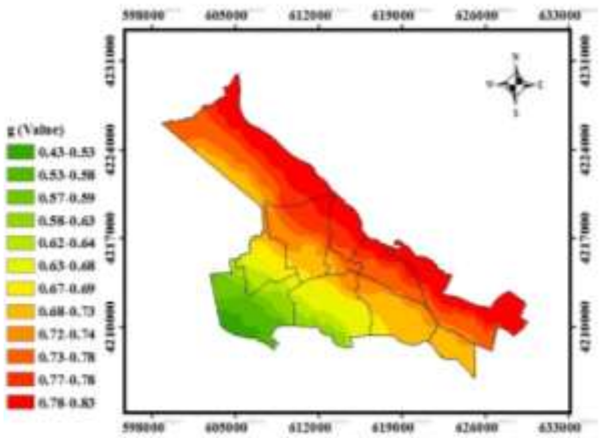

Fig 2. Seismic zonation map of Tabriz after amplification [6].

Among the studies performed to estimate the liquefaction potential using artificial intelligence methods, we can reference the research of Zhang et al. (2016), who evaluated soil liquefaction based on the concept of energy capacity and multivariate adaptive regression lines. In this study, several methods in liquefaction assessment have been developed, including those related to energy capacity for basic soil parameters such as relative density, initial adequate limit pressure, fine grain size, and soil texture characteristics [7]. Kumar Modali and Das (2016) have studied the uncertainty of the SPT-based method model for evaluating the seismic liquefaction potential of soil using multi-gene genetic programming [8].

Goharzai et al. (2017), in this study, analyzed the onset of SPT-based soil liquefaction using gene expression programming and the possible Bayesian method. They also point out that probabilistic models provide a simple but efficient decision-making tool in engineering design for quantitative evaluation of the onset threshold of liquefaction [9]. Pirhadi et al. (2018) have presented a new equation for evaluating the onset of liquefaction using artificial neural network and response surface methodology [10]. Hu and Liu (2019) have studied the evaluation of seismic liquefaction potential based on the Bayesian lattice constructed from the domain of knowledge and history data [11]. Rahbarzadeh and Azadi (2019) have studied the improvement of soil liquefaction prediction using hybrid optimization algorithms and fuzzy support vector devices. According to the experimental results, the proposed algorithm can improve classification accuracy—this practical method for predicting soil liquefaction using optimal classification parameters without a user interface [12]. Zhang and Wang (2020) present a group method to improve the prediction of earthquake-induced soil liquefaction. In this study, several data sets and soft computing methods have been used to estimate the liquefaction potential and data clustering [13].

Due to the importance and effectiveness of energy parameters in soil liquefaction, in this study, an attempt has been made to use the database obtained from previous studies by Sahebkaram Alamdari et al. (2020 and 2021) in the city of Tabriz and using the data collected in this range, to model the estimation of the accumulation energy ratio specific to the study area, using artificial neural network (ANN) and genetic expression programming (GEP).

2. Research Method

In this study, in order to estimate the accumulation energy ratio (AER) leading to liquefaction, artificial neural networks (ANN) and genetic expression programming (GEP) have been used. The following is a brief description of the methods used.

2.1. Accumulated Energy Ratio

Kokusho [1] developed an EBM that allowed for potential liquefaction to be evaluated by comparing the strain energy for liquefaction in a sand layer and the incident seismic energy, regardless of the differences in seismic motions. Comparative studies of soil models have shown that an EBM intrinsically incorporates the effects of different input motions. In contrast, they need to be included by adding suitable coefficients in an SBM. The other notable difference is that the liquefaction potentials are higher at a smaller depth in an EBM, whereas the opposite is for an SBM with uniformly deposited sands [2]. In an EBM, it is necessary first to calculate the energy dissipated by soil liquefaction. The dispersed normalized energy for a unit measure of soil for liquefaction was calculated based on the soils' penetration test results. With the standard penetration number of the soil layer known, the number must be first normalized by the fine grain content and the effective overburden stress. For details on the steps to obtain normalized strain energy and upward energy, refer to Kokusho (2013) and Kokusho and Mimori (2015).

The normalized strain energy input corresponding to each soil volume unit for liquefaction is calculated as

$$\mathrm{W}/\sigma_c' = 5.4 \times 10^{1.25\times\log(\Delta \mathrm{W}/\sigma_c')} \tag{1}$$

The upward energy in a layer can be calculated from upward SH waves for a large shake period $t = 0 \sim t_1$ as

$$\mathrm{E_u} = \rho \mathrm{V_s} \int_0^{t_1} (\dot{\mathrm{u}})^2 \mathrm{dt} \tag{2}$$

Where $\dot{u}$ denotes the seismic wave particles velocity spreading upwards, ρ is the soil density and V_s represents the S wave velocity of the layer. The potential energy ratio WH/E_{uf} is obtained by comparing the liquefaction potentials of WH and the upward energy at the end of shaking E_{uf} in the corresponding layer. At the same time, the layer with a smaller energy ratio has a higher liquefaction potential. The energy ratios of all the layers were calculated separately and then summed up together. Liquefaction takes place sequentially in layers where $\sum(WH)_i/(E_{uf})_i < 1.0$.

2.2. Artificial Neural Network (ANN)

Artificial neural networks (ANNs) have advantages such as high speed in parallel computing, immunity against input harmonic ripple, noise cancellation, resistance to a parameter change, and generalizability [14]. A neural network consists of a large number of nerve cells or processing elements that are connected by synaptic weights. The structure of neural networks can be divided into two general categories: feed networks and return networks. Feeder networks have a simple structure and do not have the problem of stability of return networks. Feeder networks have different types, the most important of which are multilayer perceptron network (MLP) and network with radial basis functions (RBF) [15]. The main reason for choosing these two methods is the appropriate results obtained from the articles presented using these two methods in the study topic. However, in the present study, we evaluated different functions to obtain the best answer, and finally, we selected the results of MLP and RBF functions for presentation.

