



SPT-based liquefaction assessment with a novel ensemble model based on GMDH-type neural network

Talas Fikret Kurnaz¹ · Yilmaz Kaya²

Received: 3 December 2018 / Accepted: 10 July 2019 / Published online: 24 July 2019
© Saudi Society for Geosciences 2019

Abstract

Liquefaction is one of the most complex problems in geotechnical earthquake engineering. This paper proposes a novel ensemble group method of data handling (EGMDH) model based on classification for the prediction of liquefaction potential of soils. The database used in this study consists of 451 standard penetration test (SPT)-based case records from two major earthquakes. The input parameters are selected as SPT blow numbers, percent finest content less than 75 μm , depth of groundwater table, total and effective overburden stresses, maximum peak ground acceleration, and magnitude of earthquake for the prediction models. The proposed EGMDH model results were also compared with other classifier models, particularly the results of the GMDH model. The results of this study indicated that the proposed EGMDH model has achieved more successful results on predicting the liquefaction potential of soils compared with the other classifier models by improving the prediction performance of GMDH model.

Keywords Liquefaction · Prediction · Group method of data handling · Ensemble model

Introduction

Liquefaction is a rapid loss of shear strength in non-cohesive soils subjected to dynamic loading effects such as earthquakes. Sometimes, the shear strength drops to almost zero or significantly reduced. In both cases, liquefaction leads to many soil-based problems (Coduto 2003). The most characteristic feature of all liquefaction events is the excessive pore water pressure under undrained loading conditions. It is a well-known issue that non-cohesive dry soils tend to tighten more under static and repetitive loads. However, when the cohesionless soils are saturated, the loading in undrained conditions develops rapidly, and the tendency of the soil to tighten more causes excessive pore water pressure and decreases the effective stress (Kramer 1996).

Liquefaction is one of the most important and complex issues in geotechnical earthquake engineering. The destructive effects of liquefaction in Alaska ($M_w = 9.2$) and Niigata ($M_s = 7.5$) earthquakes in 1964 attracted the attention of researchers to this phenomenon. The most striking of liquefaction-induced damages in both earthquakes is slope failures, failures in bridge and building foundations, floating embedded structures, etc. After the following earthquakes in 1971 San Fernando, 1976 Tangshan, 1985 Mexico city, 1989 Loma Prieta, 1994 Kobe, and 1999 Golcuk (Turkey), many researchers have turned to observing the conditions that affect the liquefaction phenomenon. For this reason, determination of the factors causing liquefaction, the liquefaction potential in vulnerable areas, and the prediction of possible damages are among the most important research topics in geotechnical earthquake engineering.

The liquefaction potential depends on the geotechnical properties of the grounds, topography, seismicity, groundwater level, and geological history (Youd and Perkins 1978). Various empirical methods based on experimental and probabilistic calculations have been developed to determine the liquefaction potential (Kramer and Mayfield 2007). Liquefaction potential can be determined by laboratory tests (dynamic three-axis, dynamic cutting, shaking tests) and in situ tests (standard penetration test (SPT), cone penetration test (CPT), seismic experiments) (Kramer 1996 and Ishihara 1996; Liu

Editorial handling: David Giles

✉ Talas Fikret Kurnaz
fkurnaz@mersin.edu.tr

¹ Vocational School of Technical Sciences, Department of Transportation Services, Mersin University, Mersin, Turkey

² Department of Computer Engineering, Faculty of Engineering and Architecture, Siirt University, Siirt, Turkey

and Qiao 1984; Elgamal et al. 1989; Lambe 1981; Husmand et al. 1988; Seed and Idriss 1971; Tokimatsu and Yoshimi 1983; Iwasaki et al. 1981; Suzuki et al. 1997; Robertson and Wride 1998; Stokoe et al. 1988; Andrus and Stokoe 2000). However, due to the laboratory tests being time-consuming and expensive, the methods by which SPT and CPT data are used are more preferred. The methods based on SPT have often been more preferred by geotechnical engineers for many years in the assessment of liquefaction. In these methods, the safety of the ground against the liquefaction during the earthquake is calculated by comparing the rate of cyclic resistance (CRR) to the rate of cyclic stress (CSR) (Seed and Idriss 1971; Youd et al. 2001; Cetin et al. 2004; Idriss and Boulanger 2006; Idriss and Boulanger 2010; Boulanger and Idriss 2012).

Recently, soft computing methods especially the artificial neural networks (ANNs) have become popular in practical solutions of the geotechnical engineering problems such as bearing capacity of shallow and pile foundations, slope stability, settlement behavior, and compressibility parameters of soils (Nejad et al. 2009; Lee and Lee 1996; Kiefa 1998; Sakellariou and Ferentinou 2005; Wang et al. 2005; Kuo et al. 2009; Abdalla et al. 2015; Chenari et al. 2015; Kalinli et al. 2011; Sulewska 2011; Chik et al. 2014). Also, the liquefaction potential of soils has been tried to predict by using different artificial intelligence applications in the last 20 years (Goh 1994, 1996, 2002; Juang and Chen 1999; Rahman and Wang 2002; Baziar and Nilipour 2003; Kim and Kim 2006; Hanna et al. 2007; Chern et al. 2008; Ramakrishnan et al. 2008; Mughieda et al. 2009; Samui and Sitharam 2011; Karthikeyan et al. 2013; Muduli and Das 2015a; Muduli and Das 2015b; Erzin and Ecemis 2015; Xue and Xiao 2016; Xue and Liu 2017; Goharzaya et al. 2017; Hoang and Bui 2018).

Goh (1994) suggested ANN models to predict the liquefaction potential of soils based on actual field records using SPT data. Rahman and Wang (2002) developed fuzzy neural network models for the evaluation of liquefaction potential with SPT-based large databases of liquefaction case histories. Hanna et al. (2007) proposed a general regression neural network model to predict the liquefaction potential in soil deposits with SPT-based data including field tests from the Turkey and Taiwan major earthquakes in 1999. Ramakrishnan et al. (2008) proposed a prediction model on liquefaction susceptibility of unconsolidated sediments using ANN model including the field data on SPT tests. Samui and Sitharam (2011) proposed two machine learning methods such as ANN and SVM to predict liquefaction susceptibility of soils based on the SPT data from the 1999 Chi-Chi, Taiwan earthquake. Hoang and Bui (2018) proposed a novel soft computing model named KFDDA-LSSVM (combines kernel Fisher discriminant analysis with a least squares support vector machine) to evaluate the earthquake-induced soil liquefaction. They used 3 historical data sets based on shear velocity, CPT,

and SPT including real cases of earthquake-induced soil liquefaction.

In this paper, an alternative and novel approach is proposed using the group method of data handling (GMDH) model, which is a type of an ANN. The GMDH model was first proposed by Ivakhnenko (1971, 1976) and the GMDH network is a self-organizing, machine learning method. While GMDH is self-organizing, it creates an optimal network by trying a number of networks in different architectures depending on the number of input variables. Recently, the GMDH method has begun to be applied in some geotechnical problems (Kordnaeij et al. 2015; Ardakani and Kordnaeij 2017; Hassanlourad et al. 2017; Jirdehi et al. 2014). In this regard, a novel ensemble GMDH model (EGMDH) based on classification with different activation function bases has been developed to best explain the relationship between input and output variables on predicting the liquefaction potential of soils with SPT-based field data from 2 major earthquakes (Chi-Chi, Taiwan earthquake, 21.09.1999, $M_w = 7.6$ and Kocaeli, Turkey earthquake, 17.08.1999, $M_w = 7.4$). The results of the proposed EGMDH model were also compared with other classifier models such as GMDH, artificial neural network (ANN) (Haykin 1994), support vector machine (SVM) (Cortes and Vapnik 1995), logistic regression (LR) (Le Cessie and Van Houwelingen 1992), and random forest (RF) (Ali et al. 2012).

