



A novel ensemble model based on GMDH-type neural network for the prediction of CPT-based soil liquefaction

T. Fikret Kurnaz¹ · Yilmaz Kaya²

Received: 15 August 2018 / Accepted: 23 May 2019 / Published online: 30 May 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

This study presents a novel ensemble group method of data handling (EGMDH) model based on classification for the prediction of liquefaction potential of soils. Liquefaction is one of the most complex problems in geotechnical earthquake engineering. The database used in this study consists of 212 CPT-based field records from eight major earthquakes. The input parameters are selected as cone tip resistance, total and effective stress, penetration depth, max peak horizontal acceleration and earthquake magnitude for the prediction models. The proposed EGMDH model results were also compared to the other classifier models, particularly the results of the group method of data handling (GMDH) model. The results of this study indicated that the proposed EGMDH model has achieved more successful results on the prediction of the liquefaction potential of soils compared to the other classifier models by improving the prediction performance of the GMDH model.

Keywords Liquefaction · Soft computing · Group method of data handling · Ensemble model

Introduction

One of the most complex problems in soils that can be seen during and/or after an earthquake and whose results can be extremely damaging is known as liquefaction. Liquefaction is defined as the transformation of a granular material from a solid to a liquefied state as a consequence of increased pore-water pressure and reduced effective stress (Marcuson 1978). According to this definition, the shear resistance of the soil disappears at a certain stage and it starts to act like a liquid. Due to the large problems encountered during the soil liquefaction especially of the 27th of March 1964 Alaska and the 16th of June 1964 Niigata earthquakes, followed by the 1971 San Fernando, 1976 Tangshan, 1985 Mexico city, 1989 Loma Prieta, 1994 Kobe and 1999 Golcuk (Turkey) earthquakes (Xue and Xiao 2016; Xue and Liu 2017; Chern et al. 2008; Erzin and Ecemis 2015; Mughieda et al. 2009), many researchers have turned to observe the conditions that affect

the liquefaction phenomenon. For this reason, the 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). As a result of researches on liquefaction, useful 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 in laboratory by cyclic tri-axial, cyclic shear, shaking table tests, the standard penetration test (SPT), the cone penetration test (CPT), and seismic experiments in the field [see Kramer (1996), Ishihara (1996), Liu 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) and Andrus and Stokoe (2000)]. However, since the laboratory tests are time-consuming and expensive, methods in which SPT and CPT data are used are preferred. For many years, the methods based on SPT in the assessment of liquefaction were preferred by geotechnical engineers, but especially in the last 20 years, CPT-based methods have become widespread due

✉ T. Fikret Kurnaz
fkurnaz@mersin.edu.tr

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

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

to their properties of being fast, continuous, and accurate soil parameter measurements. Although many methods based on CPT (Robertson and Campanella 1985; Seed and De Alba 1986; Shibata and Teparaska 1988; Andrus and Youd 1989; Stark and Olson 1995; Robertson and Wride 1998; Juang et al. 2003; Idriss and Boulanger 2004; Kokusho et al. 2005) have been suggested for the evaluation of liquefaction, the most widely accepted of these is the one proposed by Robertson and Wride (1998). However, Robertson (2009) has developed an unified approach for interpretation of cone penetration tests. In addition, CPT-based soil behavior type (SBT) classification systems were updated in Robertson (2016) to use behavior based descriptions.

Recently, soft computing methods, especially the artificial neural networks (ANNs), have become popular in the practical solutions of geotechnical engineering problems such as the 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). Additionally, the liquefaction potential of soils has been tried to be predicted 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; Erzin and Ecemis 2015; Xue and Xiao 2016; Xue and Liu 2017; Hoang and Bui 2018). Goh (1994, 1996, 2002) suggested different ANN models to predict the liquefaction potential of soils based on actual field records using SPT, CPT and shear wave velocity data. Juang and Chen (1999) evaluated the liquefaction resistance of sandy soils with various ANN models using historical database including CPT measurements. Rahman and Wang (2002) developed fuzzy neural network models for the evaluation of liquefaction potential with SPT-based large databases of liquefaction case histories. Baziar and Nilipour (2003) used ANN with back propagation algorithm to determine the occurrence of liquefaction in different sites based on CPT results. 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. Chern et al. (2008) developed a fuzzy-neural network model to assess the liquefaction potential of soils with CPT field database including actual liquefaction records from more than 11 major earthquakes between 1964 and 1999. Mughieda et al. (2009) suggested three ANN models to evaluate the liquefaction potential of soils based on CPT data. Samui and Sitharam (2011) proposed two machine learning methods, ANN and SVM, to predict the liquefaction

susceptibility of soils based on the SPT data from the 1999 Chi–Chi, Taiwan earthquake. Erzin and Ecemis (2015) proposed different ANN models to predict the cone penetration and liquefaction resistance. Xue and Xiao (2016) proposed two techniques, hybrid genetic algorithm (GA) and support vector machine (SVM), to predict the liquefaction potential of soils with CPT-based field data from major earthquakes between 1964 and 1983. Xue and Liu (2017) proposed two optimization techniques, genetic algorithm (GA) and particle swarm optimization (PSO), to improve the neural network model performance on predicting the liquefaction susceptibility of soils with CPT-based field data from major earthquakes between 1964 and 1983. Hoang and Bui (2018) proposed a novel soft computing model named KFDA-LSSVM, (combines kernel Fisher discriminant analysis with a least squares SVM) to evaluate the earthquake-induced soil liquefaction. They used three 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 ANN. The GMDH model, which is a self-organizing, machine learning method, was first proposed by Ivakhnenko (1971, 1976). The GMDH creates an optimum network by trying a lot of networks in different architectures depending on the number of input variables during the self-organization. Recently, the GMDH method has begun to be applied in some geotechnical problems (Kordnaeij et al. 2015; Ardakani and Kordnaeij 2019; Hassanlourad et al. 2017; Jirdehi et al. 2014). In this regard, a novel ensemble GMDH (EGMDH) model 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 CPT-based field data from eight major earthquakes between 1964 and 1989. The results of the proposed EGMDH model were also compared with other classifier models such as ANN, GMDH, SVM, logistic regression (LR) and random forest (RF).

