

A review of artificial intelligence applications in shallow foundations

Mohamed A. Shahin*

Geotechnical engineering deals with materials (e.g. soil and rock) that, by their very nature, exhibit varied and uncertain behavior because of the imprecise physical processes associated with the formation of these materials. Modeling the behavior of such materials in geotechnical engineering applications is complex and sometimes beyond the ability of most traditional forms of physically based engineering methods. Artificial intelligence (AI) is becoming more popular and particularly amenable to modeling the complex behavior of most geotechnical engineering applications, including foundations, because it has demonstrated superior predictive ability compared to traditional methods. The main aim of this paper is to review the AI applications in shallow foundations and present the salient features associated with the AI modeling development. The paper also discusses the strengths and limitations of AI techniques compared to other modeling approaches.

Keywords: Artificial intelligence, Shallow foundations, Modeling, Neural networks, Genetic programming, Evolutionary polynomial regression

Introduction

Over the last decade, artificial intelligence (AI) has been applied successfully to virtually every problem in geotechnical engineering. Among the available AI techniques are artificial neural networks (ANNs), genetic programming (GP), evolutionary polynomial regression (EPR), support vector machines, M5 model trees, and k -nearest neighbors (Elshorbagy *et al.*, 2010). The focus of this paper will be on three AI techniques, including ANNs, GP, and EPR. These three techniques are selected because they have been proved to be the most successful applied AI techniques in geotechnical engineering especially for shallow foundations. Of these, ANNs are by far the most commonly used one. Interested readers are referred to Shahin *et al.* (2001), where the pre-2001 ANN applications in geotechnical engineering are reviewed in some detail, and Shahin *et al.* (2009) and Shahin (2013), where the post-2001 papers of AI applications in geotechnical engineering are briefly examined.

The behavior of foundations (deep and shallow) in soils is complex, uncertain, and not yet entirely understood. This fact has encouraged many researchers to apply the AI techniques for modeling the behavior of foundations. In particular, ANNs have been used for shallow foundations including settlement estimation (Chen *et al.*, 2009; Shahin *et al.*, 2002b, 2003; Sivakugan *et al.*, 1998; Soleimanbeigi

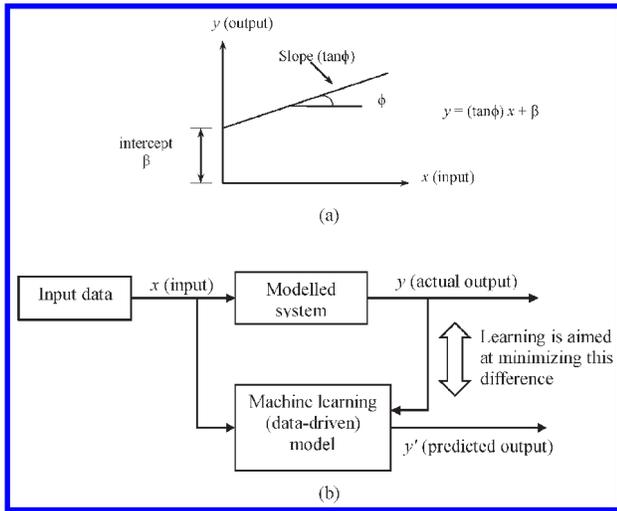
and Hataf, 2006) and prediction of ultimate bearing capacity (Behera *et al.*, 2013a, 2013b; Kalinli *et al.*, 2011; Kuo *et al.*, 2009; Padmini *et al.*, 2008; Provenzano *et al.*, 2004; Soleimanbeigi and Hataf, 2005). Likewise, GP and EPR have been investigated for settlement prediction of shallow foundations (Rezania and Javadi, 2007; Shahin, 2014; Shahnazari *et al.*, 2014) as well as ultimate bearing capacity (Adarsh *et al.*, 2012; Pan *et al.*, 2013; Shahin, 2014; Shahnazari and Tutunchian, 2012; Tsai *et al.*, 2013). The objective of this paper is to provide an overview of some of the popular AI techniques, present a review of the AI applications to date in shallow foundations, and discuss some of the current challenges and future directions.

Overview of AI

Artificial intelligence is a computational method that attempts to mimic, in a very simplistic way, the human cognition capability (e.g. emulating the operation of the human brain at the neural level) to solve engineering problems that have defied solution using conventional computational techniques (Flood, 2008). The essence of AI techniques in solving any engineering problem is to learn by examples of data inputs and outputs presented to them so that the subtle functional relationships among the data are captured, even if the underlying relationships are unknown or the physical meaning is difficult to explain. Thus, AI models are data-driven models that rely on the data alone to determine the structure and parameters that govern a phenomenon (or system) and do not make any assumptions about the physical behavior of the system. This is in

Department of Civil Engineering, Curtin University, Perth, WA 6845, Australia

*Corresponding author, email m.shahin@curtin.edu.au



1 Linear regression versus artificial intelligence (AI) modeling: a Linear regression modeling (Shahin *et al.*, 2001); b AI data-driven modeling [adapted from (Solomatine and Ostfeld, 2008)]

contrast to most physically based models that use the first principles (e.g. physical laws) to derive the underlying relationships of the system, which usually justifiably simplified with many assumptions and require prior knowledge about the nature of the relationships among the data. This is one of the main benefits of AI techniques when compared to most physically based empirical and statistical methods.

The AI modeling philosophy is similar to a number of conventional statistical models, in the sense that both are attempting to capture the relationship between a historical set of model inputs and the corresponding outputs. For example, imagine a set of x -values and corresponding y -values in two-dimensional space, where $y=f(x)$. The objective is to find the unknown function f that relates the input variable x to the output variable y . In a linear regression statistical model, the function f can be obtained by changing the slope $\tan \phi$ and intercept β of the straight line in Fig. 1a, so that the error between the actual outputs and the outputs of the straight line is minimized. The same principle is used in AI models. Artificial intelligence can form the simple linear regression model by having one input and one output (Fig. 1b). Artificial intelligence uses available data to map between the system inputs and the

corresponding outputs using machine learning by repeatedly presenting examples of the model inputs and outputs (training) to find the function $y=f(x)$ that minimizes the error between the historical (actual) outputs and the outputs predicted by the AI model.