Group method of data handling

The GMDH algorithm is a self-organizing approach based on evaluating performance on multiple input–single output data pairs. GMDH, proposed by Ivakhnenko in the 1970s (Vissikirsky et al. 2005), is an architectural class of polynomial neural network models. Since the GMDH network has a flexible structure, hybrid methods have been developed with intuitive methods such as genetic, evolutionary, and particle swarm optimization (Ghanadzadeh et al. 2012). The main implication of the GMDH model is to define an analytical function that enables weights to be obtained on a regression basis in forward feed neural networks using square neurons. In the GMDH network, neurons in a layer are bound to the next layer through a quadratic and triquadratic polynomial to form new neurons in the next layer. In this model, the input variables are mapped to the output variable. In this mapping, the goal is to construct the function $f()$ which will estimate the output value \hat{y} using the input vector $X = (X_1, X_2, X_3, \dots, X_n)$ (Kordnaeij et al. 2015). This function estimates the values as close as possible to real \hat{y} output values. When considering multiple input–single output, the function between them is expressed as follows (Ardakani and Kordnaeij 2017):

$$y_i = f(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) \quad (i = 1, 2, 3, \dots, M) \quad (1)$$

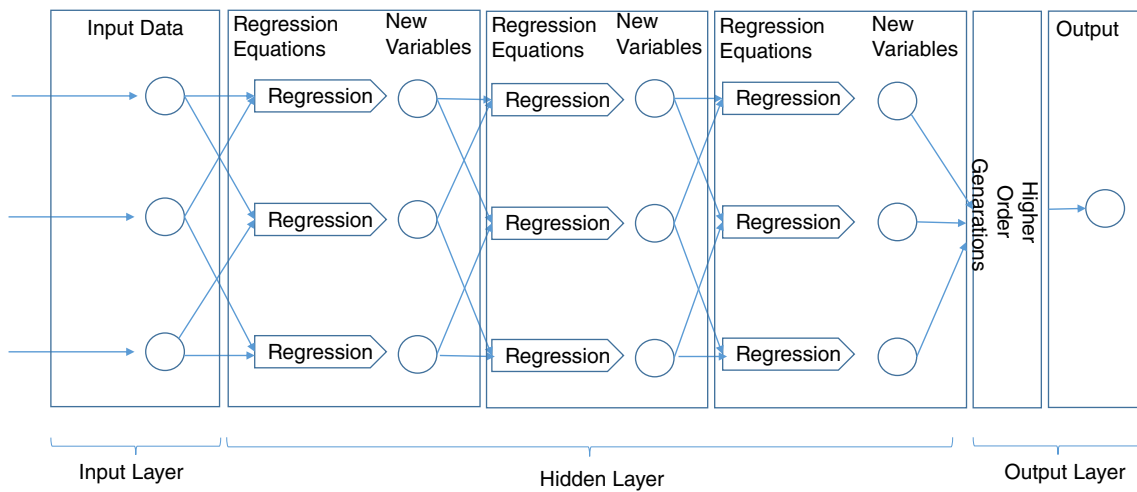


Fig. 1 GMDH network architecture

Thus, it is possible to estimate the output value \hat{y} by using the input vector $X = (X_{i1}, X_{i2}, X_{i3}, \dots, X_{in})$. Where, y_i is the dependent variable and x_i is the independent variable. The prediction equation can be written as:

$$\hat{y}_i = \hat{f}(X_{i1}, X_{i2}, X_{i3}, \dots, X_{in}) \quad (i = 1, 2, 3, \dots, M) \quad (2)$$

To solve this problem, the GMDH generates the general relation between output and input variables in the form of a mathematical definition also referred to as a reference. The aim here is to minimize the difference between the actual output values and the estimated values.

$$\sum_{i=1}^M [\hat{f}(x_{i1}, x_{i2}, x_{i3}, \dots, x_{in}) - y_i]^2 \rightarrow \text{Minimum} \quad (3)$$

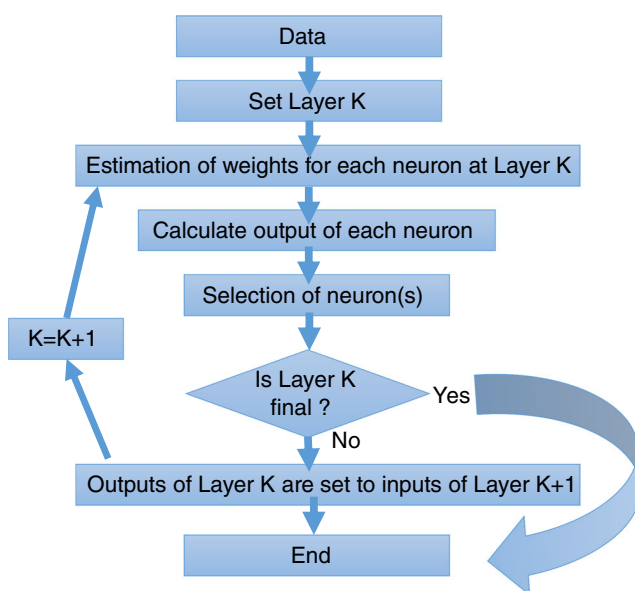


Fig. 2 The flowchart for the GMDH algorithm

The general connection between input and output variables can be expressed as a complex discrete form of a series of Volterra functions as below (Ardakani and Kordnaeij 2017; Zhu et al. 2012):

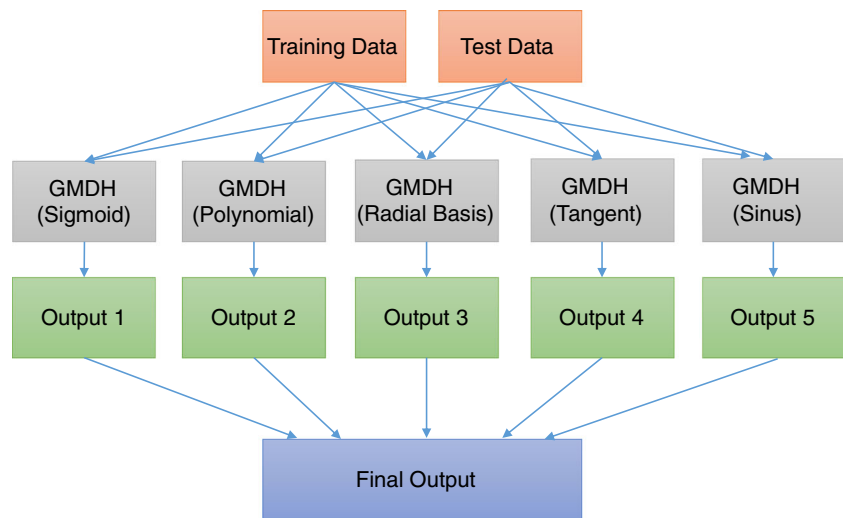
$$y = w_0 + \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n w_{ijk} x_i x_j x_k + \dots, \quad (4)$$

The above equation is known as the Kolmogorov–Gabor polynomial. This function is written as follows. GMDH uses a recursive polynomial regression procedure to synthesize any model. Polynomial regression equations can produce a high order polynomial model using effective predictors.

$$\text{Quadratic : } \hat{y} = G(x_i, x_j) = w_0 + w_1 x_i + w_2 x_j + w_3 x_i x_j + w_4 x_i^2 + w_5 x_j^2 \quad (5)$$

The mathematical relation between the input variables of the generated network and the output variable is formed by Eq. 4. The weights of the equation in Eq. 5 are calculated by regression methods. Thus, the difference between real y and estimated \hat{y} is minimized for input pairs x_i and x_j . The weights are obtained by a least squares method. In this way, the weighting coefficients of the quadratic function (G_i) are obtained so as to optimally fit the output set of all input–output data pairs. In the GMDH model, the output variables were tried to be estimated best way by taking all the input variables (two variables at a time) and creating a second-order polynomial equation (Eq. 5) in the training process. Each input vector pair (attributes) will form a second quadratic regression polynomial equation. For the first layer, the L ($L = m(m - 1)/2$) number of

Fig. 3 The algorithm of the EGMDH model



regression polynomial equations is obtain. Here, L is the number of polynomial equations to be obtained in any layer, and m is the number of variables that come to the layer. For example, if the input variable number $m = 4$, $L = 6$ regression polynomial equations will be obtained in the first layer. New variables are obtained for the next layer from the first layer using these equations. In this way, new variables are obtained for the other layers in each layer. Thus, new variables are generated which best explain the output variable from the input variables. If the minimum error value in the current layer is greater than the error value in the previous layer, the model becomes complicated. In other words, it is expected that the error value in a certain layer is smaller than the error value in the previous layer. GMDH network architecture is given in Fig. 1.

