Model. Earth Syst. Environ. (2016) 2:76
DOI 10.1007/s40808-016-0132-8

ORIGINAL ARTICLE

Comparative evaluation of artificial intelligence models for prediction of uniaxial compressive strength of travertine rocks, Case study: Azarshahr area, NW Iran

Rahim Barzegar[1] · Masoud Sattarpour[2] · Mohammad Reza Nikudel[2] · Asghar Asghari Moghaddam[1]

Received: 9 April 2016 / Accepted: 12 April 2016 / Published online: 20 April 2016

Abstract Uniaxial compressive strength (UCS) is one of the most widely used and important rock mechanical parameters in rock engineering. The main objective of the present study was to evaluate the ability of artificial intelligence models including multi-layer perceptron (MLP), Sugeno fuzzy logic (SFL), Mamdani fuzzy logic (MFL), adaptive neuro-fuzzy inference system (ANFIS) and support vector machine (SVM) to predict the UCS of travertine rocks in the Azarshahr area (NW Iran). To attempt this objective, 85 core samples of travertine rock were collected from the study area and the laboratory tests were performed to determine the P-wave velocity [V_p (km/s)], porosity (n %), Schmidt rebound hardness (R_n) and UCS of the rocks at the Rock Mechanics Laboratory in the Tarbiat Modares University. The data set including V_p (km/s), n % and R_n as the inputs and UCS as the output were divided into training (80 % of dataset) and testing (20 % of dataset) subsets to construct the models. The coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) were used to evaluate the models performance. The models accuracy followed the order SVM > ANFIS > SFL > MLP > MFL. The SVM model with RBF kernel function yielded the highest R^2 (0.9516), and the lowest RMSE (2.14 MPa) and MAE (1.351 MPa) in the testing step. Accuracy results indicated that SVM model predictions were better than MLP, SFL, MFL and ANFIS models for prediction of UCS of travertine rocks.

Keywords Artificial intelligence · Uniaxial compressive strength · Prediction · Travertine rocks · Azarshahr · Iran

Introduction

Uniaxial compressive strength (UCS) of intact rocks is an important and pertinent property for characterizing rock mass. UCS is included as a main input parameter for rock mass characterization, rock classification and failure criteria (Dehghan et al. 2010; Beiki et al. 2013). This parameter is widely used in geological, geotechnical, geophysical and petroleum engineering projects. A UCS test requires high quality core samples with regular geometry. Standard cores cannot always be extracted from weak, highly fractured, thinly bedded, foliated and/or block-in-matrix rocks (Ceyran 2014). To overcome this difficulty, various predictive models based on index tests, including mineralogical-petrographic analyses, physical properties, an elastic wave velocity test and basic mechanical tests have been developed by many researchers (Gokceoglu 2002; Sonmez et al. 2006; Dehghan et al. 2010; Yilmaz 2009; Zhang et al. 2012; Ceryan et al. 2012; Mishra and Basu 2012, 2013; Singh et al. 2012; Nefesli-oglu 2013; Yesiloglu-Gultekin et al. 2013; Ceyran 2014; Armaghani et al. 2016).

Traditionally, statistical methods such as simple and multiple regression techniques are used to found prediction models (Rzhevsky and Novick 1971; Horsrud 2001; Beiki et al. 2013). The use of these relations is often the only way to estimate strength parameters in many situations due to the absence of core for laboratory tests. The

✉ Rahim Barzegar
rm.barzegar@yahoo.com

[1] Department of Earth Sciences, Faculty of Natural Sciences, University of Tabriz, Tabriz, Iran

[2] Department of Engineering Geology, Tarbiat Modares University, Tehran, Iran

basis for these relations is the fact that many of the same factors that affect UCS also affect other physical properties. In general, these existing empirical equations give good results only in similar rocks and their reliability for other rocks should be considered with caution. In other words, most of the empirical equations are not sufficiently generic to fit all the data published on UCS and physical properties and they are not open to the general purpose use (Beiki et al. 2013).

In addition to these conventional methods, new techniques for estimating the UCS have also garnered considerable attention. Various researchers have utilized soft computing methods to estimate UCS from some rock index properties including point load, p-wave velocity and Schmidt hammer hardness (Yesiloglu-Gultekin et al. 2013; Gokceoglu 2002; Dehghan et al. 2010). Sarkar et al. (2010) conducted artificial neural network (ANN) model to predict the UCS and shear strength of different types of rocks using dynamic wave velocity, $I_{s(50)}$, slake durability index and density. Verma and Singh (2013) proposed an ANFIS model for predicting p-wave velocity and they emphasized that neuro-fuzzy method shows a good potential to model complex, nonlinear and multivariate problems. Singh and Verma (2012) performed a comparative analysis of intelligent algorithms to correlate strength and petrographic properties of some schistose rocks. Cevik et al. (2011) applied the results of laboratory experiments and a neural network to estimate the UCS of certain clay-bearing rocks. Singh et al. (2012) generalized regression neural network approach and adaptive neuro-fuzzy inference systems for prediction of UCS of rocks. Gokceoglu et al. (2009) constructed rule-based fuzzy models and multiple regression analyses to estimate the UCSs of certain clay bearing rocks. Yagiz et al. (2012) examined the effects of the cycling integer of a slake durability index test on intact rock behavior and estimated certain rock properties, such as the UCS, from rock index parameters using an ANN and regression techniques. They stated that the slake durability index (I_{d4}), p-wave velocity, density and Schmidt hammer values of rocks may be used for estimating the UCS of rocks. Armaghani et al. (2016) applied three non-linear prediction tools, namely non-linear multiple regression (NLMR), ANN and ANFIS, for estimating UCS of granitic rocks and compared their performances. They found that the prediction performance of the ANFIS models was higher than ANN and NLMR models.

The main purpose of this study is to evaluate the performance of artificial intelligence (AI) models including Multi-layer perceptron (MLP), Sugeno fuzzy logic (SFL), Mamdani fuzzy logic (MFL), Adaptive neuro-fuzzy inference system (ANFIS) and support vector machine (SVM) to predict the UCS of travertine rocks in the Azarshahr area, northwest of Iran.

Materials and methods

Multi-layer perceptron

A detailed background of ANN models can be found in Haykin (1999). However, in brief, ANNs composed of three distinct types of layers; one input layer, one or more hidden layers and one output layer. Each layer consists of a number of simple processing elements called neurons or nodes. Information data enters an ANN through the nodes of the input layer. The input layer nodes distribute the input information to the next layer (i.e., the first hidden layer). The hidden and output layer nodes process all incoming signals by applying factors to them (termed weights). Each layer also has an additional element called a bias. Bias nodes simply output a signal to the nodes of the current layer. All inputs to a node are weighted, combined and then processed through a transfer function that controls the strength of the signal released through the node's output connections. Some of the most popular transfer (activation) functions are Sigmoid, Gaussian, Hyperbolic Tangent and Hyperbolic Secant (Malekmohammadi et al. 2011).

Multi-layer perceptron (MLP) is one of the commonly used ANN approach for prediction studies. Figure 1 shows the structure of a MLP neural network model. In this Figure *i, j* and *k* denote input layer, hidden layer and output layer neurons, respectively, and *w* is the applied weight by the neuron. The explicit expression for an output value of a three layered MLP is given by Nourani et al. (2013), Barzegar and Asghari Moghaddam (2016) and Barzegar et al. (2016b):

$$y_k = f_0\left[\sum_{i=1}^{M_N} W_{kj} f_h\left(\sum_{i=1}^{N_N} W_{ji}X_i + W_{j0}\right) + W_{k0}\right] \quad (1)$$

where W_{ji} is a weight in the hidden layer connecting the *i*th neuron in the input layer and the *j*th neuron in the hidden layer, W_{j0} is the bias for the *j*th hidden neuron, f_h is the activation function of the hidden neuron, W_{kj} is a weight in the output layer connecting the *j*th neuron in the hidden layer and the *k*th neuron in the output layer, W_{k0} is the bias for the *k*th output neuron, f_0 is the activation function for the output neuron, X_i is *i*th input variable for input layer

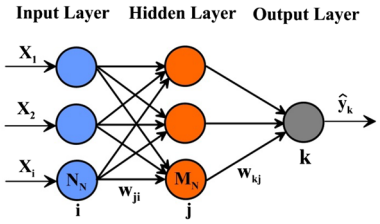

Fig. 1 Schematic diagram of a feed-forward MLP neural network

and y_j is computed output variable. NN and MN are the number of the neurons in the input and hidden layers, respectively.