If the relationship between x and y is non-linear, statistical regression analysis can be applied successfully only if prior knowledge of the nature of the non-linearity exists. On the contrary, this prior knowledge of the nature of the non-linearity is not required for AI models. In the real world, it is likely that complex and highly non-linear problems are encountered, and in such situations, traditional regression analyses are inadequate (Gardner and Dorling, 1998). In this section, a brief overview of three selected AI techniques (i.e. ANNs, GP, and EPR) is presented below.

Artificial neural networks

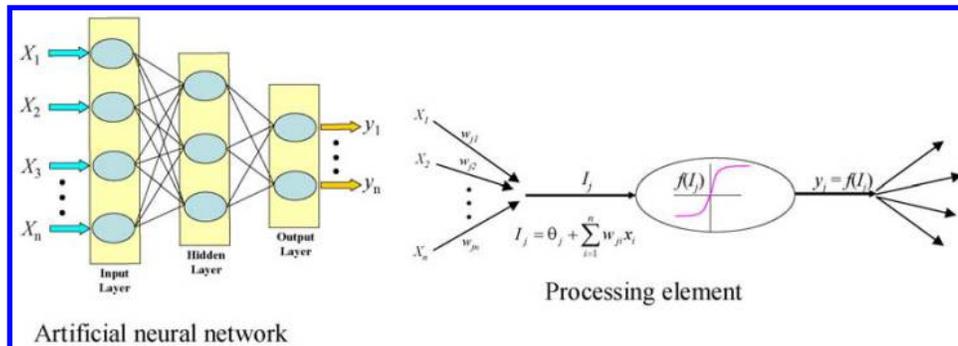
Artificial neural networks are a form of AI that attempt to mimic the function of the human brain and nervous system. Although the concept of ANNs was first introduced in 1943 (McCulloch and Pitts, 1943), research into applications of ANNs has blossomed since the introduction of the back-propagation training algorithm for feed-forward multi-layer perceptrons (MLPs) in 1986 (Rumelhart *et al.*, 1986). Many authors have described the structure and operation of ANNs (e.g. Fausett, 1994; Zurada, 1992). Typically, the architecture of ANNs consists of a series of processing elements (PEs), or nodes, that are usually arranged in layers: an input layer, an output layer, and one or more hidden layers, as shown in Fig. 2.

The input from each PE in the previous layer x_i is multiplied by an adjustable connection weight w_{ji} . At each PE, the weighted input signals are summed and a threshold value θ_j is added. This combined input I_j is then passed through a non-linear transfer function $f(\cdot)$ to produce the output of the PE y_j . The output of one PE provides the input to the PEs in the next layer. This process is summarized in equations (1) and (2), and illustrated in Fig. 2

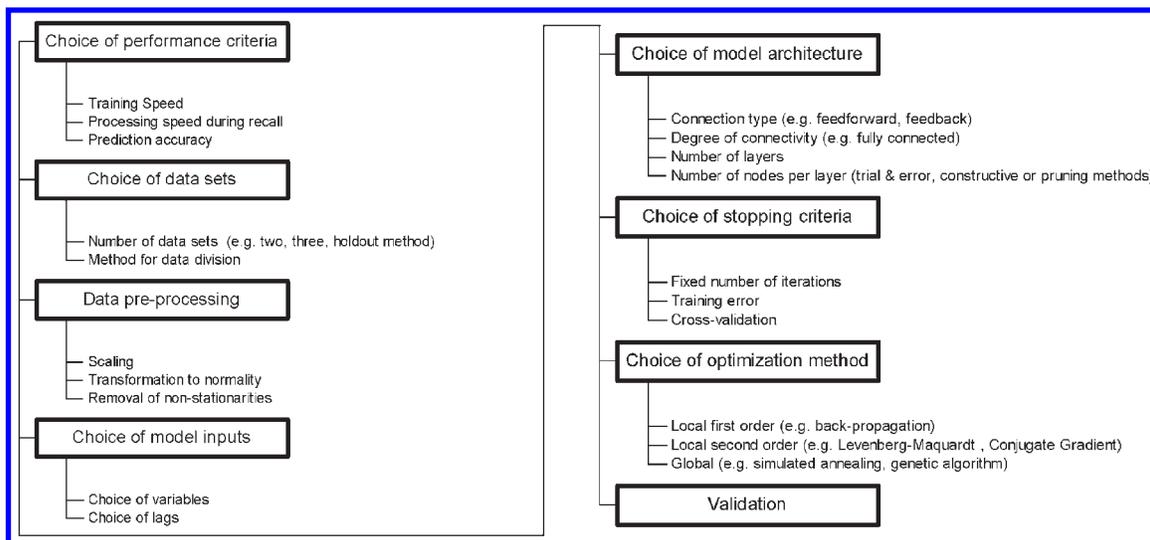
$$I_j = \sum w_{ji} \cdot x_i + \theta_j \quad \text{summation} \quad (1)$$

$$y_j = f(I_j) \quad \text{transfer} \quad (2)$$

The propagation of information in an ANN starts at the input layer, where the input data are presented. The



2 Typical structure and operation of artificial neural networks (ANNs) (Shahin *et al.*, 2009)



3 Main steps in artificial neural network (ANN) model development (Maier and Dandy, 2000)

network adjusts its weights on the presentation of a training data set and uses a learning rule to find a set of weights that produces the input/output mapping that has the smallest possible error. This process is called *learning* or *training*. Once the training phase of the model has been successfully accomplished, the performance of the trained model needs to be validated using an independent validation set. The main steps involved in the development of an ANN, as suggested by Maier and Dandy (2000), are illustrated in Fig. 3 and discussed in some depth in Shahin (2013).

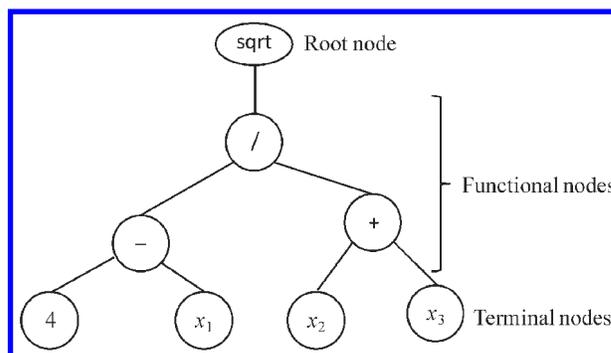
Genetic programming

Genetic programming is an extension of genetic algorithms (GA), which are evolutionary computing search (optimization) methods that are based on the principles of genetics and natural selection. In GA, some of the natural evolutionary mechanisms, such as reproduction, cross-over, and mutation, are usually implemented to solve function identification problems. Genetic programming was first introduced by Holland (1975) and developed by Goldberg (1989), whereas GP was invented by Cramer (1985) and further developed by Koza (1992). The difference between GA and GP is that GA is generally used to evolve the best values for a given set of model parameters (i.e. parameters optimization), whereas GP generates a structured representation for a set of input variables and the corresponding outputs (i.e. modeling or programming).