Each input data pair forms a regression equation. Outputs of the regression equations form new inputs to the next layer. The final output consists of the regression equations selected from all the layers. In the GMDH model, the aim is to have a minimum of error squares as specified in Eq. 6. The sum of the squares of the differences between the actual output values (y_i) and the estimated values ($G_f(x_i, x_j)$) is expected to be the smallest.

$$E = \frac{\sum_{i=1}^M (y_i - G_i(x_i, x_j))^2}{M} \rightarrow \text{minimum} \tag{6}$$

The GMDH network is constructed using all possible binary combinations of n input variables to construct the polynomial regression equation (in Eq. 4) that best predicts the independent y variable with the least squares method. From the observed $\{(y_i, x_{ip}, x_{iq}), (i = 1, 2, 3, \dots, M)\}$ samples, the first layer of the GMDH network is constructed using $n(n - 1)/2$ quadratic polynomial neurons.

$$\begin{bmatrix} x_{1p} & x_{1q} & \vdots & y_1 \\ x_{2p} & x_{2q} & \vdots & y_2 \\ \dots & \dots & \vdots & \dots \\ x_{mp} & x_{mq} & \vdots & y_m \end{bmatrix} \tag{7}$$

Here, p and q are the any two variables that come into the layer. Equation 4 can be written in matrix form as follows using the input–output variables mentioned above:

$$AW = Y \tag{8}$$

Where, W is the vector of the unknown weight coefficients of the quadratic polynomial and Y specifies the vector of the output values.

$$W = \{w_0, w_1, w_2, w_3, w_4, w_5\}^T \tag{9}$$

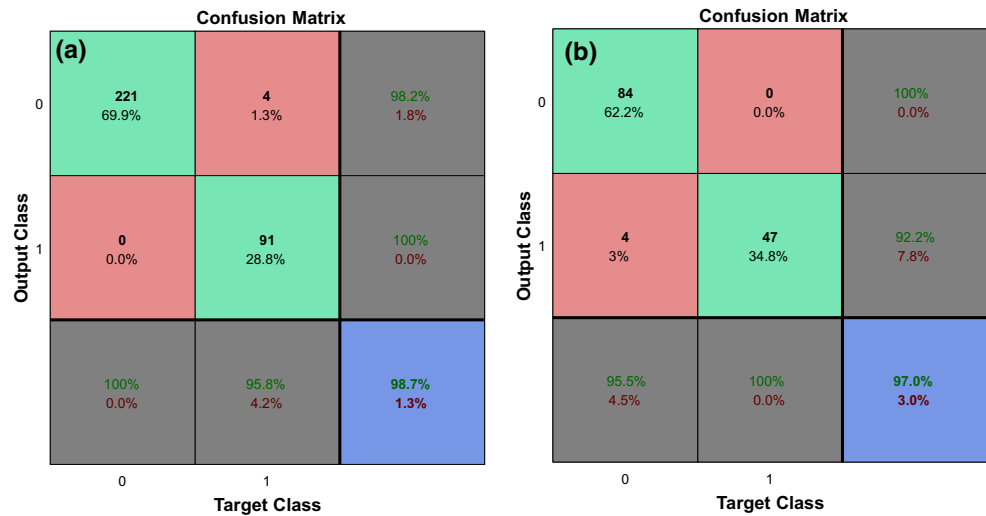
		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Fig. 4 The simplified confusion matrix

Table 1 Success rates of the GMDH model according to different training-test ratios

Train-test rate	Accuracy (%)	Precision	Recall	F measure
50–50%	96.40	0.962	0.987	0.974
60–40%	96.10	0.976	0.968	0.971
70–30%	97.00	0.955	1.000	0.976

Fig. 5 Confusion matrix by GMDH for 70–30% training-test data. **a** Training data. **b** Test data



$$Y = \{y_1, y_2, y_3, y_4, y_5, \dots, y_M\}^T \tag{10}$$

$$A = \begin{bmatrix} 1 & x_{1p} & x_{1q} & x_{1p}x_{1q} & x_{1p}^2 & x_{1q}^2 \\ 1 & x_{2p} & x_{2q} & x_{2p}x_{2q} & x_{2p}^2 & x_{2q}^2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{mp} & x_{mq} & x_{mp}x_{mq} & x_{mp}^2 & x_{mq}^2 \end{bmatrix} \tag{11}$$

The weights are solved in matrix form using multiple regression equations as follows:

$$W = (A^T A)^{-1} A^T Y \tag{12}$$

Where, W is the weight vector to be estimated, A is the input matrix, and Y is the output vector. The flowchart for the GMDH algorithm is shown in Fig. 2.

Ensemble GMDH model

The main goal in ensemble classification is to achieve a result by combining the values obtained by different classifiers. The combination of the classifiers consists of the processes of performing the classification process in the direction of the estimates resulting from the training of the resampled training sets and the classifiers separately. In general, it is stated that the accuracy of classification with the classifier obtained as a

Table 2 GMDH success rates for different numbers of hidden layers with 70–30% training-test set (with 10 neurons)

No. of hidden layer	Accuracy for training (%)	Accuracy for testing (%)
1	92.4	83.7
2	97.2	95.6
3	98.7	97.0

result of combining is better than when each classifier is used singularly. Because, while a single classifier can have a higher test error, the diversity of classifiers usually compensates for the mistakes of a single classifier. Therefore, fewer test errors are obtained with the combination of classifiers (Pal and Mather 2003). The main goal in ensemble classification is to produce a result by combining the values previously obtained by different classifiers. During this process, it is tried to calculate by giving certain weight points to the other classifiers. The main problem here is to combine different classification algorithms and decide which ratios to use. The most advantage is that it can get better values due to it uses the data of other methods together (Augusty and Izudheen 2013).

In the current study, the GMDH has been ensembled using different activation functions under the same conditions (learning rate, number of hidden layers, weights, number of neurons in hidden layer). Activation functions are used to better explain the relationship between input and output (Kondo and Ueno 2012). These activation functions are given below:

$$Sigmoid = \frac{1}{1 + e^{-y}} \tag{13}$$

$$Radial\ Basis = e^{-y^2} \tag{14}$$

$$Polynomial = y \tag{15}$$

Table 3 GMDH success rates of different neurons in hidden layers with 70–30% training-test set (with 3 hidden layers)

No. of neurons	Accuracy for training (%)	Accuracy for testing (%)
1	92.5	83.7
3	92.5	83.7
5	98.4	97.0
10	98.7	97.0
15	98.7	97.0

Table 4 EGMDH model success rates for different numbers of the hidden layers with 70–30% training-test set (with 10 neurons)

No. of hidden layer	Accuracy (%)	Precision	Recall	F measure
1	97.00	0.955	1.000	0.976
2	97.00	0.955	1.000	0.976
3	99.3	0.989	1.000	0.994

$$\text{Tangent} = \tanh(y) \tag{16}$$

$$\text{Sinus} = \sin(y) \tag{17}$$

The diagram of the proposed ensemble GMDH (EGMDH) model is shown in Fig. 3. Outputs of 5 GMDH models operated under the same conditions are produced with different activation functions. Each model makes its own output decision for example data. However, the output of EGMDH is the community decision of these 5 models.

Performance criteria

In assessing the performance of the classification models used in machine learning, the confusion matrix which compares the actual and predicted values is frequently used (Fig. 4) (Kaya 2013). Accuracy, precision, recall, and F-criterion were used to demonstrate the performance of the methods proposed in the study. These success criteria are calculated from confusion matrix as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

$$\text{Precision} = TP / (TP + FP) \tag{19}$$

$$\text{Recall} = TP / (TP + FN) \tag{20}$$

$$\text{F-criterion} = 2(\text{Recall} \times \text{Precision}) / (\text{Recall} + \text{Precision}) \tag{21}$$

In these equations, *T*, *F*, *P*, and *N* express true, false, positive, and negative, respectively. For example, *TP* indicates the number of positive samples correctly classified; *FN* indicates the number of false negative samples misclassified.

Accuracy is the most popular and simple method used to determine success and is defined as the ratio of the number of correctly classified (*TP + TN*) samples to the total number of samples (*TP + TN + FP + FN*). Precision gives the degree of precision of the classifier result and defined as the ratio of positive-labeled sample number (*TP*) to the proportion of total samples (*TP + FP*) that are classified as positive. Recall is the ratio of positively labeled samples (*TP*) to the total number of truly positive samples (*TP + FN*). The F-criterion is calculated using the precision and recall metrics. It is used to optimize the system towards the direction of precision or recall.