Fuzzy logic

The fuzzy approach based on linguistic expressions include uncertainty rather than numerical probabilistic, statistical or perturbation approaches. Fuzzy logic (FL) theory, introduced by Zadeh (1965), is an outgrowth of classical set theory. In fuzzy set theory, the requirement of sharp boundaries of classical sets is abandoned, and the membership of an object is not a matter of either affirmation or denial, but it is in general a matter of degree (Demico and Klir 2004). Fuzzy sets also have ambiguous boundaries and gradual transitions between defined sets, which are appropriate to deal with the nature of uncertainty in system and human errors (Pulido-Calvo and Gutiérrez-Estrada 2009). The degree of membership in fuzzy set is expressed in a closed unit interval [0, 1]. The exact values of 0 and 1 represent the total denial and affirmation of the membership, respectively (Ghiasi-Freez et al. 2012).

The main part of a fuzzy model is the fuzzy inference system (FIS) in which a given inputs is formulated to an output. Fuzzy logic could be constructed by the Mamdani and Sugeno methods (Mamdani 1976; Sugeno 1985). The most important differences among fuzzy inference systems are the types of the output membership functions and the implication methods. In Mamdani model the output membership functions are fuzzy sets. After the aggregation process, there is a fuzzy set for each output variable that needs defuzzification (Barzegar et al. 2016b). This method uses the min operation (^) as a fuzzy implication (Mamdani and Assilian 1975; Mamdani 1976).

Let us suppose a rule base is given in the following form:

R^i: If x is A_i and y is B_i then z is C_i, i = 1, 2,…, n

Then, $R^i = (A_i \wedge B_i) \rightarrow C_i$ is defined by:

$$\mu_{R^i} = \mu_{(A_i \text{ and } B_i \rightarrow C_i)}(x, y, z) \tag{2}$$

The input data $x = x_0$, $y = y_0$ pass through the rule, above, to produce the final output, as below (Lee 2004):

$$\begin{cases} \mu_{C_i'}(z) = \left[\mu_{A_i}(x_0) \wedge \mu_{B_i}(y_0)\right] \rightarrow \mu_{C_i}(z) \\ \mu_{C_i'}(z) = \alpha_i \wedge \mu_{C_i}(z) \\ \mu_{C'}(z) = \mu_{C_1}(z) \vee \mu_{C_2}(z) = \left[\alpha_1 \wedge \mu_{C_1}(z)\right] \vee \left[\alpha_2 \wedge \mu_{C_2}(z)\right] \\ \mu_{C'}(z) = \cup_{i=1}^{n}\left[\alpha_i \wedge \mu_{C_i}(z)\right] = \cup_{i=1}^{n} \mu_{C_i'}(z), C' = \overset{n}{\underset{i=1}{\Leftarrow}} C_i' \end{cases} \tag{3}$$

A graphical illustration of MFL is shown in Fig. 2a.

Sugeno fuzzy inference system is similar to the Mamdani method in many aspects. In the first two parts of the fuzzy inference process, fuzzifying the inputs and applying the fuzzy operator are exactly the same. Moreover, all the lemmas expressed for Mamdani fuzzy inference system are the same for SFIS. The main difference between them is that output membership functions are either linear or constant in the Sugeno (1985) model.

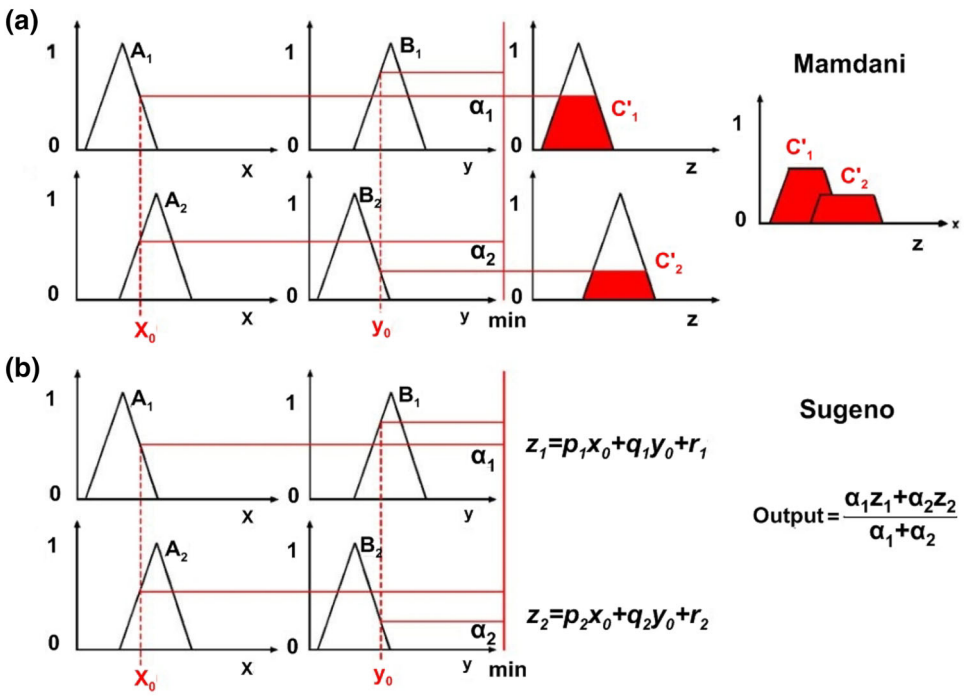

Fig. 2 Graphical illustrations of **a** MFL and **b** SFL models (Barzegar et al. 2016b)

A typical rule in a Sugeno fuzzy model has the form If Input 1 = x_0 and Input 2 = Y_0, then $z = p_{x_0} + q_{y_0} + r$.

For a zero-order Sugeno model, the output level is a constant $p = q = 0$. The output level, z_i, of each rule, is weighted by the firing strength α_i of the rule. The final output of the system is the weighted average of all rule outputs, computed as follows:

$$\text{Final output} = \sum_{i=1}^{n} \alpha_i z_i \Big/ \sum_{i=1}^{n} \alpha_i \tag{4}$$

A graphical illustration of SFL is shown in Fig. 2b.

Clustering of available data is essential for developing a fuzzy modeling system. Fuzzy c-means (FCM) and subtractive clustering (SC) are two powerful fuzzy clustering techniques, which could be used for the construction of Mamdani and Sugeno models, respectively (Ghavidel and Montaseri 2014). Each of the clusters refers to a membership function for generating the fuzzy ''if–then'' rules. The fuzzy system makes a sum of all 'then' parts and uses a defuzzification method to give the final output (Ghiasi-Freez et al. 2012).

Adaptive neuro-fuzzy inference system

The adaptive neuro fuzzy inference system (ANFIS) was introduced by Jang (1993) as a neural network functionally equivalent to a Sugeno type inference model. ANFIS uses a feed-forward network to search for fuzzy decision rules that perform well on a given task. Using a given input–output data set, ANFIS creates an FIS for which membership function parameters are adjusted using either a back propagation algorithm alone or a combination of a back propagation algorithm and a least-squares method (Abdulshahed et al. 2015). This allows the fuzzy systems to learn from the data being modeled.