Group method of data handling (GMDH)

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, 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 2019):

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

where y_i is dependent variable and x_i is independent variable. Thus, it is possible to estimate the output value \hat{y} using the input vector $X = (X_1, X_2, X_3, \dots, X_n)$. The prediction equation can be written as:

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

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:

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

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 2019; Zhu et al. 2012):

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

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:

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

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 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, it is tried to estimate the output variables

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 $L = m(m-1)/2$ number of regression polynomial equations are obtained. 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 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 $G_i(x_i, x_j)$ is expected to be the smallest:

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

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:

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

Here, p and q are the any two variables that come into the layer. The Eq. (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:

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

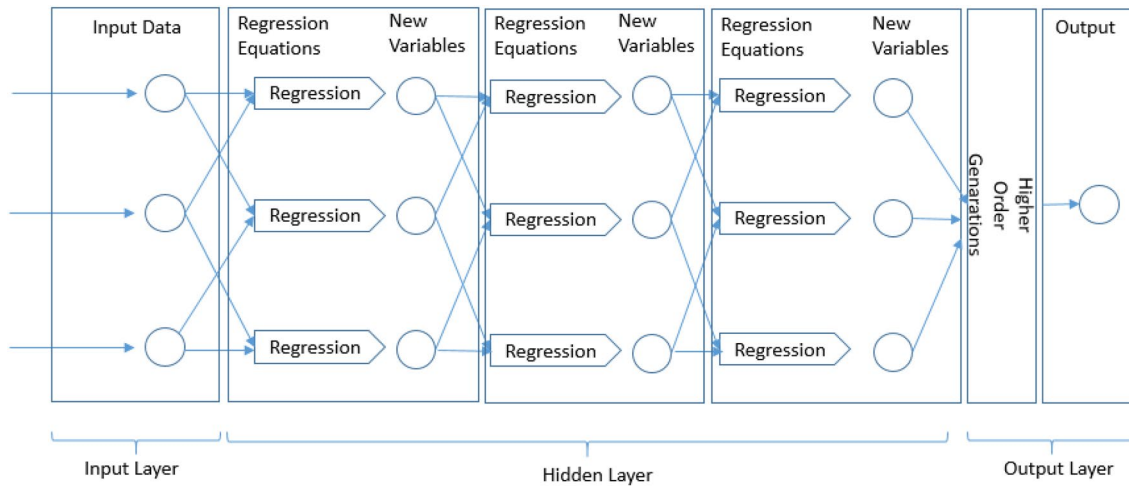


Fig. 1 GMDH network architecture

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

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

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.

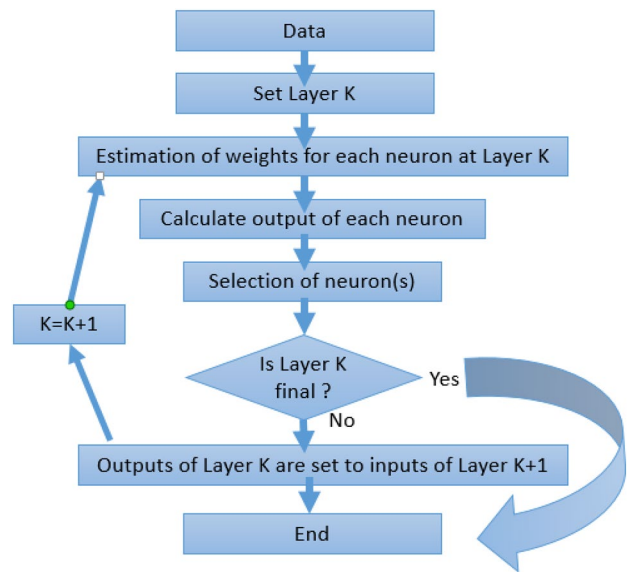


Fig. 2 The flowchart for the GMDH algorithm

Ensemble GMDH model

The main goal in the 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 result of combining is better when each classifier is used singularly. For this reason, 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 the ensemble classification is to produce a result by combining the values previously obtained by different classifiers. During this process, it is tried

to make a calculation 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. One of the most important advantages of the method is that it can get better values by combining the single models (Augusty and Izudheen 2013).

In this study, the GMDH has been ensemble 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:

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

$$\text{Radial basis} = e^{-y^2}, \tag{14}$$

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

$$\text{Tangent} = \tan h(y), \tag{16}$$

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

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

Performance criteria

Accuracy, precision, recall and *F*-criterion were used to demonstrate the performance of the methods proposed in the study. These success criteria are calculated as follows:

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

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

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

$$F\text{-criterion} = 2(\text{recall} \times \text{precision}) / (\text{recall} + \text{precision}). \tag{21}$$

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 is defined as the ratio of the positive labeled samples 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

It is known that liquefaction does not occur on all soil layers in the field. Therefore, it is necessary to first examine whether the conditions necessary for the liquefaction to occur are present in the analysis of the liquefaction hazard. 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