Data processing

The database used in the present study was preferred for predicting the presence of liquefaction by the EGMDH model on the SPT-based liquefaction assessment. In this context, 451 SPT-based field data from 2 major earthquakes in 1999 were used. Both case records were obtained from Hanna et al. (2007). Two hundred thirty-nine of the case records belong to the Chi-Chi, Taiwan earthquake ($M_w = 7.6$) and 212 of them belong to the Kocaeli, Turkey earthquake ($M_w = 7.4$). Three hundred nine of the records were classified as non-liquefied and 142 of them as liquefied. The input parameters used in the EGMDH model are SPT blow numbers ($N_{1,60}$), percent finest content less than $75 \mu\text{m}$ ($F \leq 75 \mu\text{m}$, %), depth of groundwater table (d_w), total and effective overburden stresses (σ_{vo} , σ'_{vo}), maximum peak ground acceleration (a_{max}), and magnitude of earthquake (M_w) and the output is the occurrence of liquefaction.

Fig. 6. Confusion matrix by EGMDH for 70–30% training-test data. **a** Training data. **b** Test data

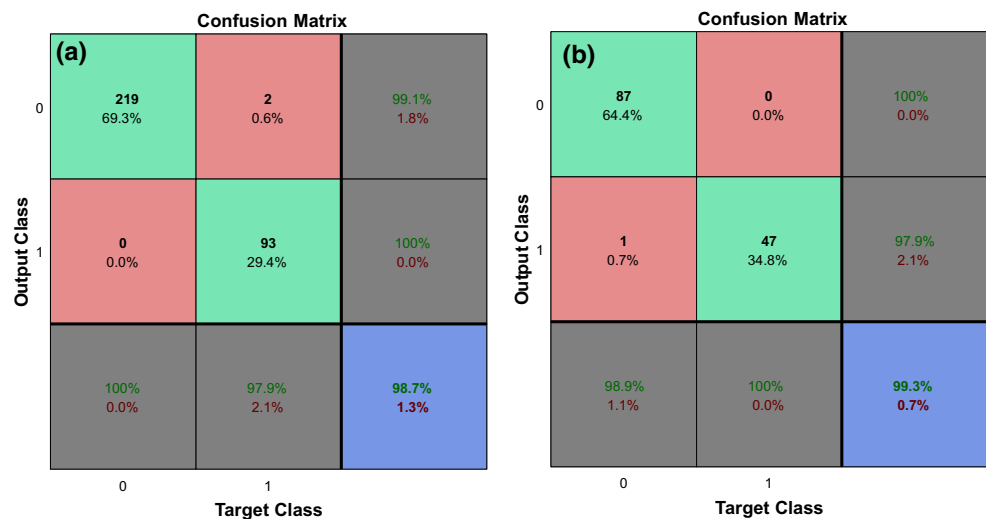


Table 5 Comparison of EGMDH with other classifier models

Model	Accuracy (%)	Precision	Recall	F measure
ANN	94.07	0.941	0.941	0.941
SVM	74.81	0.798	0.748	0.703
LR	82.96	0.836	0.830	0.820
RF	94.81	0.948	0.948	0.948
GMDH	97.00	0.955	1.000	0.976
EGMDH	99.30	0.989	1.000	0.994

SPT-based liquefaction assessment

The liquefaction depends on many factors such as particle size and distribution, geological age and sedimentation conditions, volume change potential, permeability, water table level, earthquake magnitude and duration, and distance to center. In general, loose sandy soils that are saturated with water are more sensitive to liquefaction during large earthquakes (Kramer 1996 and Coduto 2003). Liquefaction can only occur if all affecting factors such as loose soil, water saturation, and large and long-term earthquake magnitude are present at the same time.

In the literature, the most important study to investigate the liquefaction potential of soils has been presented by Seed and Idriss (1971) as a “simplified procedure.” Seed and Idriss (1971) have basically expressed the liquefaction potentials of the soils by two parameters. The first parameter is the ratio of cyclic stress (CSR) which indicates the level of cyclic loading that can be caused by the earthquake, and the second parameter is the rate of cyclic resistance (CRR) that indicates the resistance of the soil against the liquefaction. The ratio of cyclic stress generated during earthquakes (CSR) is defined as in Eq. 22.

$$CSR = 0.65 \times \frac{a_{max}}{g} \times \frac{\sigma_v}{\sigma'_v} \times r_d \tag{22}$$

Here, a_{max} is the peak horizontal acceleration at the ground surface during the earthquake; g the gravitational acceleration; σ_v and σ'_v the total and effective stress; and r_d the stress reduction coefficient. The average values are used for the r_d depending on the depth in Eq. 23 in engineering applications.

Table 6 The performance of the models on the prediction of field cases

	Liquefaction	Non-liquefaction
Field cases	47	88
GMDH results	47	84
GMDH error	–	4
EGMDH results	47	87
EGMDH error	–	1

$$r_d = \begin{cases} 1.0 - 0.00765z, & z \leq 9.15 \text{ m} \\ 1.174 - 0.0267z, & 9.15 < z \leq 23 \text{ m} \end{cases} \tag{23}$$

In order to determine the rate of cyclic resistance (CRR), Youd et al. (2001) suggested the following equation:

$$CRR = \frac{1}{34 - (N1)_{60}} + \frac{(N1)_{60}}{135} + \frac{50}{[10 \times (N1)_{60} + 45]^2} - \frac{1}{200} \tag{24}$$

Corrected SPT- N values used in the liquefaction analysis are suggested to be corrected as follows, taking into account the effect of fine grain ratio (FC) on liquefaction resistance:

$$N1_{60,CS} = \alpha + \beta N1_{60} \tag{25}$$

$$\alpha = 0 \text{ and } \beta = 1 \text{ for } FC \leq 5\% \tag{26a}$$

$$\alpha = \exp\left(1, 76 - \frac{190}{FC^2}\right) \text{ and } \beta = \left[0, 99 + \frac{FC^{1.5}}{1000}\right] \tag{26b}$$

for $5\% < FC < 35\%$

$$\alpha = 1 \text{ and } \beta = 1, 2 \text{ for } FC \geq 35\% \tag{26c}$$

Here, α and β are fine grain ratio correction coefficients; CS is the correction coefficient.

The safety factor for the liquefaction risk is defined as FS (Eq. 27). If the safety factor is less than 1, it means that the zone involves the risk of liquefaction; if the factor is greater than 1, it indicates that the zone does not involve the risk of liquefaction.

$$FS = \frac{CRR}{CSR} \tag{27}$$

Results

GMDH model

The occurrence of liquefaction in soils has been tried to estimate firstly with using the GMDH algorithm. GMDH is a nonlinear regression method, but is a model that also carries the characteristics of supervised and unsupervised artificial neural networks (ANNs). Regression is a statistical model that examines the cause-and-effect relationship between independent variables and dependent variables. Linear regression model is modeling the relationship between one or more independent variables and dependent variables. Trials with the GMDH model were conducted according to different training-test set ratios. The performance results are given in Table 1.

Due to no criterion in the literature as to which rates of training-test sets should be made, the authors have experimented with training-test sets at different rates. It can be seen in Table 1 that the trials were conducted for data sets with different ratios in the form of 50–50%, 60–40%, and 70–30%