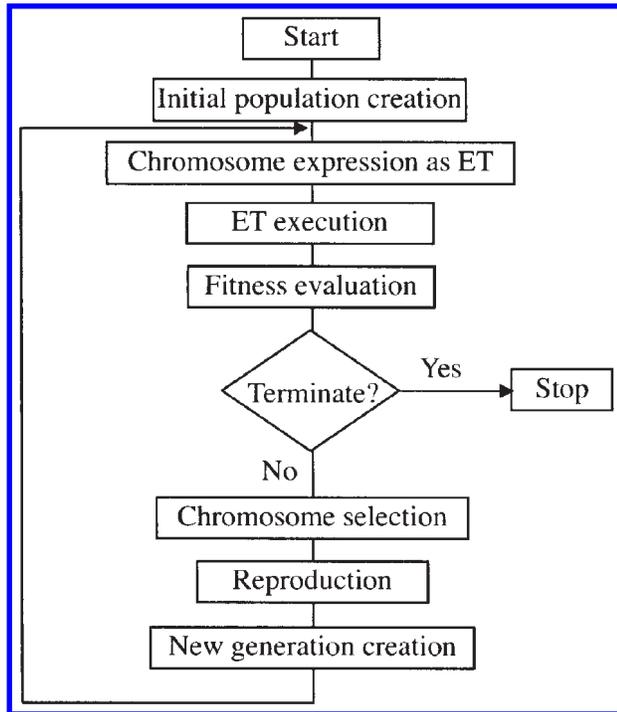
Genetic programming manipulates and optimizes a population of computer models (or programs) proposed to solve a particular problem, so that the model that best fits the problem is obtained. A detailed description of GP can be found in many publications (e.g. Koza, 1992), and a brief overview is given herein. The modeling steps by GP start with the creation of an initial population of computer models (also called *individuals* or *chromosomes*) that are composed of two sets (i.e. a set of functions and a set of terminals) that are defined by the user to suit a certain problem. The functions and terminals are selected randomly

and arranged in a tree-like structure to form a computer model that contains a root node, branches of functional nodes, and terminals, as shown by the typical example of GP tree representation in Fig. 4. The functions can contain basic mathematical operators (e.g. +, -, ×, /), Boolean logic functions (e.g. AND, OR, NOT), trigonometric functions (e.g. sin, cos), or any other user-defined functions. The terminals, on the other hand, may consist of numerical constants, logical constants, or variables.

Once a population of computer models has been created, each model is executed using available data for the problem at hand, and the model fitness is evaluated depending on how well it is able to solve the problem. For many problems, the model fitness is measured by the error between the output provided by the model and the desired actual output. A generation of new population of computer models is then created to replace the existing population. The new population is created by applying the following three main operations: reproduction, cross-over, and mutation. These three operations are applied on certain proportions of the computer models in the existing population, and the models are selected according to their fitness. Reproduction is copying a computer model from an existing population into the new population without



4 Typical example of genetic programming (GP) tree representation for the function $[(4 - x_1)/(x_2 + x_3)]^2$



5 Algorithm of gene expression programming (GEP) (Teodorescu and Sherwood, 2008)

alteration. Crossover is genetically recombining (swapping) randomly chosen parts of two computer models. Mutation is replacing a randomly selected functional or terminal node with another node from the same function or terminal set, provided that a functional node replaces a functional node and a terminal node replaces a terminal node. The evolutionary process of evaluating the fitness of an existing population and producing new population is continued until a termination criterion is met, which can be either a particular acceptable error or a certain maximum number of generations. The best computer model that appears in any generation designates the result of the GP process. There are currently three variants of GP available in the literature including the linear genetic programming (LGP), gene expression programming (GEP), and multi-expression programming (MEP) (Alavi and Gandomi, 2011). More recently, the multi-stage genetic programming (MSGP) (Gandomi and Alavi, 2011) and multi-gene genetic programming (MGGP) (Gandomi and Alvari, 2012) are also introduced. However, GEP is the most commonly used GP method in geotechnical engineering and is thus described in some detail below.

Gene expression programming was developed by Ferreira (2001) and utilizes evolution of mathematical equations that are encoded linearly in chromosomes of fixed length and expressed non-linearly in the form of expression trees (ETs) of different sizes and shapes. The chromosomes are composed of multiple genes; each gene encoded a smaller sub-program or sub-expression tree (Sub-ET). Every gene has a constant length and consists of a head and a tail. The head can contain functions and terminals (variables and constants) required to code any expression, whereas the tail solely contains terminals.

The genetic code represents a one-to-one relationship between the symbols of the chromosome and the function or terminal. The process of information decoding from chromosomes to ETs is called *translation*, which is based on sets of rules that determine the spatial organization of the functions and terminals in the ETs and the type of interaction (link) between the sub-ETs (Ferreira, 2001). The main strength of GEP is that the creation of genetic diversity is extremely simplified as the genetic operators work at the chromosome level. Another strength is regarding the unique multi-genetic nature of GEP, which allows the evolution of more powerful models/programs composed of several sub-programs (Ferreira, 2001).