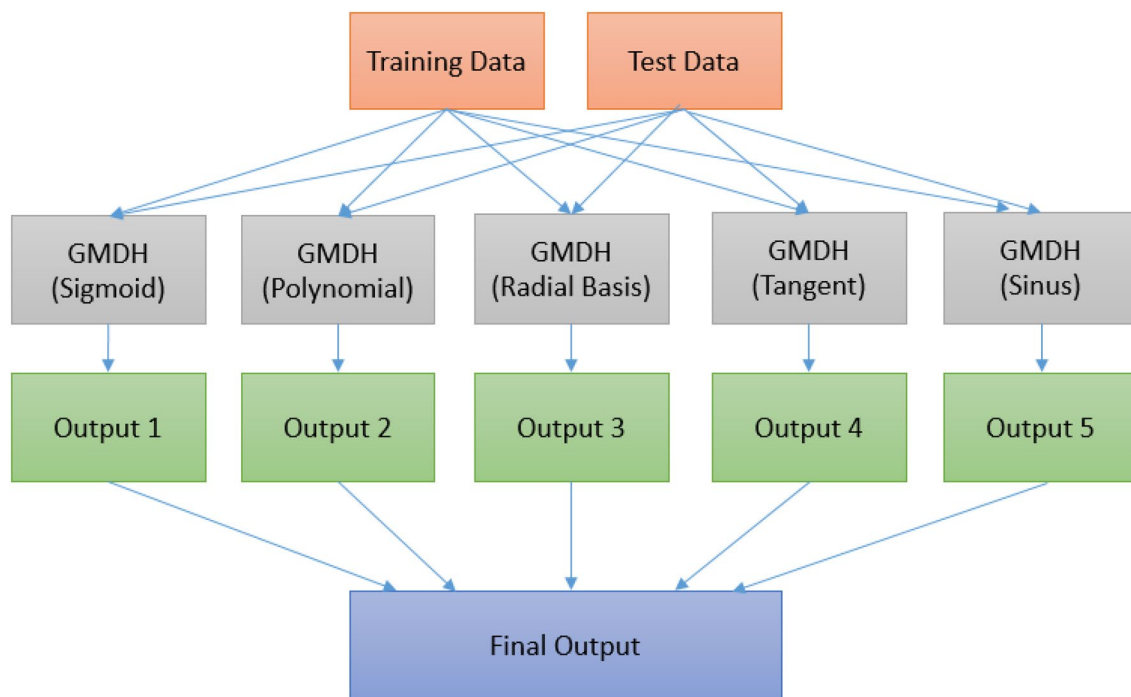


Fig. 3 The algorithm of EGMDH model

epicentral distance. In general, loose sandy soils that are saturated with water are more sensitive to liquefaction during large earthquakes [see Kramer (1996) and Coduto (2003)]. Liquefaction can only occur if all affecting factors such as loose soil, water saturation, large and long-term earthquake magnitude, etc., are present at the same time.

In this study, the database belongs to the CPT-based liquefaction assessment, which was preferred for predicting the presence of liquefaction by the EGMDH model. In this context, 212 CPT-based field data from 8 major earthquakes between 1964 and 1989 (Chern et al. 2008) were used. Chern et al. (2008) have made a liquefaction estimation study based on CPT using fuzzy- neural network by 466 CPT-based case records from major earthquakes between 1964 and 1999 in their work. Among these records, the CPT tests results from the sites where relevant earthquakes occurred are included. During the data selection for the current study, firstly the data of earthquakes with magnitudes $M > 6.0$ were preferred, among 466 data, due to the higher probability of liquefaction occurrence. It is important to achieve high success with a small number of data in the prediction of some engineering parameters of soils with artificial intelligence methods. As mentioned in the introduction section, Xue and Liu (2017) proposed optimization techniques for the prediction of liquefaction susceptibility of soils with CPT-based field data. They used only 166 of 466 case records published in Chern et al. (2008) and obtained successful results in their study. Similarly, it was aimed to use approximately half of the total data published in Chern et al. (2008) in the current study and only 212 case records were selected among the 466 field data for the CPT-based liquefaction prediction model. The case records used in this study belong to the 1964 Niigata earthquake ($M=7.5$), 1971 San Fernando Valley earthquake ($M=6.4$), 1975 Haicheng earthquake ($M=7.3$), 1976 Tangshan earthquake ($M=7.8$), 1977 Vrancea earthquake ($M=7.2$), 1979 Imperial Valley earthquake ($M=6.6$), 1983 Nihonkai-Cho earthquake ($M=7.7$) and 1989 Loma Prieta earthquake ($M=7.1$). 80 of the records were classified as non-liquefied and 132 of them as liquefied.

The input parameters used in the EGMDH model are the effective stress (σ'_{vo}), total stress (σ_{vo}), cone tip resistance (q_c), penetration depth (d), maximum peak ground acceleration (a_{max}) and magnitude of earthquake (M_w) while the output is the occurrence of liquefaction. The selected input parameters are the necessary parameters in CPT-based liquefaction assessments. In addition, these parameters were used as input parameters in most of the studies related to the CPT-based liquefaction prediction with different artificial intelligence models such as Chern et al. (2008), Goh (1996, 2002), Xue and Xiao (2016), Xue

and Liu (2017), Erzin and Ecemis (2015), Juang and Chen (1999) and Baziar and Nilipour (2003), etc.

Conventional CPT-based liquefaction assessment

The determination of the liquefaction resistance of soils based on analysis results requires calculation or estimation of two variables. 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 by Seed and Idriss (1971) as in Eq. (22):

$$CSR = 0.65 \times \frac{a_{max}}{g} \times \frac{\sigma_v}{\sigma'_v} \times r_d, \quad (22)$$

Here is, a_{max} , the peak horizontal acceleration at the ground surface during the earthquake; g , gravitational acceleration; σ_v and σ'_v total and effective stress; r_d , stress reduction coefficient. The average values are used for the r_d depending on the depth in the Eq. (23) in engineering applications (Liao and Whitman 1986):

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

Here, (z) is the depth (in meter).

The rate of cyclic resistance (CRR) is also defined by Robertson and Wride (1998) as below.