Table 7 Comparison of actual and predicted liquefaction records

No.	Z (m)	$N_{1,60}$	$F \leq 75 \mu\text{m}$ (%)	d_w (m)	σ (kPa)	σ' (kPa)	M_w	a_{max} (g)	Liquefaction	GMDH	EGMDH
1	19.2	13	60	0.8	358.7	178.2	7.6	0.18	No	No	No
2	2.2	9	15	2	39.1	37.1	7.6	0.18	Yes	Yes	Yes
3	5.8	6	54	2	110.2	73	7.6	0.18	Yes	Yes	Yes
4	7.2	14	9	2	137.5	86.4	7.6	0.18	Yes	Yes	Yes
5	8.8	7	96	2	171	104.3	7.6	0.18	No	No	No
6	11.8	23	18	2	228.5	132.3	7.6	0.18	No	No	No
7	13.2	25	7	2	257.4	147.6	7.6	0.18	No	No	No
8	14.8	26	20	2	292.1	166.5	7.6	0.18	No	No	No
9	16.2	17	10	2	323.4	184.1	7.6	0.18	No	No	No
10	17.8	21	6	2	358.3	203.3	7.6	0.18	No	No	No
11	19.2	10	29	2	388.9	220.2	7.6	0.18	No	Yes	No
12	2.2	7	59	1.1	36.4	25.7	7.6	0.18	Yes	Yes	Yes
13	3.8	10	17	1.1	66.5	40	7.6	0.18	Yes	Yes	Yes
14	5.2	9	61	1.1	93.6	53.4	7.6	0.18	Yes	Yes	Yes
15	6.8	10	15	1.1	124	68.1	7.6	0.18	Yes	Yes	Yes
16	8.2	16	15	1.1	153.5	83.9	7.6	0.18	Yes	Yes	Yes
17	9.8	13	98	1.1	186.9	101.6	7.6	0.18	No	No	No
18	11.2	20	20	1.1	212.2	113.1	7.6	0.18	No	No	No
19	12.8	25	7	1.1	244.1	129.3	7.6	0.18	No	No	No
20	14.2	19	18	1.1	275.1	146.6	7.6	0.18	No	No	No
21	15.8	24	18	1.1	308	163.8	7.6	0.18	No	No	No
22	17.2	21	10	1.1	339.4	181.4	7.6	0.18	No	No	No
23	18.8	16	98	1.1	374.6	201	7.6	0.18	No	No	No
24	2.8	6	30	2.3	43.5	38.6	7.6	0.18	Yes	Yes	Yes
25	10.1	18	23	2.3	180.5	103.9	7.6	0.18	No	Yes	No
26	13.2	11	94	2.3	241	134.1	7.6	0.18	No	No	No
27	14.8	18	67	2.3	272.3	149.7	7.6	0.18	No	No	No
28	17.8	18	20	2.3	331.9	179.9	7.6	0.18	No	No	No
29	19.2	43	29	2.3	361.1	195.3	7.6	0.18	No	No	No
30	1.2	25	81	2.3	16.4	16.4	7.6	0.18	No	No	No
31	2.8	5	99	2.3	40.1	35.1	7.6	0.18	No	No	No
32	4.2	6	19	2.3	64.4	45.7	7.6	0.18	Yes	Yes	Yes
33	5.8	7	99	2.3	96.1	61.8	7.6	0.18	No	No	No
34	7.2	7	99	2.3	121.2	73.1	7.6	0.18	No	No	No
35	8.8	7	99	2.3	150.1	86.4	7.6	0.18	No	No	No
36	10.2	6	99	2.3	175.3	97.8	7.6	0.18	No	No	No
37	11.8	8	66	2.3	203.1	110	7.6	0.18	Yes	Yes	Yes
38	13.2	21	10	2.3	229.7	122.8	7.6	0.18	No	No	No
39	14.8	19	8	2.3	262.4	139.8	7.6	0.18	No	No	No
40	16.2	16	7	2.3	291.6	155.3	7.6	0.18	No	No	No
41	17.2	8	82	2.3	313	166.8	7.6	0.18	No	No	No
42	19.2	10	97	2.3	351.7	185.9	7.6	0.18	No	No	No
43	2.8	13	80	0.99	50.9	33.2	7.6	0.38	Yes	Yes	Yes
44	4.2	8	80	0.99	78.6	47.1	7.6	0.38	Yes	Yes	Yes
45	12.2	18	95	0.99	239.3	129.3	7.6	0.38	No	No	No
46	15.2	21	78	0.99	299.5	160.1	7.6	0.38	No	No	No
47	17.2	30	20	0.99	339.1	180	7.6	0.38	No	No	No
48	3.2	15	10	0.7	84.1	59.6	7.6	0.38	Yes	Yes	Yes
49	6.2	20	32	0.7	150.4	96.4	7.6	0.38	Yes	Yes	Yes

Table 7 (continued)

No.	Z (m)	$N_{1,60}$	$F \leq 75 \mu\text{m}$ (%)	d_w (m)	σ (kPa)	σ' (kPa)	M_w	a_{max} (g)	Liquefaction	GMDH	EGMDH
50	8.6	13	57	0.7	206.1	128.6	7.6	0.38	Yes	Yes	Yes
51	10.2	18	38	0.7	237.1	143.9	7.6	0.38	Yes	Yes	Yes
52	13.2	33	58	0.7	299.8	177.2	7.6	0.38	No	No	No
53	14.2	36	19	0.7	320.6	188.2	7.6	0.38	No	No	No
54	14.8	62	11	0.7	333.6	195.3	7.6	0.38	No	No	No
55	16.2	49	26	0.7	362.6	210.5	7.6	0.38	No	No	No
56	1.2	41	22	5	24.8	24.8	7.6	0.38	No	No	No
57	5.8	9	25	5	96.8	89	7.6	0.38	Yes	Yes	Yes
58	8.8	14	43	5	161.1	123.8	7.6	0.38	Yes	Yes	Yes
59	11.8	13	85	5	221.9	155.2	7.6	0.38	No	No	No
60	13.2	17	80	5	250.4	169.9	7.6	0.38	No	No	No
61	14.8	18	97	5	282.2	186.1	7.6	0.38	No	No	No
62	17.8	31	80	5	342.5	217	7.6	0.38	No	No	No
63	19.2	38	29	5	371.3	232	7.6	0.38	No	No	No
64	1.2	27	13	1.5	21.5	21.5	7.6	0.38	No	No	No
65	3.8	15	30	1.5	74.1	51.5	7.6	0.38	Yes	Yes	Yes
66	5.8	14	30	1.5	115.1	73	7.6	0.38	Yes	Yes	Yes
67	10.8	4	97	1.5	220.2	129	7.6	0.38	No	No	No
68	11.8	16	98	1.5	240.3	139.3	7.6	0.38	No	No	No
69	13.2	14	96	1.5	267.4	152.7	7.6	0.38	No	No	No
70	14.8	18	98	1.5	299.3	168.9	7.6	0.38	No	No	No
71	16.2	18	99	1.5	326.4	182.2	7.6	0.38	No	No	No
72	17.8	28	36	1.5	358.2	198.3	7.6	0.38	No	No	No
73	18.8	63	16	1.5	378.9	209.2	7.6	0.38	No	No	No
74	7.2	6	94	1.9	147.5	95.5	7.6	0.38	No	No	No
75	8.8	63	67	1.9	179.4	111.7	7.6	0.38	No	Yes	No
76	12	14	75	1.9	245.5	146.4	7.6	0.38	No	No	No
77	13.2	42	42	1.9	269.7	158.9	7.6	0.38	No	No	No
78	14.8	36	35	1.9	302.7	176.2	7.6	0.38	No	No	No
79	16.2	30	91	1.9	332.1	191.8	7.6	0.38	No	No	No
80	17.8	33	49	1.9	366	210.1	7.6	0.38	No	No	No
81	4.2	8	32	1.14	76.7	46.7	7.6	0.67	Yes	Yes	Yes
82	5.2	21	46	1.14	96.1	56.3	7.6	0.67	Yes	Yes	Yes
83	6.2	16	15	1.14	114.1	64.5	7.6	0.67	Yes	Yes	Yes
84	8.6	18	29	1.14	163.2	90	7.6	0.67	Yes	Yes	Yes
85	11.2	23	33	1.14	216.7	118	7.6	0.67	No	No	No
86	12.2	36	32	1.14	238.2	129.7	7.6	0.67	No	No	No
87	14.2	58	32	1.14	281.5	153.4	7.6	0.67	No	No	No
88	4.2	9	54	0.75	82.5	48.7	7.6	0.67	Yes	Yes	Yes
89	7.2	11	55	0.75	144.6	81.3	7.6	0.67	Yes	Yes	Yes
90	10.2	21	52	0.75	206.5	113.8	7.6	0.67	Yes	Yes	Yes
91	8.8	15	51	0.5	176.8	95.4	7.6	0.67	Yes	Yes	Yes
92	11.2	13	40	0.5	226.6	121.7	7.6	0.67	Yes	Yes	Yes
93	16.2	30	10	0.5	334.5	180.5	7.6	0.67	No	No	No
94	17.2	28	13	0.5	356.3	192.4	7.6	0.67	No	No	No
95	19.2	32	10	0.5	400.1	216.6	7.6	0.67	No	No	No
96	2.8	9	31	1.57	55.2	43.1	7.6	0.67	Yes	Yes	Yes
97	3.8	18	19	1.57	75.8	54	7.6	0.67	Yes	Yes	Yes
98	3.8	4	31	3.6	58.9	56.9	7.6	0.67	Yes	Yes	Yes

Table 7 (continued)