2.2.1. Perceptron Multilayer Neural Network

Perceptron multilayer neural networks (MLPs) are probably the most popular in practical applications among some types of artificial neural networks. Their advantage lies in both simplicity and the relatively small number of estimated parameters [16]. An MLP is a feed neural network consisting of several layers and neurons that link the input data set to the appropriate outputs. MLP consists of an input layer, one or more hidden layers, and an output layer, with each layer attached to the next layer [17]. Each neuron is a processing element that has an activation function. Neurons are interconnected processing nodes that form an ANN. The output of each neuron is the result of a set of weighted inputs

[18]. Figure 3 shows the structure of a three-layer MLP network. In this research, the Levenberg-Marquardt optimization algorithm performs network training.

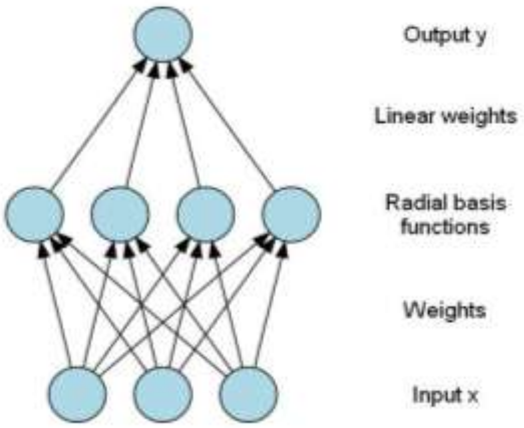

Fig.3. Structure of the radial basis function [27].

Training With Levenberg-Marquardt Algorithm

This method is a modified version of Newton's classical algorithm, used to find suitable solutions to problems requiring minimization. Similar to Newton's method, this method considers an approximation for the Hayzen matrix in weight change [19].

$$\mathbf{x_{k+1}} = \mathbf{x_k} - [\mathbf{j^T j} + \boldsymbol{\mu}\mathbf{I}]^{-1}\mathbf{j^T} \qquad (3)$$

In the above relation, x is the neural network weights, j is the Jacobin matrix of the network performance criterion that should be minimized, μ is the number that controls the training process, and e is the residual error vector. When μ is zero, the above equation is the same Newtonian method that uses the Hessian method. However, when μ is a large value, the equation becomes a gradient reduction relation with a minor interval. Newton's method has a high speed, and the results will be very close to the moment of error. This algorithm has been used in many studies due to the above characteristics. This algorithm has high efficiency and is very stable [20].

2.2.2. Radial Basis Function (RBF)

In mathematical modeling, RBF is an artificial neural network that uses basic radial functions as activation functions. The output of this network is a linear combination of radial basis functions for input and neuron parameters. These networks are used in the approximation, time series prediction, classification, and system control function. The RBF neural network is benefits from structural simplicity, higher approximate properties, and faster computational procedures [21].

RBF networks usually consist of the input layer, the hidden layer with a nonlinear RBF activity function, and the output layer [22]. The number of neurons in the hidden layer is equal to the dimension of the input data set, and the neurons in the hidden layer usually have Gaussian transfer functions whose output is proportional to the distance from the center of the neuron. The input can be a vector of a real number model. Furthermore, the output of this network is a scalar function of the input vector, which is calculated as follows [23]:

$$(x) = \sum_{i=1}^{N} a_i \rho(\|x - c_i\|) \qquad (4)$$

Where N is the number of hidden layer neurons, ci is the center vector of neuron i, and ai the weight of neuron i in the linear output neuron.

A two-step algorithm usually trains RBF networks. In the first step, the center vectors ci are selected for the RBF functions in the hidden layer. This step can be performed in several ways: centers can be randomly sampled from some set of examples, or K-means clustering can be used. Note that this step is unsupervised. According to the objective function, the second step is fitted with a linear model with wi coefficients for the hidden layer outputs [16].

2.3. Genetic Expression Programming

One of the most exciting topics in data analysis is discovering the secret relationship or relationships among the databases. For this purpose, first of all, we need an almost comprehensive and complete database. In the next stage, we need tools and techniques to discover and present the mentioned relationships accurately. As one of the most widely used tools in this field, Genetic Expression Programming can extract very diverse and complex relationships from databases and present them in the form of intuitive formulas.

Gene Expression Programming (GEP) is a generalized genetic algorithm (GA) developed based on Darwin's theory and invented by Ferreira in 1999. Genetic Expression Programming (GEP), like Genetic Algorithm (GA) and Genetic Programming (GP), is a genetic algorithm that uses a population of individuals and selects them accordingly, and genetically modifies them using one or more Applies several genetic agents. The main difference between these three algorithms is the nature of their individuals. In GA, people are linear strings with constant length (chromosomes), and in GP are nonlinear entities with different sizes and shapes (decomposition trees). In GEP, on the other hand, individuals are encoded as linear strings of constant length in (genome or chromosome) (similar to that used in genetic algorithms) and then in the form of nonlinear entities of different sizes and shapes. Showing a simple diagram or expressing a tree similar to decomposition trees in genetic programming are expressed [25, 26, 27]. The first step in the GEP algorithm is to generate the initial population of solutions. Generating the initial population can be done by a random process or using some information about the problem. GEP uses the popular wheel-roulette method to select individuals. Unlike GP and GA, GEP has several genetic operators for breeding individuals with modifications. Generation is an operation that aims to retain several suitable individuals from the present generation to the next generation. The goal of the mutation operator is to optimize intra-specific chromosomes randomly. This operator performs some flawless operations to prevent the creation of defective people in terms of rules [28]. The chromosomes are then represented as a tree expression, evaluated according to a proper function. In GEP, each gene is encoded as a tree expression. In multi-gene chromosomes, all tree expressions are linked together from their root nodes using the link function. Each gene has a coding region called the ORF (Open Interpretation Framework), which after decoding, is expressed as ET and represents a candidate solution to the problem [30, 29]. One of the essential things in GEP is to determine the proper function, and its goal is to find a solution that works well for all fitting cases as much as a specific error. The fitting function is usually evaluated by processing several objective problems, also called fits. If a satisfactory quality is found or some generations are reached, evolution stops, and the best solution ever found is reported. On the other hand, if the conditions are not stopped, the best solution of the present generation is kept (meaning elite selection), and the rest of the solutions are left to the selection process. Based on the selection, the best people have a better chance of having children. The whole process is repeated for several generations, and as the generation progresses, the population's quality is expected to improve on average [30]. Figure 4 shows schematically the main stages of gene expression programming.