In this method, it is necessary to make some adjustments in the use of the cone penetration tip resistance corrected for overburden stress (q_{c1N}) when determining the liquefaction resistance. The parameter q_{c1} is normalized as shown in the following equations using the top layer load correction coefficient (C_Q):

$$q_{c1N} = \left(\frac{q_c}{P_{a2}} \right) C_Q = \frac{q_{c1}}{P_{a2}}, \quad (24)$$

$$C_Q = (P_a / \sigma'_{vo})^n. \quad (25)$$

Here; where q_c is the measured cone tip penetration resistance C_Q is the top layer load correction coefficient for cone penetration resistance, P_a is the reference pressure in the same units as σ'_{vo} (i.e., $P_a = 100$ kPa if σ'_{vo} in kPa), and P_{a2} is a reference pressure in the same units as q_c (i.e., $P_{a2} = 0.1$ MPa if q_c is in MPa), n is coefficient varying with soil type and typically equal to 0.5. A maximum value of $C_Q = 2$ is generally applied to CPT data at shallow depths. The normalized cone penetration resistance, q_{c1N} , is dimensionless.

It was highlighted in Robertson and Wride (1998) that the CPT friction ratio (ratio of the CPT sleeve friction to the cone tip resistance) increases with increasing fines content and soil plasticity. The grain characteristics sandy soils can be estimated directly from CPT data using any of these soil behavior charts as shown in Fig. 4 (Robertson 1990). Using the CPT chart by Robertson (1990), the soil behavior type index, I_c , can be defined as follows:

$$I_c = [(3.47 - Q)^2 + (\log F + 1.22)^2]^{0.5}, \tag{26}$$

$$Q = \left(\frac{q_c - \sigma_{vo}}{P_{a2}} \right) \left(\frac{P_a}{\sigma'_{vo}} \right)^n, \tag{27}$$

$$F = [f_s / (q_c - \sigma_{vo})] \cdot 100 (\%), \tag{28}$$

where Q =normalized tip resistance, F =normalized friction ratio, the exponent n is typically equal to 1.0, σ_{vo} and σ'_{vo} are the total and effective overburden stresses, respectively. The boundaries of soil behavior type are given in terms of the index, I_c , as shown in Table 1.

The proposed equation to obtain the equivalent clean sand normalized CPT penetration resistance, $(q_{c1N})_{cs}$, is a function of both the measured penetration resistance, q_{c1N} , and the grain characteristics of the soil, as follows:

$$(q_{c1N})_{cs} = K_c q_{c1N}. \tag{29}$$

Here, (K_c) is the correction coefficient which expresses the granular structure of the soil and determined using Eqs. (30) and (31) as follows:

$$\text{If } I_c \leq 1.64, \quad K_c = 1.0, \tag{30}$$

$$\text{If } I_c > 1.64 \quad K_c = -0.403I_c^4 + 5.581. \tag{31}$$

$$I_c^3 - 21.63. \quad I_c^2 + 33.75. \quad I_c - 17.88.$$

If $I_c > 2.6$, the data should be plotted directly on the Robertson chart (and assume $q_{c1N} = Q$). However, if $I_c \leq 2.6$, the exponent to calculate Q should be changed to $n=0.5$ (i.e., essentially calculate q_{c1N} using Eq. (24), since $\sigma_{vo} < q_c$) and I_c should be recalculated based on q_{c1N} and F . If the recalculated I_c remains less than 2.6, the data should be plotted on the Robertson chart using q_{c1N} based on $n=0.5$. If, however, I_c iterates above and below a value of 2.6, depending on which value of n is used, a value of $n=0.75$ should be selected to calculate q_{c1N} [using Eq. (24)] and plot data on the Robertson chart.

The Eqs. (26), (29), (30) and (31) can be combined to estimate the equivalent clean sand normalized penetration resistance, $(q_{c1N})_{cs}$, directly from the measured CPT data. Then, using the equivalent clean sand normalized penetration resistance $(q_{c1N})_{cs}$, the CRR (for $M=7.5$) can be estimated using the following simplified equation:

$$\text{If } 50 \leq (q_{c1N})_{cs} < 160, \quad \text{CRR} = 93 \left[\frac{(q_{c1N})_{cs}}{1000} \right]^3 + 0.08, \tag{32}$$

$$\text{If } (q_{c1N})_{cs} < 50, \quad \text{CRR} = 0.833 \left[\frac{(q_{c1N})_{cs}}{1000} \right]^3 + 0.05. \tag{33}$$

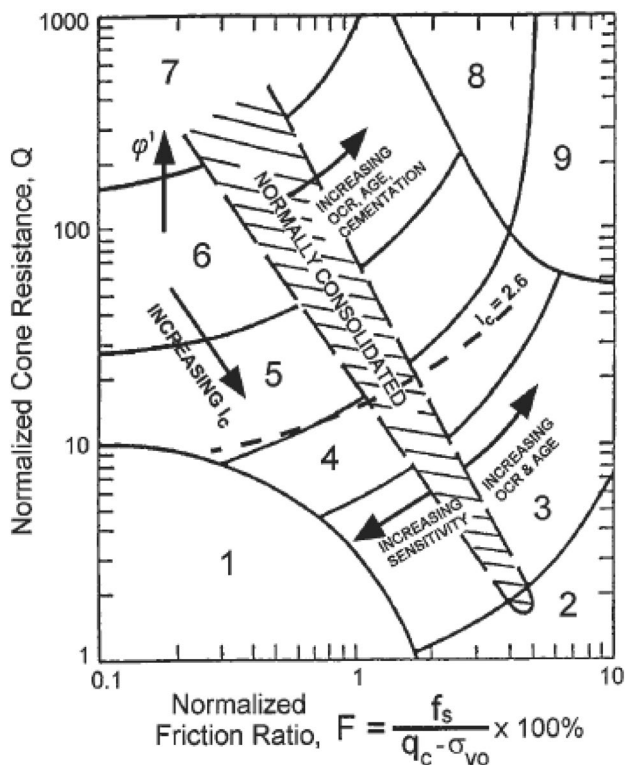


Fig. 4 Normalized CPT soil behavior type chart, as proposed by Robertson (1990)

Results

GMDH model

In this study, the occurrence of liquefaction in soils has been tried to estimate firstly using the GMDH algorithm.