No.	Z (m)	$N_{1,60}$	$F \leq 75 \mu\text{m}$ (%)	d_w (m)	σ (kPa)	σ' (kPa)	M_w	a_{max} (g)	Liquefaction	GMDH	EGMDH
99	6.2	14	36	3.6	109.2	83.7	7.6	0.67	Yes	Yes	Yes
100	9.2	41	27	3.6	174.9	120	7.6	0.67	No	No	No
101	13.2	25	67	3.6	263.3	169.2	7.6	0.67	No	No	No
102	14.2	22	52	3.6	284.6	180.6	7.6	0.67	No	No	No
103	3.8	16	19	1.03	73.7	46.5	7.6	0.67	Yes	Yes	Yes
104	7.8	11	10	1.03	157.9	91.5	7.6	0.67	Yes	Yes	Yes
105	9.8	15	23	1.03	198.9	112.8	7.6	0.67	Yes	Yes	Yes
106	18.8	26	26	1.03	387.9	213.5	7.6	0.67	No	No	No
107	19.8	28	17	1.03	408.9	224.7	7.6	0.67	No	No	No
108	1.2	11	49	3.2	19.3	19.3	7.6	0.67	No	Yes	Yes
109	5.8	16	50	3.2	108.7	83.2	7.6	0.67	Yes	Yes	Yes
110	7.2	16	40	3.2	138.4	99.2	7.6	0.67	Yes	Yes	Yes
111	10.2	13	65	3.2	200.9	132.3	7.6	0.67	Yes	Yes	Yes
112	11.8	27	27	3.2	234	149.6	7.6	0.67	No	No	No
113	13.2	30	20	3.2	263.2	165.1	7.6	0.67	No	No	No
114	14.8	35	49	3.2	296.9	183.1	7.6	0.67	No	No	No
115	18.8	24	45	3.2	386.7	233.7	7.6	0.67	No	No	No
116	1.2	3	65	0.5	21.7	14.8	7.6	0.67	Yes	Yes	Yes
117	3.8	8	61	0.5	74	41.6	7.6	0.67	Yes	Yes	Yes
118	7.5	25	45	0.5	151.7	83	7.6	0.67	Yes	Yes	Yes
119	8.8	21	44	0.5	179.1	97.7	7.6	0.67	Yes	Yes	Yes
120	11.8	27	58	0.5	242.3	131.4	7.6	0.67	No	No	No
121	13.2	17	77	0.5	270.8	146.3	7.6	0.67	No	No	No
122	14.8	38	29	0.5	304.7	164.4	7.6	0.67	No	No	No
123	16.2	13	72	0.5	334.1	180.1	7.6	0.67	No	No	No
124	17.8	13	76	0.5	367.5	197.7	7.6	0.67	No	No	No
125	19.2	11	82	0.5	396.8	213.3	7.6	0.67	No	No	No
126	3.8	14	69	1.4	70.3	46.7	7.6	0.67	Yes	Yes	Yes
127	5.8	16	22	1.4	109.3	66.2	7.6	0.67	Yes	Yes	Yes
128	7.2	19	70	1.4	139.1	82.2	7.6	0.67	Yes	Yes	Yes
129	8.8	28	71	1.4	173.1	100.5	7.6	0.67	No	No	No
130	11.8	26	51	1.4	236.6	134.6	7.6	0.67	No	No	No
131	13.2	18	74	1.4	266.4	150.6	7.6	0.67	No	No	No
132	14.8	23	40	1.4	299.3	167.8	7.6	0.67	No	No	No
133	16.2	37	35	1.4	328.6	183.5	7.6	0.67	No	No	No
134	17.8	16	67	1.4	362.9	202	7.6	0.67	No	No	No
135	19.2	17	67	1.4	391.8	217.2	7.6	0.67	No	No	No

training-test. The highest success was achieved as 97.00% for the 70–30% training-test data set. Performance measures are calculated from the confusion matrix. The confusion matrix for the 70–30% training-test set is given in Fig. 5. It is seen on the confusion matrix that only 4 samples are classified as incorrect for both training and test sets.

The GMDH can be used in architecture built in different numbers of layers and with different numbers of neurons in each layer. The performance measures obtained as a result of trials with different hidden layer numbers in GMDH

architectures are given in Table 2. Since the number of input variables is low, the trials have been performed for the hidden layer numbers 1, 2, and 3. As a result of these trials, it is seen that when the hidden layer number increases, the success rate also increases. The highest success is achieved when the hidden layer number is 3 (Table 2).

The performance measures obtained as a result of trials using different numbers of neurons in the hidden layers of a 3-layered GMDH model are given Table 3. As can be seen in Table 3, the performance for the presence of 5, 10, and 15

Table 8 Sensitivity analysis results

Features	InfoGainParameterEval	ChiSquaredParameterEval
SPT-N	0.0837	33.427
σ (kPa)	0.107	30.537
σ' (kPa)	0.429	165.282
d_w (m)	0	0
Z (m)	0.0898	35.6
a_{\max} (g)	0.0878	35.853
M_w	0.0253	10.151
$F \leq 75$ mm (%)	0.0253	10.151

neurons in hidden layers has not changed. The best success rate for the test set was 97.00% in the presence of 5 neurons in hidden layers.

EGMDH model

In this paper, a GMDH-based new approach was proposed in the prediction of soil liquefaction. A novel ensemble GMDH (EGMDH) model with different activation functions has been developed to best explain the relationship between input and output variables by changing the GMDH algorithm. The GMDH model was used for each of the sigmoid, radial basis, sin, tangent, and polynomial functions. Then, each GMDH classifier model with an activation function was combined to produce a common output. In general, it is stated that the classification accuracy with the classifier obtained as a result of combining is better than using each classifier singly. Because, the diversity of classifiers usually compensates for the mistakes of a single classifier which can have a higher test error when used singly. Thus, less test error is achieved with the combination of classifiers. The success rates obtained with the EGMDH model for the 70–30% training-test set are given in Table 4.

It is seen that the EGMDH model is more successful than the GMDH model on the estimation of soil liquefaction when Table 4 is examined. A high classification success rate as 99.30% was obtained with EGMDH. The output confusion matrix for the EGMDH model is given in Fig. 6. As it can be seen in Fig. 5b, only one sample of “1” (liquefied) state is misclassified. All samples representing the state “0” (non-liquefied) are correctly classified.

Discussion

This study was aimed to develop a novel prediction model for the liquefaction potential of soils by using the ensemble group method of data handling (EGMDH) algorithm based on the GMDH model. For this aim, the GMDH model has been converted to an ensemble model for different activation

functions. The main goal in the ensemble classification is to achieve a result by combining the values obtained by different classifiers. The combination of classifiers consists of resampled training sets, training of classifiers separately, and realization of the classification process in the direction of the emerging estimates. The accuracy of the classification made with the classifier obtained as a result of combining is better when each classifier is used singularly.

Totally 451 SPT-based field records obtained from 2 major earthquakes were used for the prediction models. The success rate of the liquefaction prediction achieved with the GMDH model was 97.00%, while it increased to 99.30% with EGMDH. The EGMDH model is also compared with different classifier models such as ANN, SVM, LR, and RF. Performance values for all models are shown in Table 5. It is obvious that the performance of the proposed EGMDH model is more successful than other classifier models.

There are 88 “non-liquefaction” and 47 “liquefaction” cases in the test set. Both the GMDH and EGMDH models have been much more capable on the prediction of “liquefaction cases” as seen in Table 6. The data of all cases in the test set and the estimation results of the models are given in Table 7. The proposed EGMH model has only one false estimate on liquefaction status.

As mentioned in the “Data processing” section, the data used in this study were obtained from the Hanna et al. (2007) study. They proposed a GRNN model on SPT-based liquefaction assessment. The success of the GRNN model was 92.9% for the test set, 94.7% for the forecast set, and 97% for the total set of data. It is understood that the proposed EGMDH model with 99.3% success performance has achieved higher success compared with the GRNN model.

The results achieved by the proposed EGMDH model were also compared well with other artificial intelligence (AI) works on predicting liquefaction status in the literature. The success rates on the results of some studies are summarized below; Rahman and Wang (2002) proposed a fuzzy neural network model for SPT-based liquefaction prediction with 205 field records. They used 27 cases for testing and achieved a 81.5% success rate with five misclassified cases. Ramakrishnan et al. (2008) proposed a SPT-based ANN model for predicting the liquefaction susceptibility of unconsolidated sediments. They used 23 case records and the model performance was tested with 5 cases and achieved a success of 99.9%. However, the input parameters used in the proposed model were selected as the liquefaction severity index, liquefaction sensitivity index, and estimated CRR and CSR values unlike the similar studies. Samui and Sitharam (2011) proposed ANN and SVM models to predict liquefaction susceptibility of soils based on the SPT data by using 288 case records. They used only two input parameters in the models (CSR $N_{1,60}$ or PGA $N_{1,60}$). The performances of both models in the test set were found to be between 87.2% and 88.37% for

the ANN model and between 94.19% and 95.35% for the SVM model. Muduli and Das (2015a) have studied the uncertainty of the SPT-based method for assessing the seismic soil liquefaction potential using multi-gene genetic programming (MGGP). Statistical performances of the developed “best” MGGP-based CRR model were found to be as $R = 0.96$ for training and $R = 0.98$ for testing. Hoang and Bui (2018) proposed a novel soft computing model named KFDDA-LSSVM (combines kernel Fisher discriminant analysis - least squares support vector machine) on the prediction of shear velocity and CPT- and SPT-based soil liquefaction. The results of the proposed KFDDA-LSSVM were compared with the other benchmark models including LSSVM, extreme learning machine (ELM), and support vector machine (SVM). The success rates of the models obtained in the SPT-based prediction were 84.95% for KFDDA-LSSVM, 84.06% for LSSVM, 82.63% for SVM, and 80.05% for ELM.