The major steps in the GEP procedure are schematically represented in Fig. 5. The process begins with choosing sets of functions F and terminals T to randomly create an initial population of chromosomes of mathematical equations. One could choose, for example, the four basic arithmetic operators to form the set of functions, i.e. $F = \{+, -, \times, / \}$, and the set of terminals will obviously consist of the independent variables of a particular problem, for example, for a problem that has two independent variables, x_1 and x_2 , would be $T = \{x_1, x_2\}$. Choosing the chromosomal architecture, i.e. the number and length of genes and linking functions (e.g. addition, subtraction, multiplication, and division), is also part of this step. The chromosomes are then expressed as ETs of different sizes and shapes, and the performance of each individual chromosome is evaluated by comparing the predicted and actual values of the presented data. One could measure the fitness f_i of an individual chromosome i using the following expression

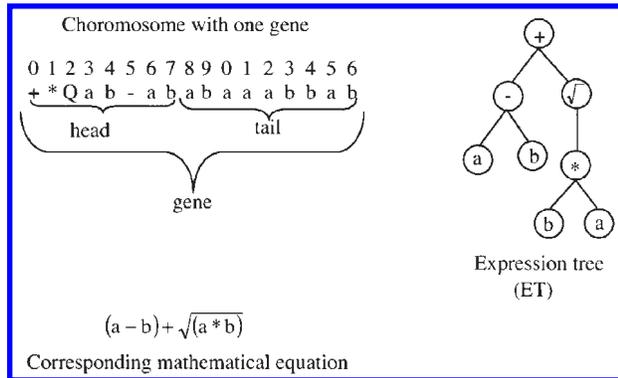
$$f_i = \sum_{j=1}^{C_t} (M - |C_{(i,j)} - T_j|) \quad (3)$$

where M is the range of selection, $C_{(i,j)}$ is the value returned by the individual chromosome i for fitness case j (out of C_t fitness cases), and T_j is the target value for the fitness case j . There are, of course, other fitness functions available that can be appropriate for different problems. If the desired results (according to the measured errors) are satisfactory, the GEP process is stopped; otherwise, some chromosomes are selected and mutated to reproduce new chromosomes, and the process is repeated for a certain number of generation or until the desired fitness score is obtained.

Figure 6 shows a typical example of a chromosome with one gene, and its ET and corresponding mathematical equation. It can be seen that, while the head of a gene contains arithmetic and trigonometric functions (e.g., $+$, $-$, \times , $/$, $\sqrt{\quad}$, \sin , \cos), the tail includes constants and independent variables (e.g. 1 , a , b , c). The ET is codified reading the ET from left to right in the top line of the tree and from top to bottom.

Evolutionary polynomial regression

Evolutionary polynomial regression is a hybrid regression technique based on evolutionary computing that was developed by Giustolisi and Savic (2006). It constructs symbolic models by integrating the soundest features of



6 Schematic representation of a chromosome with one gene and its expression tree (ET) and corresponding mathematical equation (Kayadelen, 2011)

numerical regression, with GP and symbolic regression (Koza, 1992). The following two steps roughly describe the underlying features of the EPR technique, aimed to search for polynomial structures representing a system. In the first step, the selection of exponents for polynomial expressions is carried out, employing an evolutionary searching strategy by means of GA (1989). In the second step, numerical regression using the least square method is conducted, aiming to compute the coefficients of the previously selected polynomial terms. The general form of expression in EPR can be presented as follows (Giustolisi and Savic, 2006)

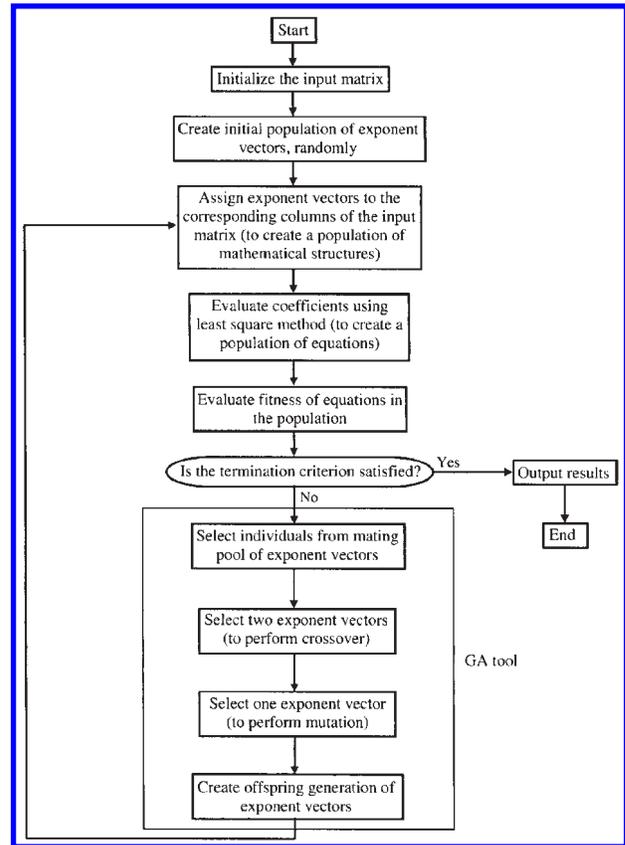
$$y = \sum_{j=1}^m F(X, f(X), a_j) + a_0 \tag{4}$$

where y is the estimated vector of output of the process, m is the number of terms of the target expression, F is a function constructed by the process, X is the matrix of input variables, f is a function defined by the user, and a_j is a constant. A typical example of EPR pseudo-polynomial expression that belongs to the class of equation (4) is as follows (Giustolisi and Savic, 2006)

$$\hat{Y} = a_0 + \sum_{j=1}^m a_j \cdot (X_1)^{ES(j,1)} \dots (X_k)^{ES(j,k)} \cdot f[(X_1)^{ES(j,k+1)} \dots (X_k)^{ES(j,2k)}] \tag{5}$$

where Y is the vector of target values, m is the length of the expression, a_j is the value of the constants, X_i is the vector(s) of the k candidate inputs, ES is the matrix of exponents, and f is a function selected by the user.

Evolutionary polynomial regression is suitable for modeling physical phenomena, based on two features (Savic *et al.*, 2006): (i) the introduction of prior knowledge about the physical system/process, to be modeled at three different times, namely before, during, and after EPR modeling calibration and (ii) the production of symbolic formulae, enabling data mining to discover patterns that describe the desired parameters. In the first EPR feature (i) above, before the construction of the EPR model, the modeler selects the relevant inputs and arranges them in a



7 Typical flow diagram of the evolutionary polynomial regression (EPR) procedure (Rezania *et al.*, 2011)

suitable format according to their physical meaning. During the EPR model construction, model structures are determined by following user-defined settings such as general polynomial structure, user-defined function types (e.g. natural logarithms, exponentials, tangential hyperbolics), and searching strategy parameters. The EPR starts from true polynomials and also allows for the development of non-polynomial expressions containing user-defined functions (e.g. natural logarithms). After EPR model calibration, an optimum model can be selected from among the series of models returned. The optimum model is selected based on the modeler’s judgment, in addition to statistical performance indicators such as the coefficient of determination. A typical flow diagram of the EPR procedure is shown in Fig. 7, and a detailed description of the technique can be found in Giustolisi and Savic (2006).