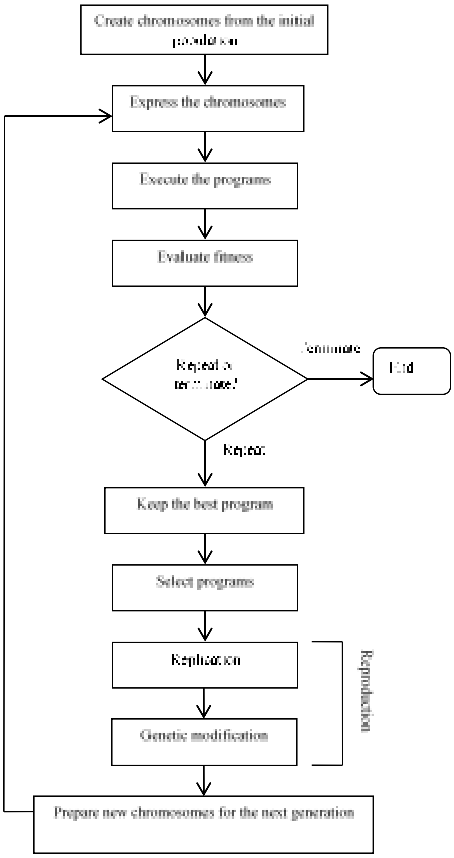

Fig.4. A gene expression programming flow chart [31].

2.4. Data Validation

One of the first validations used when collecting data was that the amount of data for different regions was a good overlap. In this way, when the number of data increases at a certain point, validation for the

values of geotechnical parameters that are not compatible with other data was removed. In order to more accurately validate the data used in the assessment of liquefaction potential, some Cetin screening criteria have been used [32]. Screening criteria by which Cetin validates the data it uses:
(1) Soil profiles that do not have complete information and are not adequately available; (2) Permanent constraints on critical soil layers that cannot be the judge if fine particles are present; b) higher plasticity and, therefore, the use of these relationships is not appropriate, (3) the sampling sites were not open ground, and (4) in the SPT speculation the impact drilling method was used.

3. Estimating Accumulated Energy Ratio

With 439 Downhole samples extracted and the results of estimating the accumulated energy ratio in the city of Tabriz, artificial intelligence has been used for modeling. From the geotechnical characteristics of the soil and seismicity of the area, the parameters of soil shear wave velocity (V_s), effective enclosed stress (σ_c'), fine-grained percentage (FC), SPT number (N), and maximum seismic velocity (V_{max}) as modeling input to the purpose is to estimate the energy ratio selected. An example of the data used is provided in Table 1. Soil geotechnical parameters obtained from the field (Standard penetration test and Downhole shear wave velocity test) and laboratory experiments. The collected data are the results of studies conducted for different lines of Tabriz metro and consulting companies active and approved in the field of geotechnical studies. In some places where the data density was lower or there was a need to compare the results to verify the data, we performed machine drilling and performed laboratory and field experiments. The maximum earthquake velocity parameter is also used to convert the acceleration maps obtained from the seismic zonation of the study area from previous studies by Sahebkaram Alamdari et al. (2020), which were also used in the potential liquefaction zoning mentioned in Figure 1., was extracted [5 and 6]. In the study of 98 cases using the experimental method of estimating the liquefaction potential based on Kokusho's (2013) energy, the results of which are presented in the form of zoning in Figure (1), have shown the liquefaction potential. Their accumulated energy ratio is less than one; in 341 cases, the energy ratio was higher than one, indicating the absence of the liquefaction potential. Therefore, the input parameters of the modeling are the same as those used in the Kokusho method. The data used for SPT tests and shear wave velocity at the site and soil identification tests were obtained in the laboratory. The parameter related to the earthquake speed, as mentioned in the text of the article, has been obtained from the seismic zooming for the city of Tabriz, which was done in a previous study by researchers.

Table 1: Sample of data used for AER modeling

Number of Data	N	V_{max}	V_s	FC	σ_c'	AER
1	22	110	300	15	46.6	0/78
2	24	110	310	25	72.6	1.96
3	21	110	345	28	98.6	2.89
4	19	110	370	23	124.6	3.46
5	27	110	410	33	15.6	4.75
6	7	110	280	15	37	0.12
7	9	110	290	19	60.3	0.33
8	9	110	310	18	83.6	0.77
9	15	110	335	22	107	1.25
10	13	110	365	25	31.3	1.63
11	17	110	400	17	153.6	2.04
12	42	110	375	11	33.3	2.00
13	36	110	405	12	56.6	3.53
14	45	110	435	10	80	5.43
15	41	110	470	9	103.3	6.88
16	26	225	325	8	48.3	0.64
17	28	225	355	5	73.6	1.32
18	22	225	390	9	99	1.71
19	25	225	430	11	124.3	2.17
20	29	225	480	12	149.6	2.74
21	25	225	345	18	68.6	0.75
22	27	225	385	25	93.3	1.72
23	31	225	405	32	118	3.06
24	29	225	440	38	142.6	4.32

Table 2 shows the average size of the parameters used in the modeling and the method to normalize them. All data must be normalized in order for all variables to be proportional to each other, as well as to improve program execution speed. The values of our variables are normalized between zero and one.

Table 2
Variables used in model development.

input					
Parameters	V_S	V_{max}	σ_c'	FC	N
Mean	326.71	130.64	87.97	7.5	16.94
How to Normalize	$\frac{V_S - 139}{532}$	$\frac{V_{max} - 40}{320}$	$\frac{\sigma_c' - 1.6}{460}$	$\frac{FC}{99}$	$\frac{N - 2.8}{63.7}$
Parameters' ratio	139-671	40-360	16-462	0-49	2.8-66.5

Statistical indicators such as coefficient of determination (R^2), root mean square error (RMSE), and mean square error (MSE), whose relationships are presented below, have been used to validate the models. In the following equations, $\hat{y}$ is the predicted value, and $\hat{y}$ is the average output value.