The abovementioned studies were performed with different or same case records, different input parameters or numbers, and different approach methods. The common aspect of achievements in all methods is that they will be a good alternative to traditional calculation methods on determining the liquefaction susceptibility of soils. It is obvious that the proposed EGMDH model in the present study can be used as an effective alternative model on predicting the liquefaction potential just like the other successful models.

Sensitivity analysis

A sensitivity analysis of the proposed model was carried out to evaluate the input parameters influence on the model output. Sensitivity analysis is the selection of appropriate parameters for a classification algorithm. The parameters in the data sets are one of the most important factors affecting the classification performance. The low number of parameters may, in some cases, cause classes not to be properly separated. In the case of a high number of parameters, it leads to problems such as an increase in training time and a decrease in the accuracy rate of the parameters. Therefore, it is necessary to determine the correct number of parameters. Considering the multiplicity of sample numbers in the data sets, it is almost impossible to perform the parameter determination manually. Thus, different approaches have been proposed for parameter selection procedures (Das and Basudhar 2008). In the present study, our data set consists of 8 inputs and one output parameter. InfoGainParameterEval and ChiSquaredParameterEval weighting methods were used to determine the effect of input parameters on the output parameter. The InfoGainParameterEval method uses the information gain that each input parameter generates with the output parameter (Lee and Lee 2006). The ChiSquaredParameterEval method uses the chi-square statistic between input parameters and output parameter (Aggarwal 2013). The weight values obtained for each parameter are given in Table 8. It is found out that the most

effective input parameter according to both methods is the effective overburden stress (σ'), while the most ineffective input parameter is the depth of groundwater table (d_w) (Table 8).

Conclusions

An ensemble model based on GMDH-type neural network was proposed for the prediction of SPT-based liquefaction assessment in this paper. The proposed novel approach achieved successful results with efficient (almost 100%) accuracy in predicting the liquefaction potential of soils. All the models used, including the proposed new ensemble model in the present study, have been much more effective on the prediction of “liquefaction” cases compared with the “non-liquefaction” cases. Despite the fact that there are many studies in the literature on the prediction of liquefaction with different artificial intelligence techniques, the authors believe that new models for predicting the liquefaction phenomena will continue to be developed just as the EGMDH model, proposed in this study.

References

- Abdalla JA, Attom MF, Hawileh R (2015) Prediction of minimum factor of safety against slope failure in clayey soils using artificial neural network. *Environ Earth Sci* 73(9):5463–5477
- Aggarwal M (2013) Performance analysis of different feature selection methods in intrusion detection. *Int J Sci Technol Res* 2(6):225–231
- Ali J, Khan R, Ahmad N, Maqsood I (2012) Random forests and decision trees. *Int J Comput Sci Issues* 9(5):272
- Andrus RD, Stokoe KH II (2000) Liquefaction resistance of soils from shear wave velocity. *J Geotech Geoenviron Eng* 126(11):1015–1025
- Ardakani A, Kordnaej A (2017) Soil compaction parameters prediction using GMDH-type neural network and genetic algorithm. *Eur J Environ Civ Eng* 23(4):449–462. <https://doi.org/10.1080/19648189.2017.1304269>
- Augusty SM, Izudheen S (2013) Ensemble classifiers A survey: evaluation of ensemble classifiers and data level methods to deal with imbalanced data problem in protein- protein interactions. *Rev Bioinformatics Biometrics* 2(1):1–9
- Baziar MH, Nilipour N (2003) Evaluation of liquefaction potential using neural-networks and CPT results. *Soil Dyn Earthq Eng* 23(7):631–636
- Boulanger RW, Idriss IM (2012) Probabilistic standard penetration test-based liquefaction-triggering procedure. *J Geotech Geoenviron Eng ASCE* 138(10):1185–1195
- Cetin KO, Seed RB, Der Kiureghian A, Tokimatsu K, Harder JLF, Kayen RE, Moss RES (2004) SPT-based probabilistic and deterministic assessment of seismic soil liquefaction potential. *ASCE J Geotech Geoenviron Eng* 130(12):1314–1340
- Chenari RJ, Tizpa P, Rad MRG, Machado SL, Fard MK (2015) The use of index parameters to predict soil geotechnical properties. *Arab J Geosci* 8(7):4907–4919
- Chern SG, Lee CY, Wang CC (2008) CPT-based liquefaction assessment by using fuzzy-neural network. *J Mar Sci Technol* 16(2):139–148