Artificial intelligence applications in shallow foundations

This section provides an overview of most AI applications, including ANNs, GP, and EPR, that have appeared to date in relation to examining the relative success or otherwise of AI in shallow foundations. It should be noted that it is not intended to cover every single application or scientific paper that can be found on this topic but rather the intention is to provide a general overview of some of

the more relevant applications in engineering problems of shallow foundations. Some works are selected to be described in some detail, whereas others are acknowledged for reference purposes.

Based on the author's experience, there are several factors in the use of AI techniques that need to be systematically investigated when developing AI models for geotechnical engineering problems, so that model performance can be improved. These factors include the determination of adequate model inputs, data division, data preparation, model validation, model robustness, model transparency and knowledge extraction, and model uncertainty. Some of these factors have received recent attention, whereas others require further research. Discussion of these factors is beyond the scope of this paper but can be found in Shahin (2013). Some of these factors are briefly discussed in the applications presented below.

Settlement estimation

The design of foundations is generally controlled by the criteria of bearing capacity and settlement, the latter often being the governing factor in shallow foundations on cohesionless soils. The estimation of settlement of shallow foundations (on cohesionless soils in particular) is very complex, uncertain, and not yet entirely understood. This fact has encouraged a number of researchers to apply the ANN technique to the settlement of shallow foundations on cohesionless soils. For example, Sivakugan *et al.* (1998) carried out a preliminary study on a small set of data to explore the possibility of using ANNs to predict the settlement of shallow foundations on sands. A neural network was trained with five inputs representing the net applied pressure, average blow count from the standard penetration test (SPT), width of the foundation, shape of the foundation, and depth of the foundation. The output was the settlement of the foundation. With the aid of cascade-correlation, a network with 1 hidden layer and 11 hidden nodes was found optimal. The results obtained by the neural network were compared with the methods proposed by Terzaghi and Peck (1967) and Schmertmann (1970). Based on the results obtained, it was shown that the traditional methods of Terzaghi and Peck, and Schmertmann overestimate the settlements by about 2.2 and 3.4 times, respectively, as shown in Fig. 8a. In contrast, the predictions using the ANN model were good (see Fig. 8b). Using the same neural network features, Arnold (1999) extended the work done by Sivakugan *et al.* (1998) with a database containing a larger number of data cases. His work, although relatively superficial, found that the best network consisted of 18 hidden layer nodes with correlation coefficients equal to 0.954, 0.955, and 0.944 for the training, testing, and validation sets, respectively. It should be noted that 18 hidden layer nodes are considered to be large for a network with five input variables, which may affect the generalization ability of the model.

The ANN models developed above for settlement prediction of shallow foundations on cohesionless soils have been built on either a limited number of data cases or have suffered from the lack of a comprehensive procedure for testing their robustness and generalization ability. In an attempt to develop more well-established models, Shahin

et al. (2002a, 2002b) carried out a comprehensive study to predict the settlement of shallow foundations on cohesionless soils utilizing ANNs. Using a large database of actual measured settlements and MLPs trained with the back-propagation algorithm, Shahin *et al.* (2002b) developed an ANN model that was found to outperform the most commonly used traditional methods. The model was trained using five inputs representing the footing width, net applied footing load, average blow count obtained from the SPT over the depth of influence of the foundations as a measure of soil compressibility, footing geometry (length to width of footing), and footing embedment ratio (embedment depth to footing width). The single model output was the foundation settlement. The results between the predicted and measured settlements obtained by utilizing ANNs were compared with three traditional methods, namely Meyerhof (1965), Schultze and Sherif (1973), and Schmertmann *et al.* (1978). Comparisons of the results obtained using the ANN and the above traditional methods in the validation set are given in Table 1 and presented in Fig. 9. It is evident from Table 1 that the ANN model performs better than the traditional methods for all performance measures considered. It is also evident from Fig. 9 that the ANN model performs reasonably well over the full range of measured settlements considered. In contrast, the traditional methods appear to work well for only small settlements, in the range of 10–20 mm. The method of Schmertmann *et al.* (1978) tends to over-predict larger settlements, whereas the method of Schultze and Sherif (1973) tends to under-predict severely larger settlements and the method of Meyerhof (1965) appears both to over- and under-predict larger settlements. In an attempt to facilitate the use of the obtained ANN model and to make it more accessible, Shahin *et al.* (2002a) translated the model into a tractable and relatively simple formula suitable for hand calculation. The derived formula can be used to calculate the settlement as follows (Shahin *et al.*, 2002a)

$$S_p^{\text{ANN}} = 0.6 + \left[\frac{120.4}{1 + e^{(0.312 - 0.725 \tanh x_1 + 2.984 \tanh x_2)}} \right] \quad (6)$$