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \bar{y})^2} \tag{5}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \tag{6}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2 \tag{7}$$

3.1. Simulation of Accumulated Energy Ratio Using Artificial Neural Network (ANN)

Simulation for this purpose, we coded two methods of Perceptron Multilayer Function (MLP) and Radial Base Function (RBF) in MATLAB.

Perceptron Multilayer Function (MLP)

In the multilayer perceptron function, among the various methods of error backpropagation training. We selected the Levenberg-Marquardt algorithm, a function of the square root derivative for use in the present study, for faster convergence in training medium-sized networks. The back propagation algorithm changes the network weights and bias values so that the performance function decreases more rapidly. Used 70% of the data for training, 25% for testing, and used 5% of the data for validation. Used the random division command to select the data, to be included in this category. As activity functions in hidden layers, the hyperbolic tangent and linear function provide the best answer among other functions. To train and stop training for the neural network, Levenberg-Marquardt selected the three-layer network, the first layer with five neurons, the second layer with three neurons, and the third layer with one neuron. With a fixed number of inputs as the best answer by examining statistical indicators. Figure 5 shows the simulation results of estimating the accumulation energy rate by the multilayer function of perceptron (MLP) in the artificial neural network. As can be seen from the values of statistical indicators, MLP has been able to provide a perfect simulation with a deficient error between our input and output data. In addition, the values obtained for the simulated accumulation energy rate matched very well with the calculated values.

Radial Basis Function (RBF)

RBF networks require more neurons than standard feedforward networks with backpropagation algorithms, but most of these networks can train shorter than the time required for feed networks. These networks perform better when multiple input vectors are available. RBF networks are a type of feedforward neural network. RBF network is a single layer. 70% of the data was used for training and 30% for testing, and the method of selecting them was still a random division. The rate of increased radius 0.1 and maximum of 30 neurons provided the best answer among the various modeling's. Figure 6 shows the modeling results of estimating the accumulation energy ratio by the radial basis function (RBF) in the artificial neural network. As can be seen from the values of statistical indicators, RBF has provided a good simulation with a low error between our input and output data. Also, the values obtained for the simulated accumulation energy rate are mainly consistent with the calculated values.

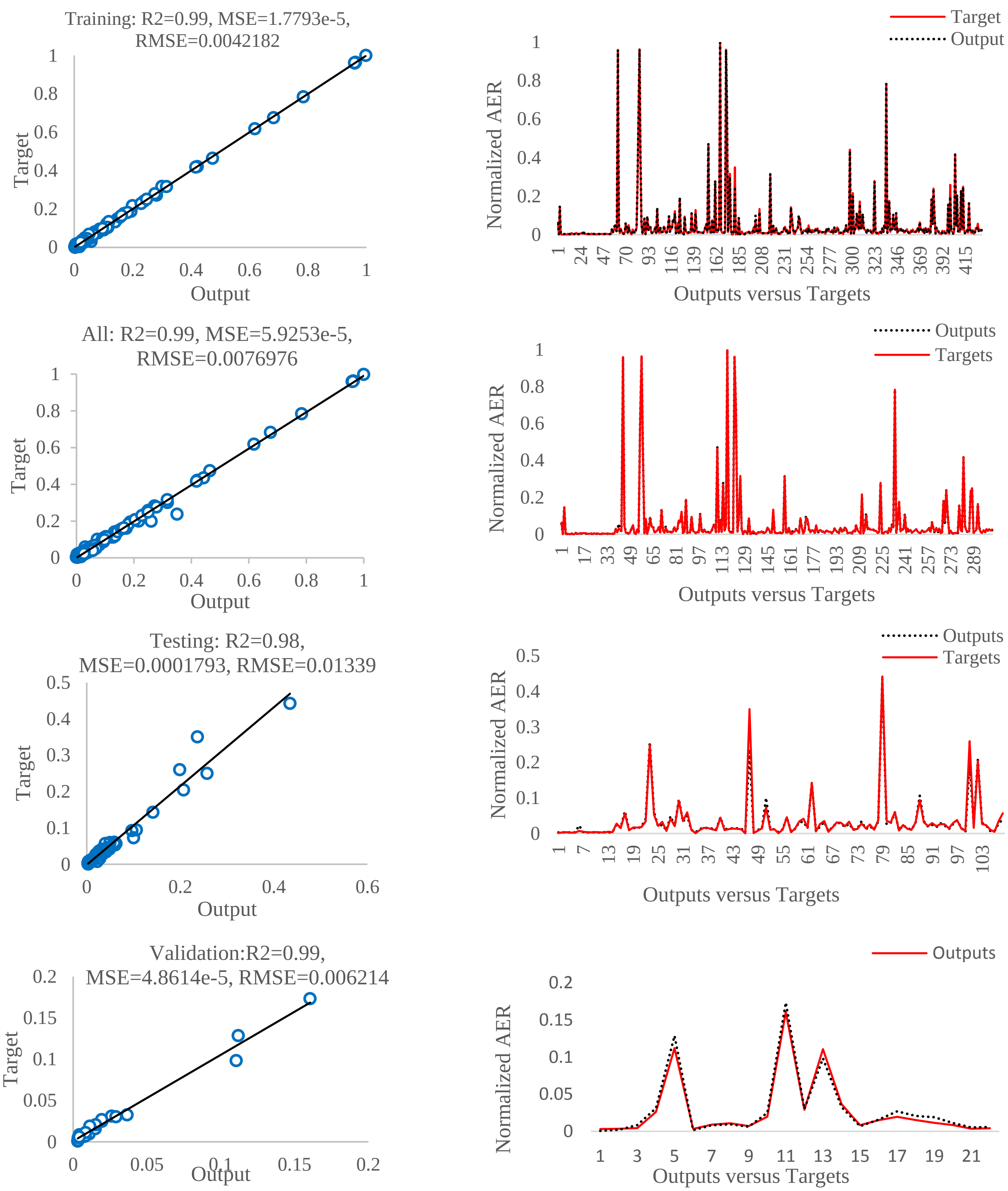

Fig.5. Results for estimating the accumulation energy ratio by the multilayer perceptron (MLP) function of the artificial neural network (ANN).