The equivalent ANFIS architecture of the Sugeno inference system is shown in Fig. 3. The entire system consists of five layers, and the relationship between the input and output of each layer is summarized as follows:

Layer 1: Every node i in this layer is an adaptive node with a node output, O, defined by:

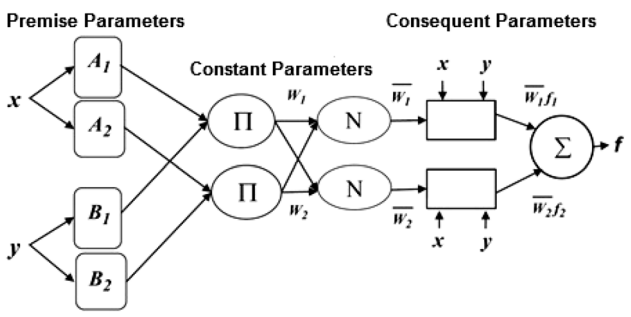

Fig. 3 A typical ANFIS architecture (Jang 1993)

$$\begin{aligned} O_{1,i} &= \mu_{A_i}(x) \quad \text{for } i = 1, 2, \text{or} \\ O_{1,i} &= \mu_{B_{i-2}}(y) \quad \text{for } i = 3, 4 \end{aligned} \tag{5}$$

where, x (or y) is the input to the node, and A_i (or B_{i-2}) is a fuzzy set associated with this node, and characterized by the shape of the node's membership function (μ).

This function must be continuous and piecewise differentiable, such as, for example, a Gaussian function. If such is used as a membership function, $\mu_{A_i}(x)$ can be computed as:

$$\mu_{A_i}(x) = e^{-\frac{1}{2}\left(\frac{x-C_i}{\sigma_i}\right)^2} \tag{6}$$

where, {σi, ci} are parameter sets.

Parameters in this layer are referred to as premise (antecedent) parameters.

Layer 2: Every node in this layer is a fixed node labeled as Π, which multiplies the incoming signals and output product. For instance,

$$O_{2,i} = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y) \quad i = 1, 2 \tag{7}$$

with each output node representing the firing strength of a rule.

Layer 3: Every node in this layer is a circular node labeled as N. The ith node calculates the ratio of the ith rule's firing strength to the sum of all rules' firing strengths.

$$O_{3,i} = \overline{w} = \frac{w_i}{w_1 + w_2} \quad i = 1, 2 \tag{8}$$

This layer's outputs are termed normalized firing strengths.

Layer 4: Node i in this layer computes the contribution of the ith rule towards the model output, with the following node function:

$$O_{4,i} = \overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \tag{9}$$

where, w_i is the output of layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set. Parameters in this layer are referred to as consequent parameters.

Layer 5: The single node in this layer is a fixed node, labeled P, that computes the overall output as the summation of all incoming signals.

$$O_{5,i} = \sum_{i=1}^{i=2} \overline{w_i} f_i = \frac{\sum_{i=1}^{i=2} w_i f_i}{\sum_{i=1}^{i=2} w_i} \tag{10}$$

Support vector machine

Support vector machine (SVM) is one of the soft computing learning algorithms and is primarily used in pattern recognition, prediction, classification, and regression analysis and its application exhibited better performance than the earlier developed methodologies, e.g. neural

network and other conventional statistical models (Vapnik et al. 1996; Joachims 1998; Huang et al. 2002; Shamshirband et al. 2015). The details of theory and evolution of SVM developed by Vapnik can be found in (Vapnik and Vapnik 1998; Vapnik 2000).

SVM was developed according to the statistical machine learning development as well as structural risk minimization to reduce the upper bound generalization error compared to local training error, which is a common technique in the previously used machine learning methodologies.

SVM functions according to Vapnik's theory are represented in Eqs. (11–15). $R = \{x_i, d_i\}$in is used to assume a set of data points. x_i indicates the input space vector of the data sample. Also, d_i and n are the target value and data size, respectively. SVM approximates the function as represented in Eqs. (11) and (12):

$$f(x) = w\varphi(x) + b \tag{11}$$

$$R_{SVMs}(C) = \frac{1}{2}\|w\|^2 + C\frac{1}{n}\sum_{i=1}^{n} L(x_i, d_i) \tag{12}$$

In Eq. (11),$\varphi(x)$ indicates high dimensional space characteristic that mapped the input space vector x. Also, w and b are a normal vector and scalar, respectively. In addition, $C\frac{1}{n}\sum_{i=1}^{n} L(x_i, d_i)$ stands error or risk. Factors b and w are measured by minimization of regularized risk equation following by introduction of positive slack variables ξ_i and ξ_i* that indicate upper and lower excess deviation:

$$\text{Minimize} \quad R_{SVMs}\left(w, \xi^{(*)}\right)\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i + \xi_i^*\right) \tag{13}$$

$$\text{Subject to} \begin{cases} d_i - w\varphi(x_i) + b_i \le \varepsilon + \xi_i \\ w\varphi(x_i) + b_i - d_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, i = 1, \ldots, l \end{cases} \tag{14}$$

where $\frac{1}{2}\|w\|^2$ is the regularization term, C represents the error penalty feature utilized to control the trade-off between the empirical error (risk) and regularization term, ε represents the loss function associated to approximation accuracy of the trained data point and the number of factors in the training data set which is defined as the l.

Optimality constraints and Lagrange multiplier which can be used to solve Eq. (11) are consequently obtained using a generic function as follows:

$$f\left(x, \beta_i, \beta_i^*\right) = \sum_{i=1}^{n}\left(\beta_i - \beta_i^*\right)K(x, x_i) + b \tag{15}$$

In Eq. (15), $K(x, x_i) = \varphi(x_i)\varphi\left(x_j\right)$ and the term K is defined as the kernel function, which is dependent on the two inner vector x_i and x_j in the feature space $\varphi(x_i)$ and $\varphi(x_j)$, respectively. The structure of the SVM model is shown in Fig. 4.

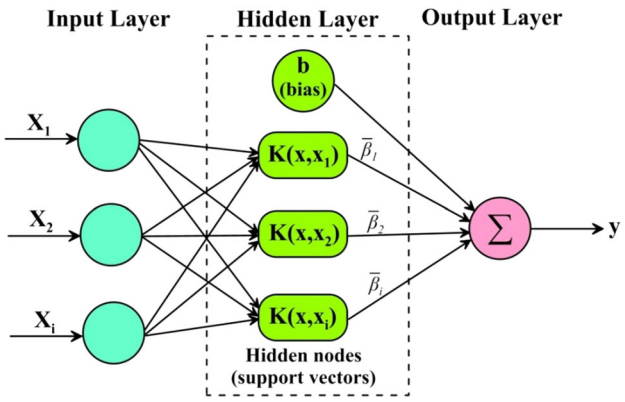

Fig. 4 The structure of the SVM model

The kernel function, denoted by K, as a straight-forward computation technique (hereafter) can be used to generate a nonlinear learning machine. The method is employed to calculate the inner product in a feature space that serve as a function to original input points. The adaptability of SVM to use kernel functions is important where it discreetly alters the information into a higher dimensional feature space. The obtained results in such a space typify the outcomes of the lower dimensional, original input space (Shamshirband et al. 2015).

Sigmoid, lineal, polynomial, and radial basis functions are the four basic kernel functions which are provided by SVM. Over time, the radial basis function (RBF) has been repeatedly proven to be the ideal function in its category due to its ability for efficient, simple, reliable, and adaptable computation for the purpose of optimization especially for adaptability in handling the parameters which are complex (Rajasekaran et al. 2008; Yang et al. 2009; Wu and Wang 2009). Only the solution of a set of linear functions are required for the training of RBF kernel equation rather than the lengthy and complicated demanding quadratic programming problem (Shamshirband et al. 2014; Mohammadi et al. 2015). Accordingly, the radial basis equation with parameter σ is adopted. The nonlinear radial basis kernel function is defined as Eq. 16:

$$K\left(x_i, x_j\right) = \exp\left(-\gamma\|x_i - x_j\|^2\right) \tag{16}$$

where x_i and x_j are vectors in the input space, i.e., vectors of features computed from training or testing samples. In addition, the accuracy of predictions using RBF kernel function depends on the selection of its three factors (γ, ε, and C).