Table 1 Boundaries of soil behavior type (Robertson 1990)

Soil behavior type index, I_c	Zone	Soil behaviour type (see Fig. 4)
$I_c < 1.31$	7	Gravelly sand to dense sand
$1.31 < I_c < 2.05$	6	Sands: clean sand to silty sand
$2.05 < I_c < 2.60$	5	Sand mixtures: silty sand to sandy silt
$2.60 < I_c < 2.95$	4	Silt mixtures: clayey silt to silty clay
$2.95 < I_c < 3.60$	3	Clays: silty clay to clay
$I_c > 3.60$	2	Organic soils: peats

GMDH is a nonlinear regression method, but is also model that carries the characteristics of supervised and unsupervised ANNs. Regression is a statistical model that examines the cause-and-effect relationship between independent variables and dependent variables. The linear regression model is modeling the relationship between one or more independent variables and dependent variables.

The input and output variables in all models were normalized with the following equation in this study:

$$A_{\text{normalized}} = \frac{A_{\text{actual}} - A_{\text{min}}}{A_{\text{max}} - A_{\text{min}}} \tag{34}$$

Here, $A_{\text{normalized}}$ represents the normalized values of A , A_{actual} expresses the actual values of the variable, A_{max} and A_{min} express the largest and smallest values in the variables. 132 samples classified as 1 (liquefied) while 80 samples classified as 0 (non-liquefied) in the database of 212 samples. Trials with GMDH model were conducted according to different training—test set ratios. The performance results are given in Table 2.

Due to no criteria in 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 on Table 1 that the trials were conducted for data sets with different ratios in the form of 50–50%, 60–40%, 70–30% and 80–20% training-test. The highest success was achieved as 95.08% for the 70–30% training-test data set. The success rate for the 80–20% data set was slightly lower, although the success rate seemed to increase as the training data rate increased. 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 that the success rate is 100% for output “1” (liquefied class). All of the 40 test samples seem to be correctly classified. However, since 18 of the 21 samples are correctly classified, the overall success rate was observed to be 85.70% for output “0” (non- liquefied class).

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 3. Since the number of input variables is low, the trials have been performed for the hidden

Table 2 Success rates of GMDH model according to different training-test ratios

Train-test rate	Accuracy (%)	Precision	Recall	F-measure
50–50%	85.44	0.8478	0.8444	0.8460
60–40%	89.02	0.9093	0.8686	0.8809
70–30%	95.08	0.9286	0.9651	0.9435
80–20%	95.00	0.9000	0.9688	0.9283

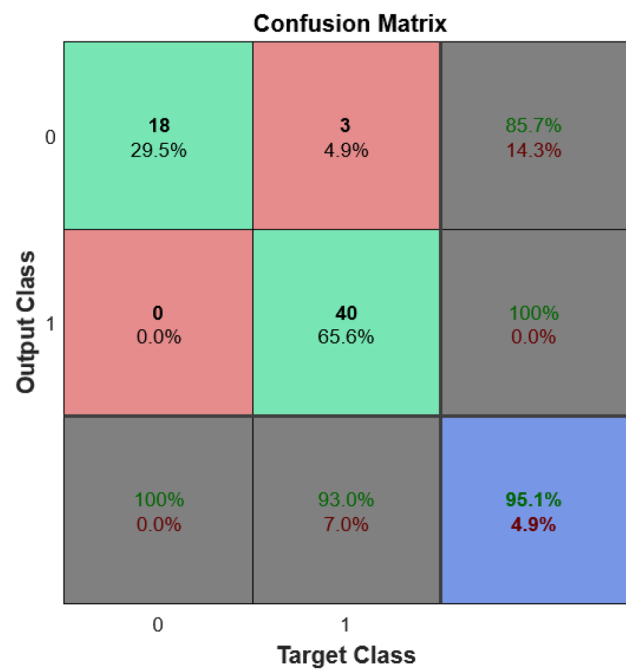


Fig. 5 Confusion matrix by GMDH for 70–30% training-test data

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 3).

The performance measures obtained as a result of trials using different numbers of neurons in the hidden layers of a three-layered GMDH model are given Table 4. The performance for three or more neurons was unchanged in the hidden layers. The best success rate was 95.08% (Table 4).

Table 3 GMDH success rates for different hidden layer numbers with 70–30% training-test set (with 10 neurons)

#Hidden layer	Accuracy	Precision	Recall	F-measure
1	90.16	0.8750	0.9302	0.8911
2	91.80	0.8913	0.9419	0.9082
3	95.08	0.9286	0.9651	0.9435

Table 4 GMDH success rates of different neurons in hidden layers with 70–30% training-test set (with 3 hidden layer)

#Neurons	Accuracy	Precision	Recall	F-measure
1	90.16	0.8750	0.9302	0.8911
3	95.08	0.9286	0.9651	0.9435
5	95.08	0.9286	0.9651	0.9435
10	95.08	0.9286	0.9651	0.9435
15	95.08	0.9286	0.9651	0.9435

Table 5 The test set of 70–30% training-test data set and estimated values with GMDH