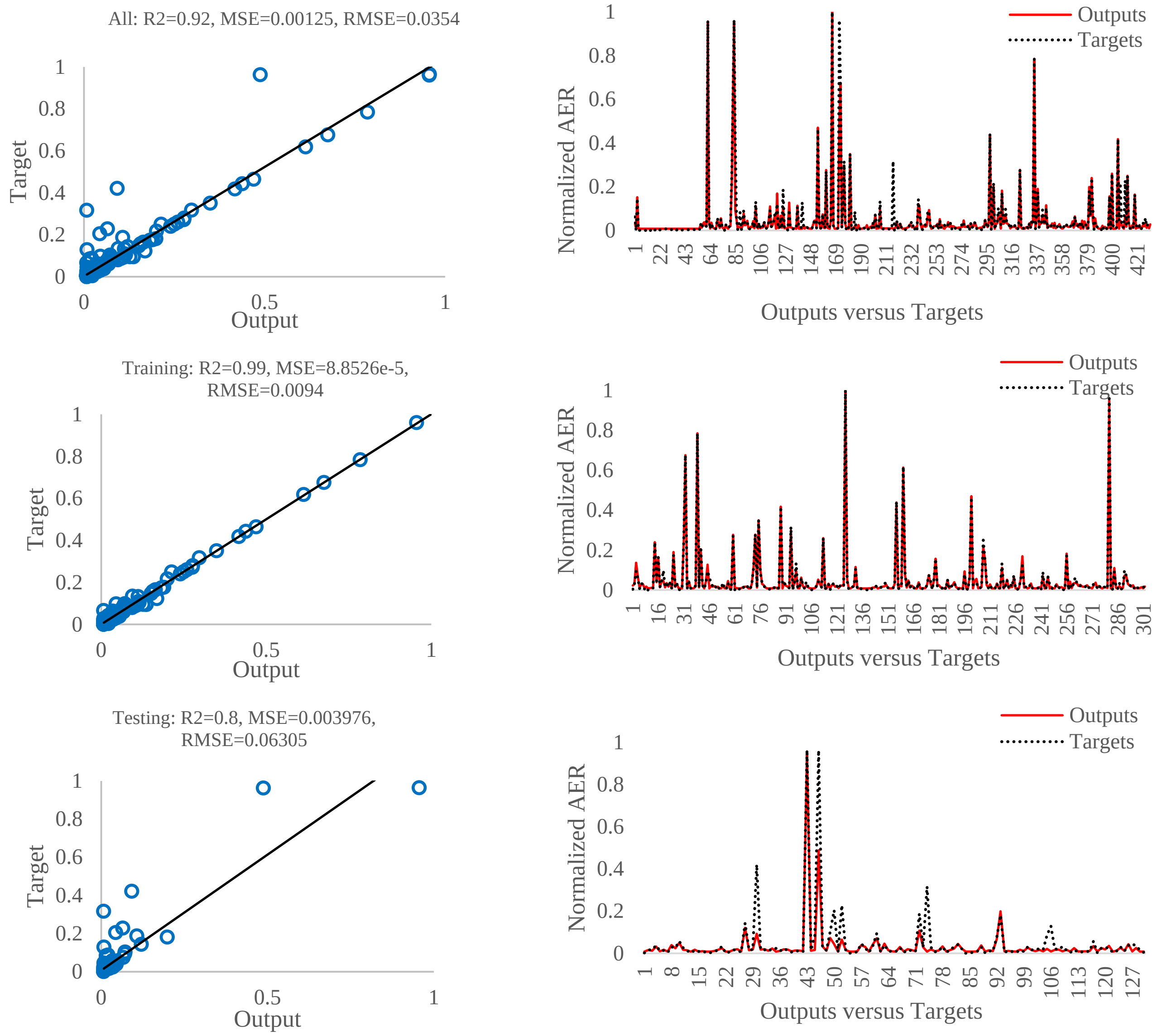

Fig.6. Results of estimating the accumulation energy ratio by the radial basis function (RBF) of the artificial neural network (ANN).

3.2. Determining the Ratio of Storage Energy using Genetic Expression Programming (GEP)

In order to use the GEP method for this research, GeneXProTools 5.0 software was used. A performance analysis was done to set the model parameters. In GEP, values of setting parameters have a significant influence on the fitness of the output model. These include the number of chromosomes, number of genes, gene head size, and the rate of genetic operators. This approach involved using different settings and conducting runs in steps. During each step, runs were carried out, and the values of one of the parameters mentioned above were varied, whereas the values of the other parameters were kept constant. At the end of each run, the statistical indicators for both training and testing sets were recorded to identify the values that give the best results [33]. The parameters used in the development of the best modeling obtained are presented in Table 3. The most crucial issue in GEP modeling is the number of genes selected according to the accuracy and application of the model. The most acute effects of a high or low number of genes are reducing the accuracy of the equation (reducing the number of genes) or lengthening it (increasing the number of genes) [34]. According to the above conditions and creating a workable equation with high accuracy, eight genes were considered using trial and error. In addition, a crucial function must be used to maintain the relationship between the written mapping performances of each tree diagram for models with more than one gene. In addition, several equations were created using different connection functions such as "+", "×", "÷", "-", "x^2" and their combinations, and finally, the best equation was selected .

Table 3
GEP software settings

Parameters	Achieved functions, values and rates
Linking function	Addition (+)
Number of chromosomes	40
Number of genes	8
Gene head size	3
Recombination rate	0.02
Mutation rate	0.01

GEP divided data randomly into two parts for training and experimentation. Of the 439 datasets, programming used 307 for training. Then, programming used the remaining 132 datasets to evaluate the obtained equation. GEP equation presented based on five input parameters.