- Chik Z, Aljanabi QA, Kasa A, Taha MR (2014) Tenfold cross validation artificial neural network modeling of the settlement behavior of a stone column under a highway embankment. *Arab J Geosci* 7(11): 4877–4887. <https://doi.org/10.1007/s12517-013-1128-6>
- Coduto DP (2003) *Geotechnical engineering, principles and practice*. Prentice-Hall, New Delhi, pp 137–155
- Cortes C, Vapnik V (1995) Support vector networks. *Mach Learn* 20(3): 273–297
- Das SK, Basudhar PK (2008) Prediction of residual friction angle of clays using artificial neural network. *Eng Geol* 100(3–4):142–145
- Elgamal AW, Dobry R, Adalier K (1989) Small-scale shaking table tests of saturated layered sand-silt deposits, 2nd U.S-Japan Workshop on soil liquefaction, Buffalo, N.Y., NCEER Rep. No. 890032, 233–245
- Erzin Y, Ecemis N (2015) The use of neural networks for CPT-based liquefaction screening. *Bull Eng Geol Environ* 74(1):103–116
- Ghanadzadeh H, Ganji M, Fallahi S (2012) Mathematical model of liquid–liquid equilibrium for a ternary system using the GMDH-type neural network and genetic algorithm. *Appl Math Model* 36: 4096–4105
- Goh ATC (1994) Seismic liquefaction potential assessed by neural networks. *J Geotech Eng* 120(9):1467–1480
- Goh ATC (1996) Neural network modeling of CPT seismic liquefaction data. *J Geotech Eng* 122(1):70–73
- Goh ATC (2002) Probabilistic neural network for evaluating seismic liquefaction potential. *Can Geotech J* 39:219–232
- Goharzaya M, Noorzada A, Ardakania AM, Jalal M (2017) A worldwide SPT-based soil liquefaction triggering analysis utilizing gene expression programming and Bayesian probabilistic method. *J Rock Mech Geotech Eng* 9(4):683–693
- Hanna AM, Ural D, Saygili G (2007) Neural network model for liquefaction potential in soil deposits using Turkey and Taiwan earthquake data. *Soil Dyn Earthq Eng* 27(6):521–540
- Hassanlourad M, Ardakani A, Kordnaeij A, Mola-Abasi H (2017) Dry unit weight of compacted soils prediction using GMDH-type neural network. *Eur Phys J Plus* 132:357
- Haykin S (1994) *Neural network: a comprehensive foundation*. MacMillan College Publishing Co, New York
- Hoang ND, Bui DT (2018) Predicting earthquake-induced soil liquefaction based on a hybridization of kernel Fisher discriminant analysis and a least squares support vector machine: a multi-dataset study. *Bull Eng Geol Environ* 77(1):191–204
- Husmand B, Scott F, Crouse CB (1988) Centrifuge liquefaction tests in a laminar box. *Geotechnique* 38(2):253–262
- Idriss IM, Boulanger RW (2006) Semi-empirical procedures for evaluating liquefaction potential during earthquakes. *Int J Soil Dyn Earthquake Eng* 26:115–130
- Idriss IM, Boulanger RW (2010) SPT-based liquefaction triggering procedures. Rep. UCD/CGM-10/02, Dept. of Civil and Environmental Engineering, Univ. of California, Davis, CA
- Ishihara K (1996) *Soil behaviour in earthquake geotechnics*. The Oxford Engineering Science Series, Oxford
- Ivakhnenko AG (1971) Polynomial theory of complex systems. *IEEE Trans Syst Man Cybern A Syst Hum* 1:364–378. <https://doi.org/10.1109/TSMC.1971.4308320>
- Ivakhnenko AG (1976) The group method of data handling in prediction problems. *Sov Autom Control Avtomatika* 9:21–30
- Iwasaki T, Tokida K, Tatsuoka F (1981) Soil liquefaction potential evaluation with use of the simplified procedure. *International Conference on Recent Advances in Geotechnical Earthquake Engineering and Soil Dynamics*, St. Louis, pp 209–214
- Jirdehi RA, Mamoudan HT, Sarkaleh HH (2014) Applying GMDH-type neural network and particle swarm optimization for prediction of liquefaction induced lateral displacements. *Appl Appl Math Int J* 9(2):528–540
- Juang CH, Chen CJ (1999) CPT-based liquefaction evaluation using artificial neural networks. *Comput Aided Civil Infrastruct Eng* 14(3): 221–229
- Kalinli A, Acar MC, Gunduz Z (2011) New approaches to determine the ultimate bearing capacity of shallow foundations based on artificial neural networks and ant colony optimization. *Eng Geol* 117(1–2): 29–38. <https://doi.org/10.1016/j.enggeo.2010.10.002>
- Karhikeyan J, Kim D, Aiyer BG, Samui P (2013) SPT-based liquefaction potential assessment by relevance vector machine approach. *Eur J Environ Civ Eng* 17(4):248–262. <https://doi.org/10.1080/19648189.2013.781546>
- Kaya Y (2013) A new intelligent classifier for breast cancer diagnosis based on a rough set and extreme learning machine: RS+ ELM. *Turk J Electr Eng Comput Sci* 21(Sup. 1):2079–2091
- Kiefa MAA (1998) General regression neural networks for driven piles in cohesionless soils. *Geotech Geoenviron Eng* 124(12):1177–1185
- Kim YS, Kim BT (2006) Use of artificial neural networks in the prediction of liquefaction resistance of sands. *J. Geotech. Geoenviron. Eng. ASCE* 132(11):1502–1504. <https://doi.org/10.1061/ASCE1090-02412006132:111502>
- Kondo T, Ueno J (2012) Feedback GMDH-type neural network and its application to medical image analysis of liver cancer. In *42th ISCIE international symposium on stochastic systems theory and its applications*, pages 81–82
- Kordnaeij A, Kalantary F, Kordtabar B, Mola-Abasi H (2015) Prediction of recompression index using GMDH-type neural network based on geotechnical soil properties. *Soils Found* 55(6):1335–1345
- Kramer SL (1996) *Geotechnical earthquake engineering*. Prentice-Hall, Upper Saddle River, p 653
- Kramer SL, Mayfield RT (2007) The return period of soil liquefaction. *J Geotech Geoenviron Eng* 133(7):802–813
- Kuo YL, Jaksa MB, Lyamin AV, Kaggwa WS (2009) ANN-based model for predicting the bearing capacity of strip footing on multi-layered cohesive soil. *Comput Geotech* 36(3):503–516. <https://doi.org/10.1016/j.compgeo.2008.07.002>
- Lambe PC (1981) *Dynamic centrifuge modelling of a horizontal sand stratum*, ScD Thesis, Dept. Of Civil Engineering, Massachusetts Institute of Technology, Cambridge, USA
- Le Cessie S, Van Houwelingen JC (1992) Ridge estimators in logistic regression. *J R Stat Soc: Ser C: Appl Stat* 41(1):191–201
- Lee I, Lee J (1996) Prediction of pile bearing capacity using artificial neural networks. *Comput Geotech* 18(3):189–200. [https://doi.org/10.1016/0266-352X\(95\)00027-8](https://doi.org/10.1016/0266-352X(95)00027-8)
- Lee C, Lee GG (2006) Information gain and divergence-based feature selection for machine learning-based text categorization. *Inf Process Manag* 42(1):155–165
- Liu H, Qiao T (1984) Liquefaction potential of saturated sand deposits underlying foundation of structure, *Proceeding of 8th World Conference on Earthquake Engineering*, San Francisco, 3, 199–206
- Muduli PK, Das SK (2015a) Model uncertainty of SPT-based method for evaluation of seismic soil liquefaction potential using multi-gene genetic programming. *Soils Found* 55(2):258–275
- Muduli PK, Das SK (2015b) First order reliability method for probabilistic evaluation of liquefaction potential of soil using genetic programming. *Int J Geomech ASCE* 15(3):04014052
- Mughieda O, Bani HK, Safieh B (2009) Liquefaction assessment by artificial neural networks based on CPT. *Int J Geotech Eng* 2:289–302
- Nejad FP, Jaksa MB, Kakhi M, McCabe BA (2009) Prediction of pile settlement using artificial neural networks based on standard penetration test data. *Comput Geotech* 36(7):1125–1133
- Pal M, Mather PM (2003) An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sens Environ* 86: 554–565
- Rahman MS, Wang J (2002) Fuzzy neural network models for liquefaction prediction. *Soil Dyn Earthq Eng* 22:685–694

- Ramakrishnan D, Singh TN, Purwar N, Badre KS, Gulati A, Gupta S (2008) Artificial neural network and liquefaction susceptibility assessment: a case study using the 2001 Bhuj earthquake data, Gujarat, India. *Comput Geosci* 12:491–501
- Robertson PK, Wride CE (1998) Evaluating cyclic liquefaction potential using the cone penetration test. *Can Geotech J* 35(3):442–459
- Sakellariou MG, Ferentinou M (2005) A study of slope stability prediction using neural networks. *Geotech Geol Eng* 24(3):419–445
- Samui P, Sitharam TG (2011) Machine learning modelling for predicting soil liquefaction susceptibility. *Nat Hazards Earth Syst Sci* 11:1–9
- Seed HB, Idriss IM (1971) Simplified procedure for evaluating soil liquefaction potential. *Journal of Soil Mech Foundation Div ASCE* 97(9):1249–1273
- Stokoe KH, Roesset JM, Bierschwale JG, Aouad M (1988) Liquefaction potential of sands from shear wave velocity. *Proceedings of Ninth World Conference on Earthquake Engineering, Tokyo, Japan, 3*, 213–218.
- Sulewska MJ (2011) Applying artificial neural networks for analysis of geotechnical problems. *Comput Assist Mech Eng Sci* 18:231–241
- Suzuki Y, Koyamada K, Tokimatsu K (1997) Prediction of liquefaction resistance based on CPT tip resistance and sleeve friction. *Proceedings XIV International Conference of Soil Mechanics and Foundation Engineering, Hamburg, Germany*, 603–606
- Tokimatsu K, Yoshimi Y (1983) Empirical correlation of soil liquefaction based on SPT N-value and fines content. *Soils Found* 23(4):56–74
- Vissikirsky VA, Stepashko VS, Kalavrouziotis IK, Drakatos PA (2005) Growth dynamics of trees irrigated with wastewater: GMDH modeling, assessment, and control issues. *Instrum Sci Technol* 33(2):229–249
- Wang HB, Xu WY, Xu RC (2005) Slope stability evaluation using back propagation neural networks. *Eng Geol* 80:302–315
- Xue X, Liu E (2017) Seismic liquefaction potential assessed by neural networks. *Environ Earth Sci* 76:192. <https://doi.org/10.1007/s12665-017-6523-y>
- Xue X, Xiao M (2016) Application of genetic algorithm-based support vector machines for prediction of soil liquefaction. *Environ Earth Sci* 75:874. <https://doi.org/10.1007/s12665-016-5673-7>
- Youd TL, Perkins DM (1978) Mapping liquefaction-induced ground failure potential. *J Geotech Eng Div* 104(4):443–446
- Youd TL, Idriss IM, Andrus RD, Arango I, Castro G, Christian JT, Dobry R, Liam Finn WD, Harder LF Jr, Hynes ME, Ishihara K, Koester JP, Laio SSC, Marcuson WF III, Martin GR, Mitchell JK, Moriwaki Y, Power MS, Robertson PK, Seed RB, Stokoe KH (2001) Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils. *J Geotech Geoenviron Eng* 127(10):817–833
- Zhu W, Wang J, Zhang W, Sun D (2012) Short-term effects of air pollution on lower respiratory diseases and forecasting by the group method of data handling. *Atmos Environ* 51:29–38