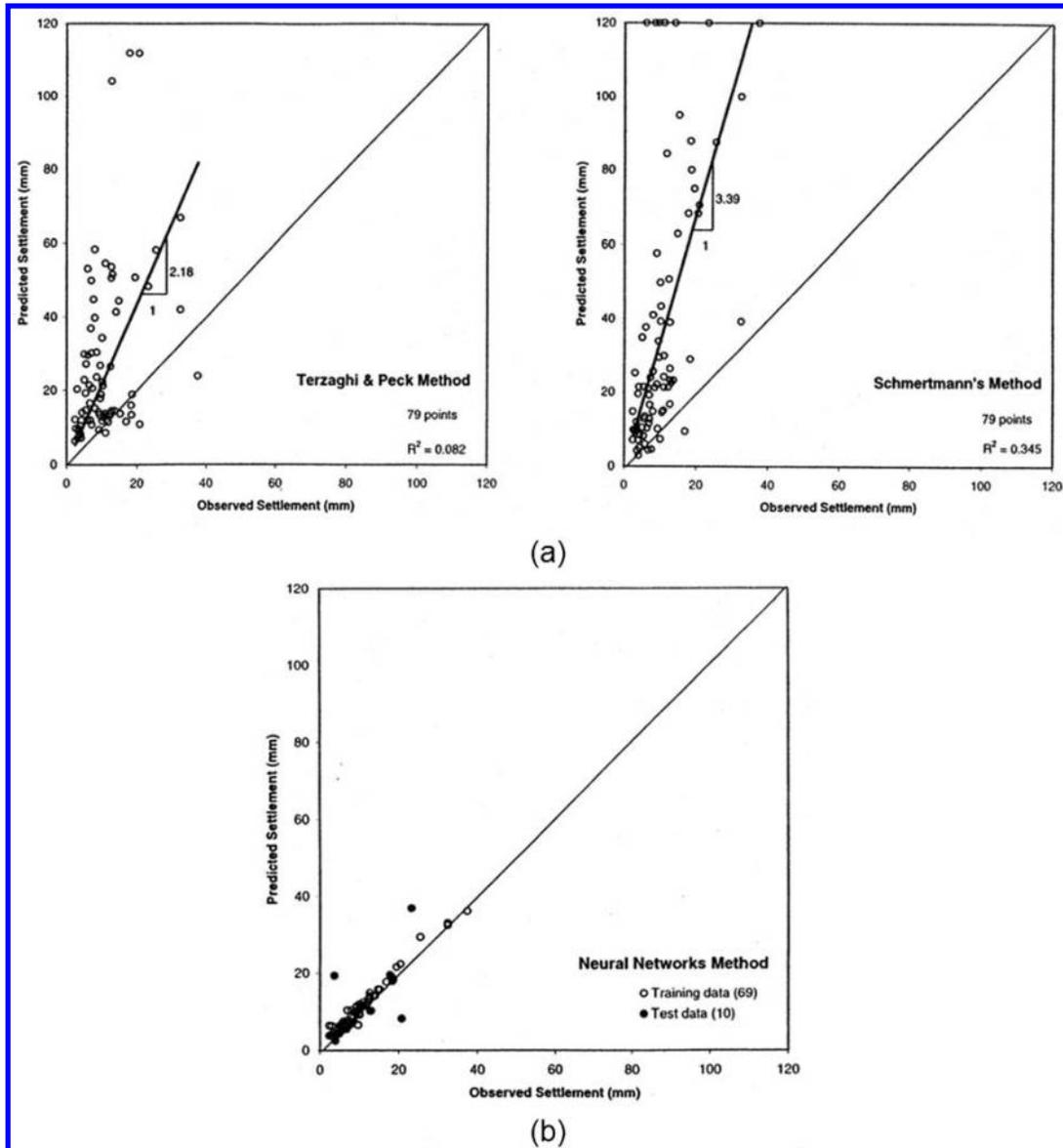
in which

$$x_1 = 0.1 + 10^{-3} \left[3.8B + 0.7q + 4.1N - 1.8 \left(\frac{L}{B} \right) + 19 \left(\frac{D_f}{B} \right) \right] \quad (7)$$

$$x_2 = 10^{-3} \left[0.7 - 41B - 1.6q + 75N - 52 \left(\frac{L}{B} \right) + 740 \left(\frac{D_f}{B} \right) \right] \quad (8)$$

where S_p is the predicted settlement (mm), B is the footing width (m), q is the net applied footing load (kPa), N is the average SPT blow count, L/B is the footing geometry, and D_f/B is the footing embedment ratio.

Using the same database of Shahin *et al.* (2002a) and similar model inputs and outputs, Rezanian and Javadi (2007) and Shahnazari *et al.* (2014) developed two different GP models, and Shahnazari *et al.* (2014) developed an



8 Settlement predictions using artificial neural networks (ANNs) and traditional methods by Sivakugan *et al.* (1998): *a* Traditional methods; *b* ANNs

additional EPR model. The formulation of the GP model developed by Rezania and Javadi (2007) is as follows

$$S_p^{GP} = \frac{q(1.80B + 4.62) - 346.15D_f}{N^2} + \frac{11.22L - 11.11}{L} \quad (9)$$

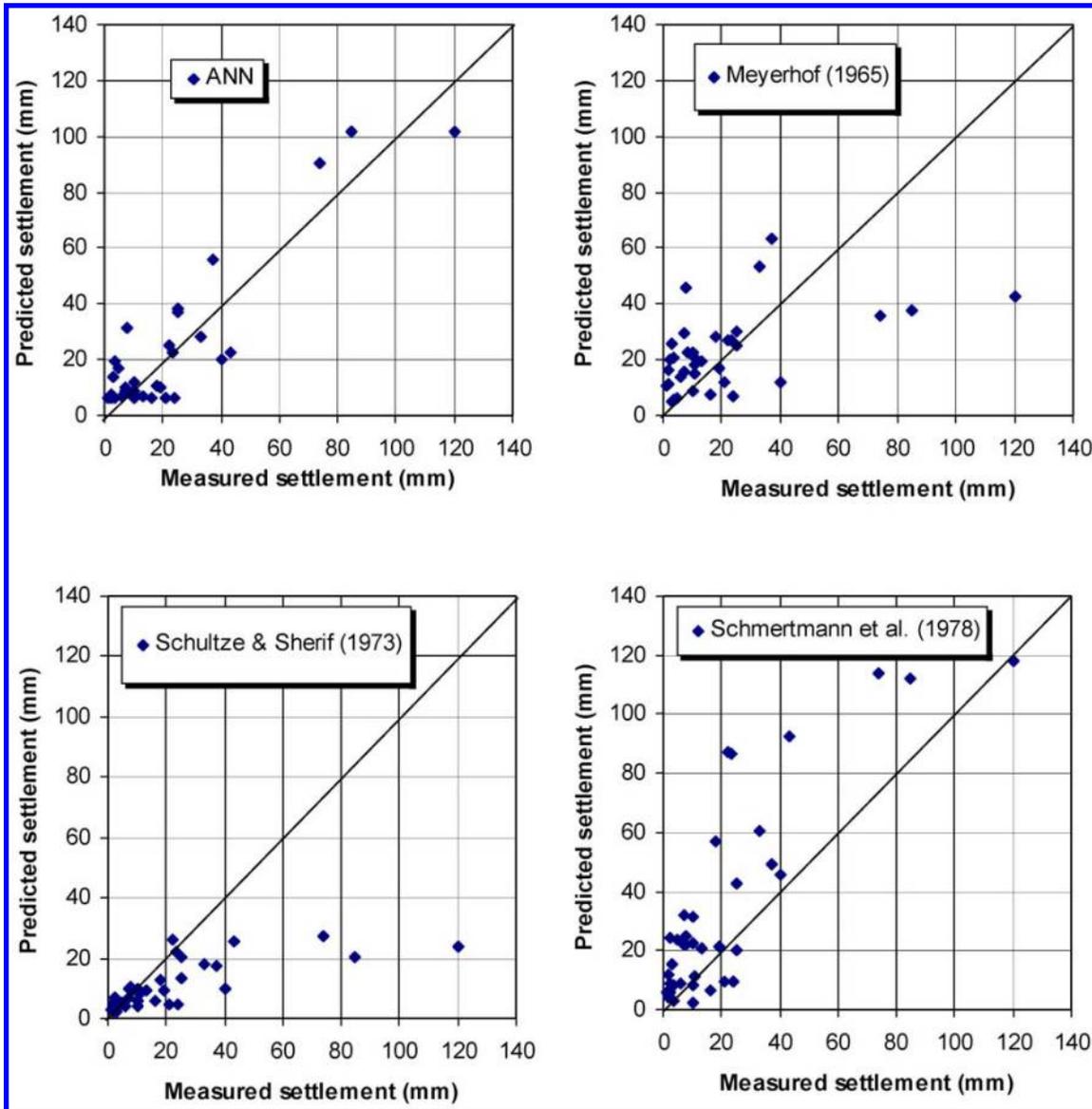
The formulations of the GP and EPR models developed by Shahnazari *et al.* (2014) are as follows