Table 2 presents the parameters used in the above equations. They are comparing the predicted and calculated values, as shown in Figure 7, a good agreement between the values predicted by this equation and the calculated values. As can be seen from the statistical index values, GEP has provided a perfect simulation with a deficient error between our input and output data. In addition, the values obtained for the simulated accumulation energy ratio are very well match with the calculated values. The proposed model can establish our complex relationships between the rate of accumulated energy and their main influential factors. The proposed model is the best model obtained by genetic programming to estimate the rate of accumulated energy. So that the modeling results for training and test data in all statistical indicators, as shown in Table 4, have been acceptable and successful.

$$\mathrm{AER} = \left((\sigma'_c - V_s)^2 \times (1.76V_{max} - 1.584)\right)^2 + (FC^4 \times V_{max} \times N) - \frac{N}{4.75} + \frac{N}{\left(\frac{FC \times V_s}{V_{max} + \sigma'_c} + FC - 2.34\right)^2} \tag{8}$$

Table 4
Results of statistical indicators of relationships presented

Statistical specifications				
AER model		**R^2**	**RMSE**	**MSE**
ANN-MLP	All	0.99	0.0076	5.92e-5
	Training	0.99	0.0042	1.772e-5
	Testing	0.96	0.013	0.00017
ANN-RBF	All	0.92	0.0354	0.00125
	Training	0.99	0.0094	8.85e-5
	Testing	0.8	0.063	0.0039
GEP	All	0.93	0.028	8.1e-4
	Training	0.95	0.03	9e-4
	Testing	0.89	0.025	6.3e-4

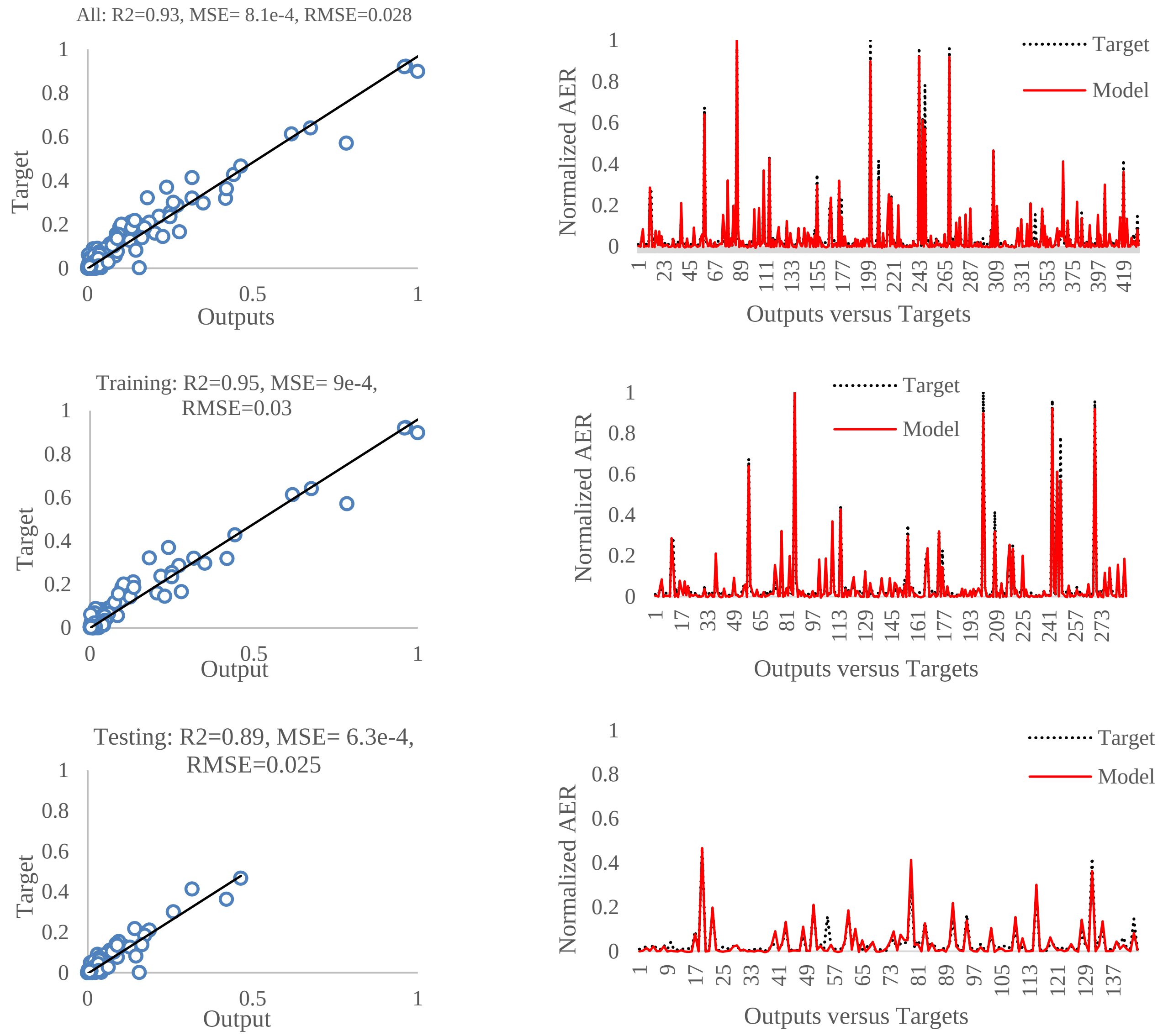

Fig.7. Results of estimating the accumulation energy ratio by genetic expression programming (GEP).

4. Sensitivity Analysis

Sensitivity analysis was performed to evaluate the model response to changes in input parameters. For this purpose, all input parameters, soil shear wave velocity (Vs), effective enclosed stress ($\boldsymbol{\sigma'_c}$), fine-grained percentage (FC), SPT number (N), and maximum earthquake velocity (V_{max}) were considered. The value of one of the input parameters has been increased by 20% to evaluate the parameters' effect on the accumulation energy ratio. In contrast, the amplitude of other input parameters has been kept constant, and this increase in the accumulation energy ratio has been calculated. This calculation is done for all input parameters, and the results of calculating the change of output parameter are shown in Table 6. In this table, negative changes mean a decrease, and positive means an increase in the impact on the accumulation energy rate parameter. It is shown that with increasing Vmax, a decrease in the accumulation energy rate parameter occurs. In addition, Table 5 shows that the accumulation energy rate parameter increases with increasing Vs, FC, N, and$\boldsymbol{\sigma'_c}$. This table shows that the maximum V_{max} earthquake velocity has the most significant impact on the rate of accumulated energy.