Model performance criteria

The following statistical indicators were selected in the performance evaluation of constructed models:

1. Coefficient of determination (R^2);
2. Root Mean Square Errors (RMSE), and
3. Mean Absolute Error (MAE)

$$R^2 = \left[\sum_{i=1}^{N}(P_i - \overline{P})(O_i - \overline{O})\right]^2 \left[\sum_{i=1}^{N}(P_i - \overline{P})^2 (O_i - \overline{O})^2\right]^{-1} \quad (17)$$

$$RMSE = \left[N^{-1}\sum_{i=1}^{N}(P_i - O_i)^2\right]^{0.5} \quad (18)$$

$$MAE = \frac{|P_i - O_i|}{N} \quad (19)$$

where N is the number of observations, P_i is the predicted values, O_i is the observed data, and $\overline{P}$ and $\overline{O}$ are the mean values for P_i and O_i, respectively.

The coefficient of determination (R^2) measures the degree of correlation among the observed and predicted values. R^2 values range from 0 to 1, with 1 indicating a perfect relationship between the data and the line drawn through them, while 0 represents no statistical correlation between the data and the line. The RMSE evaluates the variance of errors independently of the sample size. RMSE indicates the discrepancy between the observed and predicted values. A perfect fit between observed and predicted values would have an RMSE of 0. MAE is the simplest measure of prediction accuracy. It is used to measure how close predicted values are to the measured values.

Study area and the data set

Fissure-ridge type travertines are exposed in extensive area in Azarshahr area and eastern part of the Urmia Lake, NW Iran. Some of the ridges are still active in the area indicating active displacement along the faults in this region. Quaternary active Sahand volcanic complex close to the travertine ridges had reasonable influence on solution of the Jurassic and Cretaceous carbonates and stained travertine layers. The location of the travertine rocks is shown in the geological map of the study area (Fig. 5). Climate changes, strongly affects the amount of water supply in travertine ridges. This is well identified by inter-bedded paleo-soil bands sub-horizontal to the travertine layers and change of dip in travertine layers (Taghipour and Mohajjel 2013). The travertine facies are deposited at the outlet of springs.

In this study, 30 travertine rock blocks from the Azarshahr area, each sample measuring approximately 40–40–20 cm, were collected in the field for the rock mechanics tests using the core-drilling machine at the Rock Mechanics Laboratory in the Tarbiat Modares University. A total of 85 core samples were prepared from the rock blocks. The core samples had a ratio of length to diameter of ≈ 2 and the edges of the cores were cut parallel and smooth. The mechanical tests on the core samples including porosity (n %), P-wave velocity [V_p (km/s)], Schmidt rebound hardness (R_n) and UCS were performed by laboratory tests in accordance with ISRM (1981). Figure 6 shows the UCS test setup and some of core samples used for laboratory tests.

The data set including V_p (km/s), n % and R_n as the inputs and UCS as the output were divided into training (80 % of dataset) and testing (20 % of dataset) subsets to develop the AI models for prediction of UCS. The cross validation technique (Fijani et al. 2013; Barzegar et al. 2016a, b) was used to divide the data set. Some descriptive statistics including minimum and maximum values, mean values, standard deviation (SD) and the range of the data utilized as well as the correlation coefficient between the UCS and the considered input variables for both training and testing data sets are listed in Table 1.

Input and output variables for AI-based modeling are usually normalized by scaling between 0 and 1, to ensure that all variables receive equal attention during the training step of a model. The following simple linear mapping of the variables is the most common method for this purpose:

$$X_N = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (20)$$

Upon completing the training of the models, normalized model outputs were denormalized to actual values:

$$X = [X_N(X_{\max} - X_{\min})] + X_{\min} \quad (21)$$

where X_N is the normalized value of variable X, and $0 \le X_N \le 1$, while $X_{\max}$ and $X_{\min}$ are the maximum and minimum value of variable X of the original data, respectively.

Model development

MLP model

For the MLP model, the training set was further divided into 80 % training and 20 % validation, so overall, 64 % of the data was used for training, 16 % was used for validation, and 20 % was used for testing. To develop a three-layered ANN model, three inputs including V_p (km/s), n % and R_n were used in the first layer and the UCS as output