No	M_w	d (m)	σ (kPa)	σ' (kPa)	q_c (MPa)	a_{max} (g)	Actual	Predicted by GMDH
1	7.7	3.1	56.9	47.1	9.81	0.23	0	0
2	7.7	3.8	71.6	53	15.69	0.23	0	0
3	7.7	5	94.1	63.7	15.08	0.23	0	0
4	7.7	2.8	53	45.1	1.76	0.23	1	1
5	7.7	3.4	62.8	51	4.02	0.23	1	1
6	7.7	5.1	94.1	65.7	7.8	0.23	1	1
7	7.7	6	111.8	73.5	8.8	0.23	1	0
8	7.1	5.9	115.1	83.1	14.5	0.24	0	0
9	7.1	5.3	103.4	77.4	15.2	0.24	0	0
10	7.1	5.9	115.1	85.1	1.3	0.24	1	1
11	7.1	2.8	54.6	50.6	1.9	0.24	1	1
12	7.1	8.4	172.2	143.2	5.9	0.24	0	0
13	7.1	3	57	45	3.8	0.25	1	1
14	7.1	3	57	45	3	0.25	1	1
15	7.1	4.4	83.6	65.6	5.4	0.25	1	1
16	7.1	4	76	70	16.4	0.25	0	0
17	7.1	2.5	47.5	40.5	7.7	0.25	1	1
18	7.1	9.5	180.5	103.5	25	0.25	0	0
19	7.1	1.4	26.6	26.6	3	0.25	1	1
20	7.1	3.9	74.1	53.1	11.7	0.25	0	0
21	7.1	8	152	90	20	0.25	0	0
22	7.1	1.9	36.1	34.1	10.4	0.25	0	0
23	7.1	5	95	63	9	0.25	1	0
24	7.1	6.8	129.2	79.2	20.8	0.25	0	0
25	7.1	6.5	123.5	75.5	18.2	0.25	0	0
26	7.1	3.5	66.5	49.5	15.5	0.25	0	0
27	7.1	4.1	77.9	62.9	13	0.25	0	0
28	7.1	2.5	47.5	40.5	10.4	0.25	0	0
29	7.1	2.1	39.9	33.9	2.5	0.25	1	1
30	7.1	2.6	49.4	40.4	10	0.25	0	0
31	7.1	4.1	77.9	55.9	6.2	0.25	1	1
32	7.1	4.1	77.9	55.9	4.3	0.25	1	1
33	7.1	4.7	89.3	72.3	4.1	0.25	1	1
34	7.1	4.2	79.8	64.8	4.9	0.25	1	1
35	7.1	2.9	55.1	38.1	6.6	0.25	1	1
36	7.1	2	38	28	3.1	0.25	1	1
37	7.1	2.2	41.8	29.8	3.1	0.25	1	1
38	7.1	2.9	55.1	39.1	5.1	0.25	1	1
39	7.1	3.6	68.4	44.4	7.8	0.25	1	1
40	7.1	8.3	157.7	88.7	4.3	0.25	1	1
41	7.1	8.6	163.4	90.4	4.3	0.25	1	1
42	7.1	9.8	186.2	110.2	3.8	0.25	1	1
43	7.1	4.4	83.6	54.6	8.2	0.25	1	1
44	7.1	5.5	104.5	77.5	1.9	0.25	1	1
45	7.1	6	114	79.7	2.2	0.16	1	1
46	7.1	6	114	79.7	2.1	0.16	1	1
47	7.1	6.3	119.7	77.5	4.1	0.28	1	1
48	7.1	7	133	84	8.5	0.28	1	0
49	7.1	6	114	84.6	6.4	0.28	1	1
50	7.1	9.5	180.5	116.8	12.1	0.28	0	0

Table 5 (continued)

No	M_w	d (m)	σ (kPa)	σ' (kPa)	q_c (MPa)	a_{max} (g)	Actual	Predicted by GMDH
51	7.1	2.5	47.5	42.6	3.4	0.27	1	1
52	7.1	3.3	62.7	50	6.2	0.27	1	1
53	7.1	5.7	82.4	73.4	2.2	0.53	1	1
54	7.1	7.3	114.7	86.4	4.5	0.53	1	1
55	7.1	6	99.8	79	3.9	0.53	1	1
56	7.1	2.1	28.4	22.5	1	0.21	1	1
57	7.1	3.3	57.5	32.5	1	0.21	1	1
58	7.1	2.1	32.7	27.1	1.3	0.21	1	1
59	7.1	2.7	42.1	33.1	1	0.21	1	1
60	7.1	7.3	116.3	96.9	4	0.49	1	1
61	7.1	6.8	128.3	118	3.7	0.53	1	1

Table 6 EGMDH model success rates for different numbers of the hidden layers with 70–30% training-test set (with 10 neurons)

#Hidden layer	Accuracy	Precision	Recall	<i>F</i> -measure
1	93.44	0.9091	0.9535	0.9256
2	95.08	0.9286	0.9651	0.9435
3	98.36	0.9737	0.9884	0.9806

Table 5. It is observed that there are three incorrect estimates for samples 7, 23 and 48.

EGMDH model

In this study, 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 6.

It is seen that the EGMDH model is more successful than the GMDH model on the estimation of soil liquefaction when Table 5 is examined. A high classification success rate as 98.36% was obtained with EGMDH. The output confusion matrix for the EGMDH model is given in Fig. 6. As it can be seen in Fig. 6, only one sample of “0” (non-liquefied) state is misclassified. All samples representing the state “1” (liquefied) are correctly classified. The input–output variables of the test samples for the 70–30% training-test set and the output variables estimated by EGMDH are given in Table 7.