Table 5
The accumulation energy rate parameter change corresponds to a 20% increase in the average value of the input parameters

Impact of parameter changes on AER	V_S	V_{max}	σ'_c	FC	N
	0.046352	-0.08023	0.030999	0.004818	0.00393

5. Assessment of the Results

In this research, we used the Multilayer simulation of perceptron (MLP) and radial base function (RBF) in artificial neural networks and modeling of gene expression programming (GEP) to estimate the accumulation energy ratio, which is an energy-based parameter for estimating liquefaction potential. The input parameters used were soil shear wave velocity (Vs), effective enclosed stress ($\boldsymbol{\sigma'_c}$), fine-grained percentage (FC), SPT number (N), and maximum earthquake velocity (Vmax). From the data of laboratory experiments that have led to log boreholes of geotechnical characteristics of the studied soil area and field tests of downhole shear wave and standard penetration test after validation, 439 data have been selected for modeling. In the study of 98 of these cases, using the experimental method of estimating the liquefaction potential based on Kokusho energy (2013), the results of which are presented in the form of zoning in Figure (1), have shown the liquefaction potential. Their energy ratio is less than 1; In 341 cases, the energy ratio was higher than 1, which indicates the non-liquefaction potential.

Table 5 and figure 8 present the aggregation of the results obtained from modeling accumulation energy rates using artificial neural networks and gene expression programming. As can be seen from the comparison of statistical indices, the results of modeling using the multilayer perceptron (MLP) in the artificial neural network were better than modeling with the radial base function (RBF) and gene expression programming. In MLP, as activity functions in hidden layers, the hyperbolic tangent and linear function provide the best answer among other functions. A three-layer network has been selected as the best answer by examining statistical indicators to train and stop training for the Levenberg-Marquardt neural network. The first layer comprises five neurons, the second layer comprises three neurons, and the third layer is composed of one neuron with a fixed number of inputs. In the RBF, by examining statistical indicators, the rate of increased radius 0.1 and maximum of 30 neurons provided the best answer among the various modeling's. However, the results of modeling using genetic expression programming (GEP) are significant due to the ability to provide a specific formula for estimating the accumulated energy ratio. The model presented by GEP uses 40 chromosomes, eight genes, three gene head sizes, and a 0.01 mutation rate. The statistical characteristics of this modeling are based on the use of local data, and specific to the study area is acceptable. The sensitivity analysis also showed that the maximum velocity of the earthquake is the most influential parameter in the accumulation energy ratio.

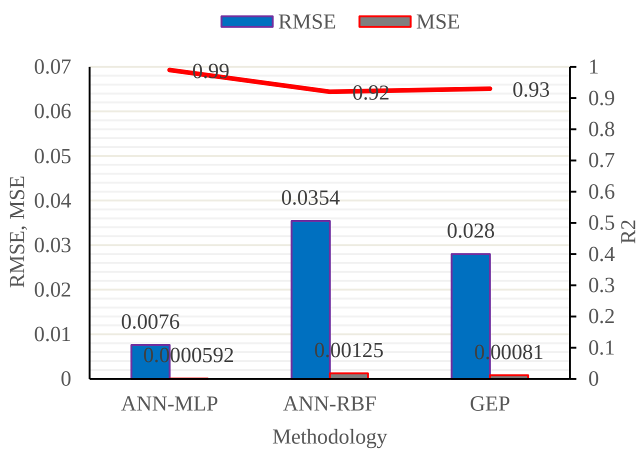

Fig.8. Comparison of R^2, RMSE and MSE statistical indices for accumulation energy ratios obtained from ANN-MLP, ANN-RBF and GEP modeling

6.Conclusions

The results obtained from neural network modeling were performed by the multilayer perceptron (MLP) function to determine the accumulated energy ratio, a criterion for estimating energy-based liquefaction potential; it perfectly matches the outputs and can estimate the results with an acceptable percentage. The relationship presented by genetic expression programming (GEP), which is trained with local data, can also yield acceptable results from estimating the accumulated energy ratio in the study area. Considering the independent and accessible relationship trained with data specific to the study area is another advantage. Due to the importance and high impact of earthquake stimulus parameters, it is crucial to pay more attention to energy-based methods to estimate the liquefaction potential, which is also one of the objectives of the present study. Proximity and sometimes construction has been done on the main fault north of Tabriz, can be very effective in estimating this natural factor that occurs due to earthquakes as accurately as possible. As can be seen from the sensitivity analysis of the parameters affecting the results of modeling the accumulation energy ratio, the parameters of earthquake speed and SPT number have the most significant impact on the outputs. These issues show more and more attention to the more accurate estimation of the probable earthquake strength and more accurate field testing of standard penetration to obtain the most accurate estimate of the liquefaction potential.

References

[1] Kokusho, T. “Liquefaction potential evaluations: energy-based method versus stress-based method”, Canadian Geotechnical Journal; 2013, 50: 1088–1099.

[2] Kokusho, T., Mimori, Y. “Liquefaction potential evaluations by energy-based method and stress-based method for various ground motions”, Soil Dynamic sand Earthquake Engineering; 2015, 75: 130–146.

[3] Kokusho, T., Kaneko, Y. “Energy evaluation for liquefaction-induced strain of loose sands by harmonic and irregular loading tests”, Soil Dynamics and Earthquake Engineering; 2018, 114: 362-377.

[4] Masulli, F., Studer, L. “Time series forecasting and neural networks”, In: Proc. Int. Joint Conf. on Neural Networks (IJCNN), Washington, DC. New York, 1999.

[5] Sahebkaram Alamdari, A., Dabiri, R., Jani, R., and Behrouz Sarand, F. “Seismic Zoning of Tabriz City Area by Stochastic Finite Fault Model with Acts of Site Impact”, Soils and Rocks; 2021, 44(1): 1-13.