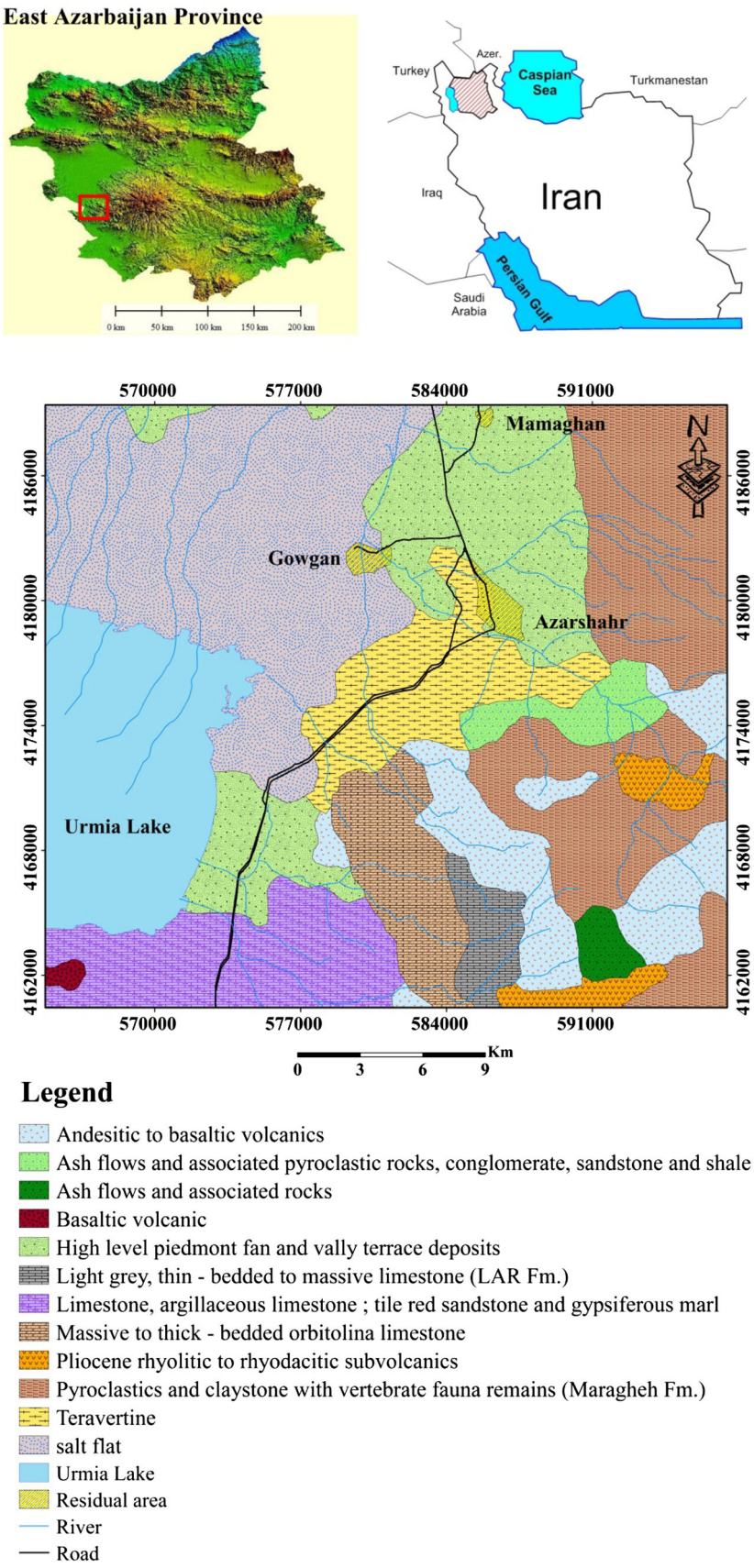

Fig. 5 Geological map of the study area

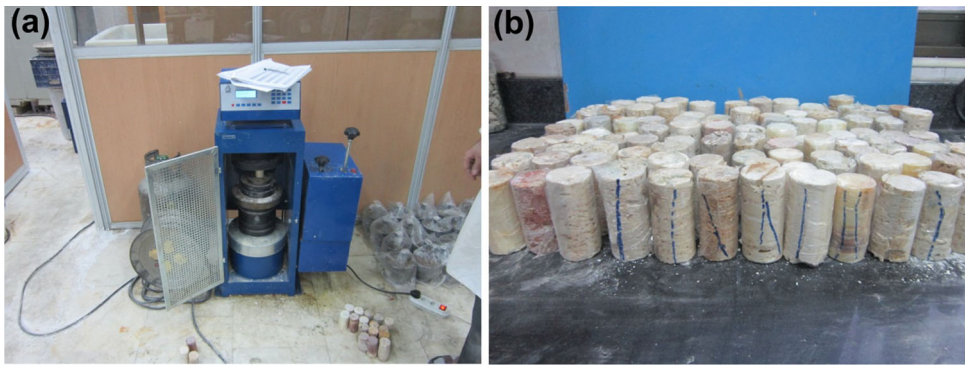

Fig. 6 **a** The uniaxial compressive strength test setup, **b** some of core samples used for laboratory tests

Table 1 Descriptive statistics of data sets utilized for training and testing

	Minimum	Maximum	Mean	SD	Range	Correlation with UCS
Training data						
V_p (km/s)	3.50	5.82	4.92	0.61	2.32	0.41
n %	0.49	9.80	4.99	3.16	9.31	−0.54
R_n	33.00	55.00	45.80	5.67	22.00	0.59
UCS (MPa)	37.50	67.50	54.90	9.73	30.00	1.00
Testing data						
V_p (km/s)	3.30	5.58	4.53	0.74	2.28	0.62
n %	0.50	9.50	5.00	2.54	9.00	−0.66
R_n	35.00	55.00	43.50	6.52	20.00	0.54
UCS (MPa)	41.70	67.80	52.10	8.92	26.10	1.00

was utilized in the last layer. The feed-forward neural network was trained with Levenberg–Marquardt algorithm (TrainLM). The number of hidden neurons for MLP models was selected via a trial and error method (Belayneh et al. 2014, 2016; Adamowski and Sun 2010; Barzegar et al. 2016a, b). However, Wanas et al. (1998) and Mishra and Desai (2006) empirically considered the equations e.g. log (N), where N is the number of training samples and $2n + 1$, where n is the number of input neurons to determine the number of hidden neurons. In this study, the optimal number of hidden neurons was determined to be between log (N) and $(2n + 1)$. By using the Wanas et al. (1998) method, two hidden neurons and by using the Mishra and Desai (2006) method, seven hidden neurons were considered; thereafter the optimal number was chosen via trial and error. The number of neurons in the hidden layer was 5. TANSIG and PURELIN functions were used as the transfer functions in the hidden layer. The performance plot (Fig. 7a) shows the value of the function, in terms of training, validation, and testing behaviors, versus the iteration number. The best validation performance was at epoch 7 based on the mean square error equally to 9.4315×10^{-3}. The magnitude of the gradient and the number of validation checks used to terminate network training are illustrated in Fig. 7b. At an epoch of 13 iterations, the gradient was 1.0112×10^{-3}, barely above the 1×10^{-4} threshold below which training will stop, and at six the validation checks also indicated training should stop. When the training of the model was completed, the testing data set served as model input and UCS values were predicted.

Fuzzy logic model

In the present study, two different types of fuzzy logic model including Sugeno Fuzzy Logic (SFL) and Mamdani Fuzzy Logic (MFL) models were used to predict USC of the travertine rocks. The subtractive clustering method was applied for classifying the input and output datasets in the SFL model. The important parameter in subtractive clustering which controls number of clusters and fuzzy if-then rules is clustering radius (Barzegar et al. 2016a, b). This parameter can take values between the range of [0, 1]. Specifying a smaller cluster radius (say 0.1) will usually yield more and smaller clusters in the data resulting in more rules. In contrast, a large cluster radius (say 0.9) yields a few large clusters in the data resulting in few rules (Chiu 1994). The optimal clustering radius was searched by performing the clustering process several times and gradually increasing the clustering radius from 0 to 1 and it was

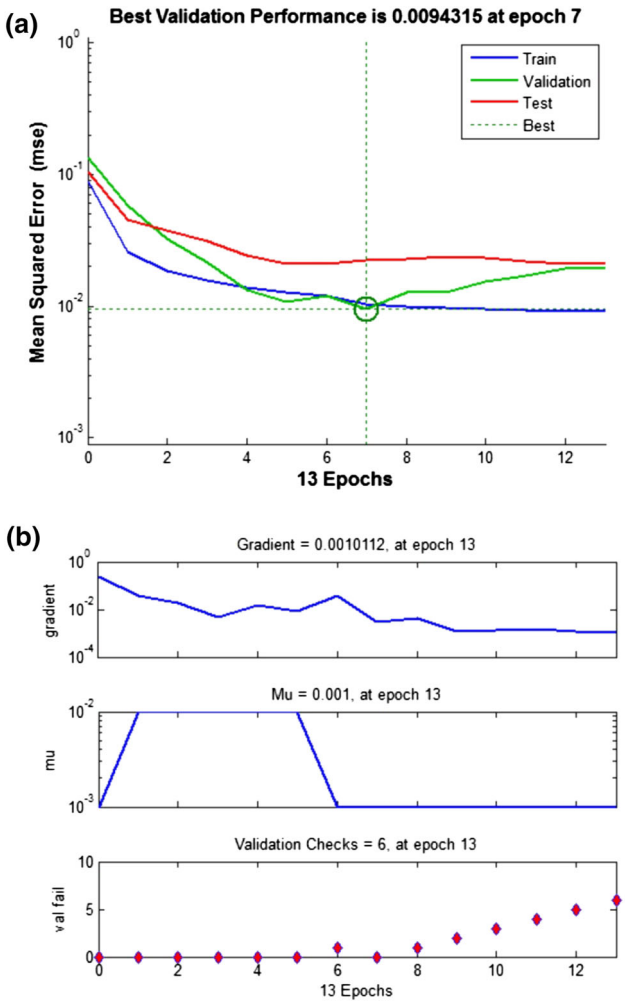

Fig. 7 Training state and performance of the developed MLP neural network model for prediction of UCS of travertine rocks

selected at the minimum RMSE. The optimal clustering radius was 0.48. Thus, the SFL model was established by eight Gaussian membership functions (clusters) for input and output data resulting in eight rules.

For MFL model, the FCM clustering method was used for extraction of clusters and fuzzy if–then rules. The MFL model was established by ten Gaussian membership functions (clusters) for input and output data resulting in ten rules.

ANFIS model

To develop the ANFIS model, hybrid algorithm which is the combination of the least-squares method and the back propagation gradient descent method was applied to optimize and adjust the Gaussian membership function parameters and coefficients of the output linear equations (Zounemat-Kermani and Teshnehlab 2008; Fijani et al. 2013). In this study, the Gaussian membership function was used because it generated the least error in the fuzzification of the data collected for the components. The number of epochs and error tolerance were set to 500 and 0, respectively. Subtractive fuzzy clustering, based on a measure of the density of data points in the feature space (Chiu 1994), was used to establish the rule-based relationship between the input and output variables. The best ANFIS model performance was achieved after 100 epochs of training when the clustering radius was set to 0.5. Four Gaussian membership functions were extracted for the input variables of the ANFIS model (Fig. 8).