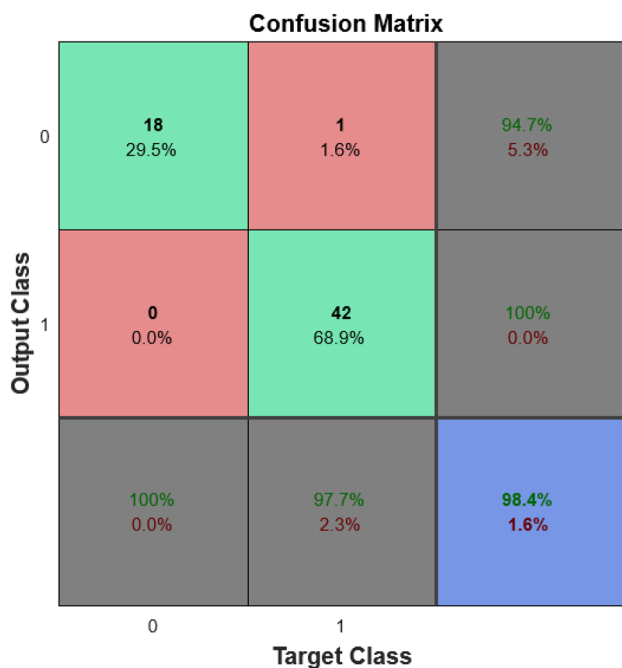


Fig. 6 Confusion Matrix by EGMDH for 70–30% training-test data

The input–output variables of the test samples for the 70–30% training-test set are given in Table 5. Additionally, the output variables estimated by the GMDH are shown in

Table 7 Test set of 70–30% training-test data set and estimated values with EGMDH

No	M_w	d (m)	σ (kPa)	σ' (kPa)	q_c (MPa)	a_{\max} (g)	Actual	Predicted by EGMDH
1	7.7	3.1	56.9	47.1	9.81	0.23	0	0
2	7.7	3.8	71.6	53	15.69	0.23	0	0
3	7.7	5	94.1	63.7	15.08	0.23	0	0
4	7.7	2.8	53	45.1	1.76	0.23	1	1
5	7.7	3.4	62.8	51	4.02	0.23	1	1
6	7.7	5.1	94.1	65.7	7.8	0.23	1	1
7	7.7	6	111.8	73.5	8.8	0.23	1	0
8	7.1	5.9	115.1	83.1	14.5	0.24	0	0
9	7.1	5.3	103.4	77.4	15.2	0.24	0	0
10	7.1	5.9	115.1	85.1	1.3	0.24	1	1
11	7.1	2.8	54.6	50.6	1.9	0.24	1	1
12	7.1	8.4	172.2	143.2	5.9	0.24	0	0
13	7.1	3	57	45	3.8	0.25	1	1
14	7.1	3	57	45	3	0.25	1	1
15	7.1	4.4	83.6	65.6	5.4	0.25	1	1
16	7.1	4	76	70	16.4	0.25	0	0
17	7.1	2.5	47.5	40.5	7.7	0.25	1	1
18	7.1	9.5	180.5	103.5	25	0.25	0	0
19	7.1	1.4	26.6	26.6	3	0.25	1	1
20	7.1	3.9	74.1	53.1	11.7	0.25	0	0
21	7.1	8	152	90	20	0.25	0	0
22	7.1	1.9	36.1	34.1	10.4	0.25	0	0
23	7.1	5	95	63	9	0.25	1	1
24	7.1	6.8	129.2	79.2	20.8	0.25	0	0
25	7.1	6.5	123.5	75.5	18.2	0.25	0	0
26	7.1	3.5	66.5	49.5	15.5	0.25	0	0
27	7.1	4.1	77.9	62.9	13	0.25	0	0
28	7.1	2.5	47.5	40.5	10.4	0.25	0	0
29	7.1	2.1	39.9	33.9	2.5	0.25	1	1
30	7.1	2.6	49.4	40.4	10	0.25	0	0
31	7.1	4.1	77.9	55.9	6.2	0.25	1	1
32	7.1	4.1	77.9	55.9	4.3	0.25	1	1
33	7.1	4.7	89.3	72.3	4.1	0.25	1	1
34	7.1	4.2	79.8	64.8	4.9	0.25	1	1
35	7.1	2.9	55.1	38.1	6.6	0.25	1	1
36	7.1	2	38	28	3.1	0.25	1	1
37	7.1	2.2	41.8	29.8	3.1	0.25	1	1
38	7.1	2.9	55.1	39.1	5.1	0.25	1	1
39	7.1	3.6	68.4	44.4	7.8	0.25	1	1
40	7.1	8.3	157.7	88.7	4.3	0.25	1	1
41	7.1	8.6	163.4	90.4	4.3	0.25	1	1
42	7.1	9.8	186.2	110.2	3.8	0.25	1	1
43	7.1	4.4	83.6	54.6	8.2	0.25	1	1
44	7.1	5.5	104.5	77.5	1.9	0.25	1	1
45	7.1	6	114	79.7	2.2	0.16	1	1
46	7.1	6	114	79.7	2.1	0.16	1	1
47	7.1	6.3	119.7	77.5	4.1	0.28	1	1
48	7.1	7	133	84	8.5	0.28	1	1
49	7.1	6	114	84.6	6.4	0.28	1	1
50	7.1	9.5	180.5	116.8	12.1	0.28	0	0

Table 7 (continued)

No	M_w	d (m)	σ (kPa)	σ' (kPa)	q_c (MPa)	a_{max} (g)	Actual	Predicted by EGMDH
51	7.1	2.5	47.5	42.6	3.4	0.27	1	1
52	7.1	3.3	62.7	50	6.2	0.27	1	1
53	7.1	5.7	82.4	73.4	2.2	0.53	1	1
54	7.1	7.3	114.7	86.4	4.5	0.53	1	1
55	7.1	6	99.8	79	3.9	0.53	1	1
56	7.1	2.1	28.4	22.5	1	0.21	1	1
57	7.1	3.3	57.5	32.5	1	0.21	1	1
58	7.1	2.1	32.7	27.1	1.3	0.21	1	1
59	7.1	2.7	42.1	33.1	1	0.21	1	1
60	7.1	7.3	116.3	96.9	4	0.49	1	1
61	7.1	6.8	128.3	118	3.7	0.53	1	1

Discussion and conclusions

Due to the occurrence during the earthquake and many factors influencing it, the liquefaction is one of the most complex soil problems. Therefore, the determination or prediction of the liquefaction potential of soils has a very great importance. In this study, it was aimed to develop a novel prediction model for the liquefaction potential of soils using the ensemble group method of data handling (EGMDH) algorithm based on the GMDH model. Totally 212 CPT-based field records obtained from 8 major earthquakes were used for this study. Recently, the GMDH model has been used in many geotechnical problems with high success.