$$S_p^{GP} = \frac{2.5B(N/B - 1 + (B+1)/(D_f + 0.16B)) + (2B - N)/L + q/N}{(N + (D_f/B)(B - L/B) + B/N)} \quad (10)$$

$$S_p^{EPR} = \frac{7.2q}{N^2} - \frac{190}{NL} + 3.4 \frac{B\sqrt{q} - 10\sqrt{BD_f}}{N} - 0.3\sqrt{Bq} + 16 \quad (11)$$

The above GP and EPR models represented by equations (9)–(11) were compared with the traditional methods and were found to outperform most available methods.

All of the above AI models were also compared to each other in terms of the coefficient of determination (R^2), root mean squared error (RMSE), and mean absolute error (MAE), and the results are given in Table 2. It can be seen that the EPR- and GP-based models are the most precise models in predicting the settlement of shallow foundations on cohesionless soils. It can also be seen that the accuracy of the ANN model proposed by Shahin *et al.* (2002a) and the GP model developed by Rezania and Javadi (2007) is high but not as high as the EPR and GP models of Shahnazari *et al.* (2014). Overall, the statistical results indicate that the application of EPR and GP methods provides a more potential improvement over the ANN model.



9 Settlement predictions using artificial neural networks (ANNs) and traditional methods by Shahin *et al.* (2002b)

The use of reinforcement to increase the bearing capacity and reduce the settlement of shallow foundations is a common construction technique; however, a few practical methods have been developed to compute the settlement of shallow foundations on reinforced cohesionless soils (Soleimanbeigi and Hataf, 2006). In lieu of this fact, Soleimanbeigi and Hataf (2006) examine the potentiality of ANNs to investigate the settlement of shallow foundations

on reinforced cohesionless soils. A number of 123 data records obtained from both laboratory and field measurements were used for ANN model development and verification. The ANN model inputs included the footing size, soil properties, and reinforcement characteristics, whereas the single output was the settlement of shallow reinforced foundation. The results showed a good degree of model accuracy (i.e. $r=0.876$, $RMSE=2.23$ mm, and

Table 1 Performance of artificial neural network (ANN) model and traditional methods in the validation set for settlement estimation (Shahin *et al.*, 2002b)

Performance measure	ANN (Shahin <i>et al.</i> , 2002b)	Meyerhof (1965)	Schultze and Sherif (1973)	Schmertmann <i>et al.</i> (1978)
r	0.905	0.440	0.729	0.798
RMSE/mm	11.04	25.72	23.55	23.67
MAE/mm	8.78	16.59	11.81	15.69

RMSE: root mean squared error; MAE: mean absolute error.

MAE=1.61 mm) for the validation set. To provide a practical tool for design purposes, the authors developed several design charts that were drawn from the developed ANN model, which can be readily used for fast determination of settlement of shallow reinforced foundations.

Bearing capacity prediction

Bearing capacity is one of two criteria that governs the design of shallow foundations, and has been examined by several AI researchers, especially using ANNs. Provenzano *et al.* (2004) explored the potential of using neurofuzzy networks, namely ANFIS (adaptive network-based fuzzy inference system), for predicting the load-settlement behavior of shallow foundations, subjected to vertical centered and eccentric loads. Neurofuzzy networks are ANN modeling technique that combines the explicit linguistic knowledge representation of fuzzy systems with the learning power of neural networks (Brown and Harris, 1995). Neurofuzzy networks can be trained by processing data samples to perform input/output mappings, similar to the way traditional neural networks do, with the additional benefit of being able to provide a set of production *if-then* linguistic fuzzy rules that describe the model input/output relationships in a transparent way, such as

IF(x_1 is high AND x_2 is low) THEN (y is high), $c=0.9$

where x_1 and x_2 are input variables, y is the corresponding output variable, and $c=0.9$ is the rule confidence that indicates the degree to which the above rule has contributed to the output. The neural network was trained and validated with data obtained from results of small scale model experiments. The model was trained to predict the load transmitted by the footing to the soil, as the single model input. The model inputs included the seepage gradient, load eccentricity ratio (i.e. eccentricity/footing width), and non-dimensional settlement of footing (i.e. settlement/footing width). The model was found to be robust as it was able to reproduce the relationship between settlement and vertical load with a good

Table 2 Statistical performances of different artificial intelligence (AI) models in the validation set for settlement estimation (Shahnazari *et al.*, 2014)

Reference	Method	R^2	RMSE	MAE
Shahin <i>et al.</i> (2002b)	ANN	0.851	10.25	7.14
Rezania and Javadi (2007)	GP	0.826	11.07	6.77
Shahnazari <i>et al.</i> (2014)	GP	0.878	9.27	6.03
Shahnazari <i>et al.</i> (2014)	EPR	0.871	11.07	6.77

RMSE: root mean squared error; MAE: mean absolute error; ANN: artificial neural network; GP: genetic programming; EPR: evolutionary polynomial regression.