SVM model

In this study, the SVM model was developed in data regression (DTREG) soft computing software. The models were created by using Epsilon-SVR kernel type. The both grid and pattern search and tenfold cross-validation resampling method were employed to find optimal parameter values. During grid search, the program (DTREG) evaluates values of each parameter within the predefined search area. On the other hand, a pattern search (also known as a line search or a compass search) starts from the center of the search area and tries steps in both directions for each parameter. The center of search area is then moved to the new point if a better model fit is obtained (Jain and Bhatia 2013; Sonebi et al. 2016). The process is repeated until the specified tolerance rate is reached. Grid search is computationally expensive because the model must be evaluated at many points within the grid for each parameter (Al-Anazi and Gates 2010).

RBF kernel function was adopted for prediction of UCS. Model parameters such as C with the search range 0.1–5000, kernel parameter γ with 0.001–50, ε (Epsilon) with 0.001–100 were selected for the developing the SVM model. The pattern search technique was selected with ten search intervals and the tolerance for stopping the iterative optimization process was considered 1e-008. During the grid or pattern searches, cross-validation was performed to evaluate the fitting of so as to avoid over fitting of the model. The optimal values of parameters for the SVM model were $C = 1.389$, $\gamma = 6.691$ and $\varepsilon = 0.001$. After training the SVM model, the testing data set was served to the model to predict the UCS of the travertine rocks.

Results

The performance of MLP, SFL, MFL, ANFIS and SVM models was compared in this part of the study. After training the proposed models, the models were tested with

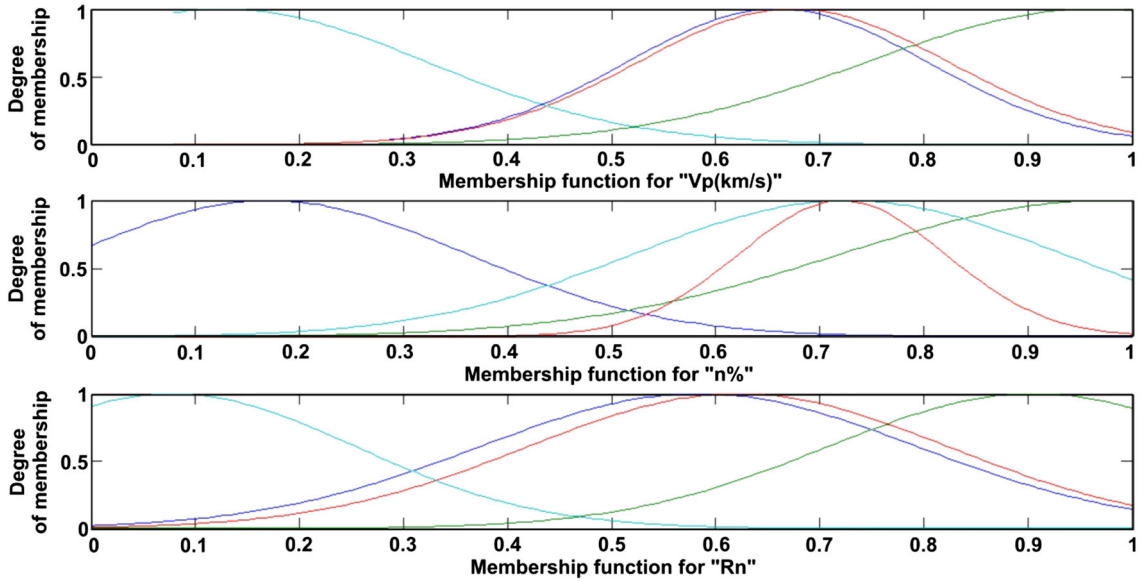

Fig. 8 Sugeno-FIS generated Gaussian membership functions for input variables of the ANFIS model

Table 2 The results of the proposed models in the training and testing steps for prediction of UCS (Mpa) of travertine rock

Model	Training step			Testing step		
	R^2	RMSE (MPa)	MAE (MPa)	R^2	RMSE (MPa)	MAE (MPa)
MLP	0.9109	2.9359	2.2024	0.838	3.5162	2.2778
SFL	0.9199	2.5067	1.8028	0.8492	3.5152	2.598
MFL	0.8042	4.349	3.2672	0.7648	4.3756	2.783
ANFIS	0.9431	2.3161	1.6298	0.9061	2.6839	1.6388
SVM	0.9888	1.0266	0.5813	0.9516	2.14	1.351

17 testing samples. The results of the developed models in the training and testing steps are presented in Table 2. All the models for the prediction of UCS revealed satisfactory results in terms of the statistical performance criteria. Therefore, these models were acceptable for prediction of UCS of travertine rocks in the Azarshahr area. The proposed models obtained relatively lower prediction errors in training step as compared to the testing step indicated that these models exhibited relatively better generalization as compared to the prediction. The R^2, RMSE, and MAE values of the MLP model in training step were 0.9109, 2.9359 and 2.2024 MPa, respectively, whereas those were 0.838, 3.5162 and 2.2778 MPa, in the testing step. In the training step, the SFL model resulted in the R^2 of 0.9199, RMSE of 2.5067 MPa, and MAE of 1.8028 MPa, whereas, for the testing data, the corresponding values were 0.8492, 3.5152 and 2.598 MPa, respectively. The SFL model performance was slightly better than MLP model, whereas, the accuracy of the MFL model was less than MLP model.

The R^2, RMSE, and MAE values of the ANFIS model in training step were 0.9431, 2.3161 and 1.6298 MPa, respectively, whereas those were 0.9061, 2.6839 and 1.6388 MPa, in the testing step. Generally, the ANFIS model performed better than the MLP model in prediction of UCS values based on performance criteria. This result is concur with the studies of Barzegar et al. (2016a, b), Rajaee et al. (2009), Adamowski and Chan (2011), Nourani et al. (2011), Moosavi et al. (2013), Fijani et al. (2013), Emamgholizadeh et al. (2014) and Parmar and Bhardwaj (2015). This may be related to the effect of fuzzification of the input through membership functions (Barzegar et al. 2016a, b).

The R^2, RMSE, and MAE values of the ANFIS model in training step were 0.9888, 1.0266 and 0.5813 MPa. The SVM model with RBF kernel function provided the best results in the testing step showing the highest R^2 (0.9516), and the lowest RMSE (2.14 MPa) and MAE (1.351 MPa). Measured and predicted values of UCS by MLP, SFL, MFL, ANFIS and SVM models in the testing step are compared in Fig. 9. It was concluded that the SVM model outperformed the other developed models and this result was related to dimensional independence,