[6] Sahebkaram Alamdari, A., Dabiri, R., Jani, R., and Behrouz Sarand, F. “Evaluation of Liquefaction Potential by Energy-based and Stress-based Methods and Gene Expressing Programming (Case study: Tabriz City)”, Indian Geotechnical Journal; 2021, (in review).

[7] Zhang, J., Juang, C. H., Martin, J. R., Huang, H. W. “Inter-region variability of Robertson and Wride method for liquefaction hazard analysis”, Engineering Geology; 2016, 203: 191–203.

[8] Muduli, P. K., Das, S. K. “Model uncertainty of SPT-based method for evaluation of seismic soil liquefaction potential using multi-gene genetic programming”, Soils and Foundations; 2015, 55(2): 258–275.

[9] Goharzay, M., Noorzad, A., Ardakani, A. M., Jalal, M. “A worldwide SPT-based soil liquefaction triggering analysis utilizing gene expression programming and Bayesian probabilistic method”, Journal of Rock Mechanics and Geotechnical Engineering, 2017, 9(4): 683-693

[10] Pirhadi, N., Tang, X., Yang, Q., Kang, F. “A New Equation to Evaluate Liquefaction Triggering Using the Response Surface Method and Parametric Sensitivity Analysis”, Sustainability; 2018, 11(1): 1-14.

[11] Hu, J., Liu, H. “Identification of ground motion intensity measure and its application for predicting soil liquefaction potential based on the Bayesian network method”, Engineering geology; 2019, 248: 34-49.

[12] Rahbarzare, A., Azadi, M. “Improving prediction of soil liquefaction using hybrid optimization algorithms and a fuzzy support vector machine”, Bulletin of Engineering Geology and the Environment; 2019, 78: 4977–4987.

[13] Zhang, J., Wang, Y. “An ensemble method to improve prediction of earthquake-induced soil liquefaction: a multi-dataset study”, Neural Computing and Applications; 2020, 33: 1533–1546.

[14] Batzel, T. D., Lee, K.Y. “A diagonally recurrent neural network approach to sensorless operation of the permanent magnet synchronous motor”, IEEE Power Engineering Society Summer Meeting; 2000, 4: 2441-2445.

[15] Bhushan, B., Singh, M., Hage, Y. “Identification and control using MLP, Elman, NARXSP and radial basis function networks: a comparative analysis”, Artificial Intellengce Review; 2012, 37: 133–156.

[16] Haykin, S. “Neural Networks, a Comprehensive Foundation”, Macmillan College Publishing Co., New York, USA, 1999.

[17] Hsu, K., Gupta, H. V. “Sorooshian S. Artificial neural network modeling of the rainfall-runoff process”, Water Resource Research; 1995, 31(10): 2517–2530.

[18] Ghritlahre, H. K., Prasad, R. K. “Application of ANN technique to predict the performance of solar collector systems – a review”, renew. Sustain Energy Revew; 2018, 84: 75–88.

[19] Wilamowski, B. M., Chen, Y., Malinowski, A. “Efficient algorithm for training neural networks with one hidden layer”, International Joint Conference on Neural Networks, Washington, DC, USA City, 1999, 1725–1728.

[20] Hagan, M. T., Menhaj, M. B. “Training feedforward networks with the Marquardt algorithm”, IEEE Transactions on Neural Networks; 1994, 5: 989–993.

[21] Zendehboudi, A., Tatar, A. “Utilization of the RBF network to model the nucleate pool boiling heat transfer properties of refrigerantoil mixtures with nanoparticles” J. Mol. Liq.; 2017, 247: 304–312.

[22] Howlett, R. J., Jain, L. D. “Radial basis function networks: 1. Recent developments in theory and applications”, Heidelberg (Germany): Physica-Verlag, A Springer-Verlag Company, (2001).

[23] Howlett, R. J., Jain, L. D. “Radial basis function networks: 2. Recent developments in theory and applications”, Heidelberg (Germany): Physica-Verlag, A Springer-Verlag Company, (2001).

[24] Bortman, M., Aladjem, M. “A Growing and Pruning Method for Radial Basis Function Networks”, IEEE Transactions on Neural Networks; 2009, 20(6): 1039-1045.

[25] Koza J. R. “Genetic Programming: On the Programming of Computers by Means of Natural Selection” MIT Press, Cambridge, MA. (1992).

[26] Ferreira C. “Gene expression programming: a new adaptive algorithm for solving problems”, Complex Systems; 2001, 13(2): 87–129.

[27] Ferreira, C. “Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence”, Springer-Verlag, Germany, 2th ed., 2006, 478.

[28] Lopez, H. S., Weinert, W. R. “An Enhanced Gene Expression Programming Approach for Symbolic Regression Problems”, International Journal of Applied Mathematics in Computer Science; 2004, 14: 375-384.

[29] Ferreira, C. “A Quick Introduction to Gene Expression Programming. From GEP Tutorials: A Gepsoft Web Resource”, http://www.gene-expression-programming, 2010.

[30] Sahebkaram Alamdari, A., Najafi. A. “The Study of the Liquefaction Probability and Estimation of the Relative Importance of Effective Parameters Using Fuzzy Clustering and Genetic Programming”, Journal of Civil and Environmental Engineering; 2018, 47(4): 37-46.

[31] Cetin, K. O., Seed, R. B., Kayen, R. E., Moss, R. E. S., Bilge, H. T., Ilgac, M., Chowdhury, K. “Summary of SPT based field case history data of Cetin 2016 Database”, Report No: METU / GTENG 08/16-01. (2016).

[32] Ferreira, C. “Gene expression programming and the evolution of computer programs. In: Castro, L.N.D., Zuben, F.J.V. (Eds.)”, Recent Developments in Biologically Inspired Computing; 2004, 82–103.

[33] Alkroosh, I., Nikraz, H. “Predicting axial capacity of driven piles in cohesive soils using intelligent computing”, Engineering Applying Artificial Intelligence; 2011, 25(3): 618-627.

[34] Ahangari, K., Moeinossadat, S. R., Behnia D. “Estimation of tunnelling-induced settlement by modern intelligent methods”, Soils and Foundations; 2015, 55(4): 737–748.