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.

In this study, the success rate of the liquefaction prediction achieved with the classical GMDH model was 95.08%,

while it increased to 98.36% 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 8. It is obvious that the worst performance was obtained with the SVM model and the most successful performance belongs to the EGMDH model proposed in this study. Despite the fact that there are many studies in literature on the prediction of liquefaction with different artificial intelligence techniques, as also mentioned in this study, the authors believe that new models for the predicting of the liquefaction phenomena will continue to be developed just as the EGMDH model, proposed in this study.

Acknowledgements This academic work was linguistically supported by the Mersin Technology Transfer Office Academic Writing Center of Mersin University.

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
- Andrus RD, Stokoe KH II (2000) Liquefaction resistance of soils from shear wave velocity. *J Geotech Geoenviron Eng* 126(11):1015–1025
- Andrus RD, Youd TL (1989) Penetration test in liquefiable gravels. In: *Proceedings of the 12th international conference on soil mechanics and foundation engineering*, Rotterdam, the Netherlands, pp 679–682
- Ardakani A, Kordnaeij A (2019) Soil compaction parameters prediction using GMDH-type neural network and genetic algorithm. *Eur J Environ Civ Eng* 23(4):449–462
- 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 Bioinform Biom* 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

Table 8 Comparison of EGMDH with other classifier models

Model	Accuracy	Precision	Recall	F -measure
ANN	93.860	0.939	0.939	0.939
SVM	75.000	0.800	0.750	0.714
LR	87.730	0.878	0.877	0.876
RF	92.920	0.929	0.929	0.929
GMDH	95.080	0.928	0.965	0.943
EGMDH	98.360	0.973	0.988	0.980

- 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
- 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, pp 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
- 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
- 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 (2004) Semi-empirical procedures for evaluating liquefaction potential during earthquakes. In: *Proceedings of the 11th international conference on soil dynamics and earthquake engineering and 3rd international conference on earthquake geotechnical engineering*, Berkeley, California, pp 32–56
- Ishihara K (1996) *Soil behaviour in earthquake geotechnics*. Oxford University Press, The Oxford Engineering Science Series, Oxford. ISBN 10:0198562241, ISBN 13: 978-0198562245
- Ivakhnenko AG (1971) Polynomial theory of complex systems. *IEEE Trans Syst Man Cybern Part 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. In: *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 Civ Infrastruct Eng* 14(3):221–229
- Juang CH, Yuan H, Lee DH, Lin PS (2003) Simplified cone penetration test-based method for evaluating liquefaction resistance of soils. *J Geotech Geoenviron Eng* 129(1):66–80
- 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>
- 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>
- Kokusho T, Hara T, Murahata K (2005) Liquefaction strength of fines containing sands compared with cone resistance in triaxial specimens. In: *Pre-workshop proceedings of the 2nd Japan–US workshop on testing, modelling, and simulation in geomechanics*, Campus Plaza, Kyoto, Japan, pp 280–296
- 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*, pp 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 civil engineering and engineering mechanics
- 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, Kagwa 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, Department Of Civil Engineering, Massachusetts Institute of Technology, Cambridge, USA
- 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)
- Liao SSC, Whitman RV (1986) Overburden correction factors for SPT in sand. *J Geotech Eng ASCE* 112(3):373–377
- Liu H, Qiao T (1984) Liquefaction potential of saturated sand deposits underlying foundation of structure. In: *Proceeding of 8th world conference on earthquake engineering*, San Francisco, 3, pp 199–206
- Marcuson WF III (1978) Definition of terms related to liquefaction. *J Geotech Eng Div ASCE* 104(9):1197–1200
- 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 (1990) Soil classification using the cone penetration test. *Can Geotech J* 27(1):151–158

- Robertson PK (2009) Interpretation of cone penetration tests—a unified approach. *Can Geotech J* 46(11):1337–1355
- Robertson PK (2016) Cone penetration test (CPT)-based soil behaviour type (SBT) classification system—an update. *Can Geotech J* 53(12):1910–1927
- Robertson PK, Campanella RG (1985) Liquefaction potential of sands using the CPT. *J Geotech Eng* 111(3):384–403
- 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, De Alba P (1986) Use of SPT and CPT tests for evaluating the liquefaction resistance of sands. In: *Proceedings of the insitu*, ASCE, New York, pp 281–302
- Seed HB, Idriss IM (1971) Simplified procedure for evaluating soil liquefaction potential. *J Soil Mech Found Div ASCE* 97(9):1249–1273
- Shibata T, Teaparaska W (1988) Evaluation of liquefaction potentials of soils using cone penetration tests. *Soils Found* 28(2):49–60
- Stark TD, Olson SM (1995) Liquefaction resistance using CPT and field case histories. *J Geotech Eng* 121(12):856–869
- Stokoe KH, Roesset JM, Bierschwale JG, Aouad M (1988) Liquefaction potential of sands from shear wave velocity. In: *Proceedings of ninth world conference on earthquake engineering*, Tokyo, Japan, vol 3, pp 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. In: *Proceedings XIV international conference of soil mechanics and foundation engineering*, Hamburg, Germany, pp 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
- 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

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.