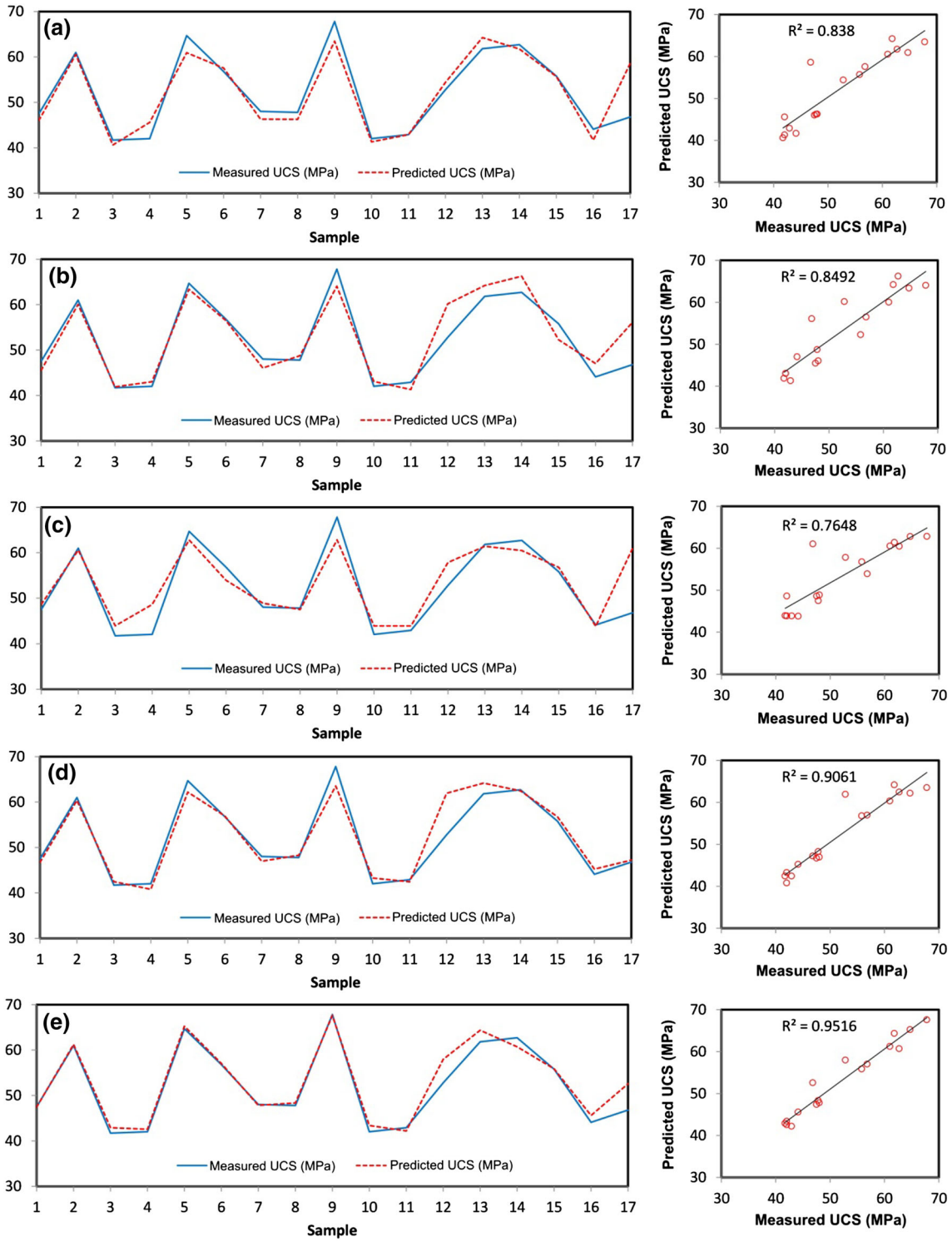

Fig. 9 Measured and predicted UCS (MPa) for travertine rocks using **a** MLP, **b** SFL, **c** MFL **d** ANFIS and **e** SVM models in the testing step

limited number of freedom, global optimum and higher generalization capability of the SVM (Scholkopf and Smola 2002; Leong et al. 2010). The models accuracy followed the order SVM > ANFIS > SFL > MLP > MFL.

Conclusions

In the present study, five AI models including Multi-layer Perceptron (MLP), Sugeno Fuzzy Logic (SFL), Mamdani Fuzzy Logic (MFL), Adaptive Neuro-Fuzzy Inference

System (ANFIS) and Support Vector Machine (SVM) were evaluated for prediction of UCS of travertine rocks in the Azarshahr area (NW Iran) based on V_p (Km/S), n % and R_n. The coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) were used to evaluate the models performance. The R^2, RMSE, and MAE values of the MLP model in testing step were 0.838, 3.5162 and 2.2778 MPa, respectively, whereas those were 0.9061, 2.6839 and 1.6388 MPa, respectively for the ANFIS model. The ANFIS model outperformed the MLP model in prediction of UCS values based on performance criteria. This was attributed to the effect of fuzzification of the input through membership functions. The SVM model with RBF kernel function yielded the highest R^2 (0.9516), and the lowest RMSE (2.14 MPa) and MAE (1.351 MPa) in the testing step. Accuracy results indicated that SVM model predictions were better than MLP, FL and ANFIS models. It was related to dimensional independence, limited number of freedom, global optimum and higher generalization capability of the SVM.

References

Abdulshahed AM, Longstaff AP, Fletcher S (2015) The application of ANFIS prediction models for thermal error compensation on CNC machine tools. Appl Soft Comput 27:158–168

Adamowski JF, Chan HG (2011) A wavelet neural network conjunction model for groundwater level forecasting. J Hydrol 407(1–4):28–40

Adamowski JF, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J Hydrol 390(1–2):85–91

Al-Anazi AF, Gates ID (2010) Support vector regression for porosity prediction in a heterogeneous reservoir: a comparative study. Comput Geosci 36:1494–1503

Armaghani DJ, Mohammad ET, Hajihassani M, Yagiz S, Motaghedi S (2016) Application of several non-linear prediction tools for estimating uniaxial compressive strength of granitic rocks and comparison of their performances. Eng Comput 32(2):189–206

Barzegar R, Asghari Moghaddam A (2016) Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. Model Earth Syst. doi:10.1007/s40808-015-0072-8

Barzegar R, Adamowski J, Moghaddam AA (2016a) Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran. Stoch Environ Res Risk Assess. doi:10.1007/s00477-016-1213-y

Barzegar R, Asghari Moghaddam A, Baghban H (2016b) A supervised committee machine artificial intelligent for improving DRASTIC method to assess groundwater contamination risk: a case study from Tabriz plain aquifer, Iran. Stoch Environ Res Risk Assess 30(3):883–899

Beiki M, Majdi A, Givshad AD (2013) Application of genetic programming to predict the uniaxial compressive strength and elastic modulus of carbonate rocks. Int J Rock Mech Min Sci 63:159–169

Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B (2014) Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet-support vector regression models. J Hydrol 508:418–429

Belayneh A, Adamowski J, Khalil B, Quilty J (2016) Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought. Atmos Res 172–173(15):37–47

Ceryan N, Okkan U, Kesimal A (2012) Application of generalized regression neural networks in predicting the unconfined compressive strength of carbonate rocks. Rock Mech Rock Eng 45:1055–1072

Cevik A, Sezer EA, Cabalar AF, Gokceoglu C (2011) Modelling of the uniaxial compressive strength of some clay-bearing rocks using neural network. Appl Soft Comput 11(2):2587–2594

Ceyran N (2014) Application of support vector machines and relevance vector machines in predicting uniaxial compressive strength of volcanic rocks. J Afr EarthSci 100:634–644

Chiu S (1994) Fuzzy model identification based on cluster estimation. J Intell Fuz Sys 2:267–278

Dehghan S, Sattari Gh, Chehreh Chelgani S, Aliabadi MA (2010) Prediction of uniaxial compressive strength and modulus of elasticity for Travertine samples using regression and artificial neural networks. Min Sci Technol 20:41–46

Demico RV, Klir GJ (2004) Fuzzy logic in geology. Elsevier Academic Press, London, p 347

Emamgholizadeh S, Kashi H, Marofpoor I, Zalaghi E (2014) Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. Int J Environ Sci Technol 11(3):645–656

Fijani E, Nadiri AA, Asghari Moghaddam A, Tsai F, Dixon B (2013) Optimization of DRASTIC method by supervised committee machine artificial intelligence to assess groundwater vulnerability for Maragheh-Bonab plain aquifer Iran. J Hydrol 530:89–100

Ghavidel SZ, Montaseri M (2014) Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud basin. Stoch Environ Res Risk Assess 28:2101–2118

Ghiasi-Freez J, Kadkhodaie-Ilkhchi A, Ziaii M (2012) Improving the accuracy of flow units prediction through two committee machine models: an example from the South Pars Gas Field, Persian Gulf Basin, Iran. Comput Geosci 46:10–23

Gokceoglu C (2002) A fuzzy triangular chart to predict the uniaxial compressive strength of the Ankara agglomerates from their petrographic composition. Eng Geol 66(1–2):39–51

Gokceoglu C, Sonmez H, Zorlu K (2009) Estimating the uniaxial compressive strength of some clay bearing rocks selected from Turkey by nonlinear multivariable regression and rule-based fuzzy models. Expert Syst 26(2):176–190

Haykin S (1999) Neural networks: a comprehensive foundation. Prentice-Hall, New Jersey, p 842

Horsrud P (2001) Estimating mechanical properties of shale from empirical correlations. SPE Drill Complet 16:68–73

Huang C, Davis L, Townshend J (2002) An assessment of support vector machines for land cover classification. Int J Remote Sens 23(4):725–749

ISRM (1981) Rock characterization testing and monitoring. In: Brown ET (ed) ISRM suggested methods. Pergamon Press, Oxford, p 211

Jain S, Bhatia MPS (2013) Performance investigation of support vector regression using meteorological data. Int J Database Theor Appl 6(4):109–118

Jang JSR (1993) ANFIS: adaptive network based fuzzy inference system. IEEE Trans Syst, Man, Cybern 23(3):665–685

Joachims T (1998) Text categorization with support vector machines: learning with many relevant features. In: Nédellec C, Rouveirol

C (eds) Proceedings of of ECML-98, 10th European conference on machine learning, Springer, Chemnitz, DE, 1398, pp 137–142
Lee KH (2004) First course on fuzzy theory and applications. Springer, Berlin, p 335
Leong MK, Lin SW, Chen HB, Tsai FU (2010) Predicting mutagenicity of aromatic amines by various machine learning approaches. Toxicol Sci 116(2):498–513
Malekmohammadi I, Bazargan-Lari MR, Kerachian R, Nikoo MR, Fallahnia M (2011) Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction. Ocean Eng 38:487–497
Mamdani EH (1976) Advances in the linguistic synthesis of fuzzy controllers. Int J Man-Mach Stud 8(6):669–678
Mamdani EH, Assilian S (1975) An experiment in linguistic synthesis with a fuzzy logic controller. Int J Man Mach Stud 7:1–13
Mishra DA, Basu A (2012) Use of the block punch test to predict the compressive and tensile strengths of rocks. Int J Rock Mech Min Sci 51:119–127
Mishra DA, Basu A (2013) Estimation of uniaxial compressive strength of rock materials by index tests using regression analysis and fuzzy inference system. Eng Geol 160:54–68
Mishra AK, Desai VR (2006) Drought forecasting using feed-forward recursive neural network. Ecol Model 198(1–2):127–138
Mohammadi K, Shamshirband S, Tong CW, Arif M, Petkovic D, Ch S (2015) A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. Energy Convers Manage 92:162–171
Moosavi V, Vafakhah M, Shirmohammadi B, Behnia N (2013) A wavelet-ANFIS hybrid model for groundwater level forecasting for different prediction periods. Water Resour Manag 27(5):1301–1321
Nefeslioglu HA (2013) Evaluation of geo-mechanical properties of very weak and weak rock by using non-destructive techniques: ultrasonic pulse velocity measurements and reflectance spectroscopy. Eng Geol 160(27):8–20
Nourani V, Kisi Z, Mehdi K (2011) Two hybrid artificial Intelligence approaches for modeling rainfall-runoff process. J Hydrol 402(1–2):41–59
Nourani V, Baghanam AH, Adamowski J, Gebremichael M (2013) Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall-runoff modeling. J Hydrol 476:228–243
Parmar KS, Bhardwaj R (2015) River water prediction modeling using neural networks, fuzzy and wavelet coupled model. Water Resour Manag 29(1):17–33
Pulido-Calvo I, Gutiérrez-Estrada JC (2009) Improved irrigation water demand forecasting using a soft-computing hybrid model. Biosyst Eng 102(2):202–218
Rajaee T, Mirbagheri SA, Zounemat-Kermani M, Nourani V (2009) Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. Sci Total Environ 407(17):4916–4927
Rajasekaran S, Gayathri S, Lee TL (2008) Support vector regression methodology for storm surge predictions. Ocean Eng 35(16):1578–1587
Rzhevsky V, Novick G (1971) The physics of rocks. Mir Publisher, Moscow, p 320
Sarkar K, Tiwary A, Singh TN (2010) Estimation of strength parameters of rock using artificial neural networks. Bull Eng Geol Environ 69:599–606
Scholkopf B, Smola A (2002) In learning with Kernels: support vector machines, regularization, optimization, and beyond, 1st edn. MIT Press, Cambridge
Shamshirband S, Petkoviæ D, Anuar NB, Saboohi H, Inayat I, Akib S, Cojbasic Z, Nicolic V, Kiah MLM, Gani A (2014) Wind turbine power coefficient estimation by soft computing methodologies: comparative study. Energy Convers Manage 81:520–526
Shamshirband S, Mohammadi K, Tong CW, Zamani M, Motamedi S, Sudheer CH (2015) A hybrid SVM-FFA method for prediction of monthly mean global solar radiation. Theor Appl Climatol. doi:10.1007/s00704-015-1482-2
Singh TN, Verma AK (2012) Comparative analysis of intelligent algorithms to correlate strength and petrographic properties of some schistose rocks. Eng Comput 28:1–12
Singh R, Vishal V, Singh TN, Ranjith PG (2012) A comparative study of generalized regression neural network approach and adaptive neuro-fuzzy inference systems for prediction of unconfined compressive strength of rocks. Neural Comput Appl 23(2):499–506
Sonebi M, Cevik A, Grunewald S, Walraven J (2016) Modelling the fresh properties of self-compacting concrete using support vector machine approach. Constr Build Mater 106:55–64
Sonmez H, Gokceoglu C, Medley EW, Tuncay E, Nefeslioglu HA (2006) Estimating the uniaxial compressive strength of a volcanic bimrock. Int J Rock Mech Min Sci 43:554–561
Sugeno M (1985) Industrial application of fuzzy control. North-Holland, New York, p 269
Taghipour K, Mohajjel M (2013) Structure and generation mode of travertine fissure-ridges in Azarshahr area, Azarbaydjan, NW Iran. Iran J Geol 7(25):15–33
Vapnik V (2000) The nature of statistical learning theory. Springer, Berlin
Vapnik VN, Vapnik V (1998) Statistical learning theory. Wiley, New York, p 736
Vapnik V, Golowich SE, Smola A (1996) Support vector method for function approximation, regression estimation, and signal processing. Adv Neural Inf Process Syst 9:281–287
Verma AK, Singh TN (2013) A neuro-fuzzy approach for prediction of longitudinal wave velocity. Neural Comput Appl 22(7–9):1685–1693
Wanas N, Auda G, Kamel M S, Karray, F (1998) On the optimal number of hidden nodes in a neural network. In: Proceedings of the IEEE Canadian conference on electrical and computer engineering vol 2, pp 918–921
Wu KP, Wang SD (2009) Choosing the kernel parameters for support vector machines by the inter-cluster distance in the feature space. Pattern Recognit 42(5):710–717
Yagiz S, Sezer EA, Gokceoglu C (2012) Artificial neural networks and nonlinear regression techniques to assess the influence of slake durability cycles on the prediction of uniaxial compressive strength and modulus of elasticity for carbonate rocks. Int J Numer Anal Methods 36:1636–1650
Yang H, Huang K, King I, Lyu MR (2009) Localized support vector regression for time series prediction. Neurocomputing 72(10):2659–2669
Yesiloglu-Gultekin N, Gokceoglu C, Sezer EA (2013) Prediction of uniaxial compressive strength of granitic rocks by various nonlinear tools and comparison of their performances. Int J Rock Mech Min Sci 62:113–122
Yilmaz I (2009) A new testing method for indirect determination of the unconfined compressive strength of rocks. Int J Rock Mech Min Sci 46:1349–1357
Zadeh LA (1965) Fuzzy sets. Inf Control 8(3):338–353
Zhang L, Ding X, Budhu MI (2012) A rock expert system for the evaluation of rock properties. Int J Rock Mech Min Sci 50(1):124–132
Zounemat-Kermani M, Teshnehlab M (2008) Using adaptive neuro-fuzzy inference system for hydrological time series prediction. Appl Soft Comput 8(2):928–936