Table 3 Performance of artificial neural network (ANN) model and traditional methods in the validation set for ultimate bearing capacity of shallow foundation on reinforced cohesionless soils (Soleimanbeigi and Hataf, 2005)

Performance measure	ANN (Soleimanbeigi and Hataf (2005))	Huang and Tatsuoka (1990)	Huang and Meng (1997)	Zhao <i>et al.</i> (1998)
r	0.981	0.947	0.016	0.59
RMSE/kPa	45.7	129.3	347.8	150.7
MAE/kPa	28.2	76.6	238.1	128.6

RMSE: root mean squared error; MAE: mean absolute error.

accuracy, for footings with large values of eccentricity. However, for centered and small eccentricity footings, the model accuracy was less satisfying and the model was rather sensitive to the load conditions. On the other hand, Padmini *et al.* (2008) also investigated the applicability of ANFIS neurofuzzy for predicting the ultimate bearing capacity of shallow foundations on cohesionless soil, and undertook a comparative study with the commonly used bearing capacity theories. The result indicated that the ANFIS model was able to predict well the ultimate bearing capacity of shallow foundations and perform significantly better than the theoretical methods.

Soleimanbeigi and Hataf (2005) developed an ANN model for predicting the ultimate bearing capacity of shallow foundations on reinforced soils. In this study, the data used for model development were obtained from the literature and comprised a number of 351 records of laboratory and field measurements of bearing capacity of shallow foundations on reinforced cohesionless soils. The model results were compared with three traditional methods, i.e. Huang and Tatsuoka (1990), Huang and Meng (1997), and Zhao *et al.* (1998). The model inputs were the footing width, footing geometry (i.e. footing length/footing width), footing depth ratio (i.e. depth of embedment/footing width), internal friction angle of soil, unit weight of soil, number of reinforcement, first layer depth ratio (i.e. first reinforcement level depth/footing width), vertical spacing of reinforcement layer ratio (i.e. vertical spacing of reinforcement layers/footing width), reinforcement width ratio (i.e. reinforcement width/footing width), and reinforcement stiffness. The single model output was the bearing capacity of the foundation. The methods were compared using three performance measures including the coefficient of correlation r , RMSE, and MAE, as shown in Table 3. It is evident from Table 3 that the ANN model outperforms the traditional methods. The relative importance of the input variables for bearing capacity was also carried out by performing a sensitivity analysis, and it was found that the footing width, number of reinforcement layers, and reinforcement vertical spacing have more impact on bearing capacity than the other factors.

Kuo *et al.* (2009) used ANN-based model for predicting the bearing capacity of strip footing on multi-layered cohesive soil. In this study, the data used for model development were generated from a parametric study carried out using numerical formulation of upper and lower bound theorems. The inputs to the ANN model included the cohesion of the multi-soil layers, the stratum thickness of the soil layers, and the footing width. The model output was the ultimate bearing capacity. To

examine the accuracy of the ANN model, it was compared with some other traditional methods, namely the multi-regression and Bowles methods (Bowles, 1997). The result indicated that the ANN model was able to predict well the bearing capacity of a strip footing and significantly outperform the other methods. The sensitivity analysis carried out to test the robustness of the developed ANN model and examine the relative importance of the model inputs indicated that the bearing capacity of a footing on multi-layered soils increases as the soil cohesion increases, and as the width of the footing increases, and is strongly influenced by the soil layers located immediately beneath the footing.

More recently, Behera *et al.* (2013a) utilized ANNs to predict the ultimate bearing capacity of eccentrically inclined loaded strip footing over sand. The ANN model was developed using data obtained from extensive laboratory model tests that were carried out on a strip footing foundation lying over sand bed subjected to an eccentrically inclined load. Based on the model test results, the ANN model was developed to predict the reduction factor, which was defined as the ratio of the ultimate bearing capacity of the foundation subjected to an eccentrically inclined load to the ultimate bearing capacity of the foundation subjected to a centric vertical load. Different sensitivity analyses were also carried out to evaluate the parameters affecting the reduction factor. A prediction model equation was also established using the trained weights of the ANN model. The predictions from the ANN model, and those from two other available empirical approaches developed by Patra *et al.* (2012b) and Loukidis *et al.* (2008), were compared. The ANN model inputs included the eccentricity ratio (i.e. load eccentricity/footing width), embedment ratio (i.e. embedment depth/footing width), and ratio of load inclination (i.e. load inclination angle/friction angle of soil), whereas the output was the reduction factor as defined above. The coefficient of determination (or efficiency) R^2 for the training and validation sets was found to be equal to 0.995 and 0.993, respectively, indicating a high level of accuracy. The predictability of the ANN model was also found to be slightly better than the empirical equations used for comparison. It was observed from the relative importance study that the load inclination ratio is the most important input parameter followed by the eccentricity ratio and embedment ratio. It was also observed from the sensitivity analyses that the eccentricity and load inclination ratios are inversely related to the reduction factor, whereas the embedment ratio is directly related to the reduction factor.

Behera *et al.* (2013b), soon after, developed another ANN model to estimate the reduction factor of the ultimate bearing capacity of eccentrically inclined loaded strip footing in reinforced condition, and model performance was found to be reasonably well with R^2 equal to 0.994 and 0.988 for the training and validation sets, respectively. The model predictability was also found to be more accurate than the regression equation proposed by Patra *et al.* (2012a). Another successful application for the use of ANN for predicting the ultimate bearing capacity of shallow foundations was carried out by Kalinli *et al.* (2011).

The application of GP and EPR techniques for predicting the ultimate bearing capacity of shallow foundation is relatively recent. Adarsh *et al.* (2012) developed a GP model to predict the ultimate bearing capacity of cohesionless soils beneath shallow foundations. In this study, the model inputs were the footing width, footing depth, length-to-width ratio of footing, density of soil, and angle of internal friction of soil, and the model output was the ultimate bearing capacity. The model results were compared with three theoretical approaches, as well as an ANN model and fuzzy inference system reported in the literature. The statistical evaluation of the results showed that the GP model is better than the theoretical approaches and competed well with the other AI techniques, i.e. ANN and fuzzy inference models. Similarly, Shahnazari and Tutunchian (2012) developed another GP model for predicting the ultimate capacity of shallow foundations on cohesionless soils. In this case, the GP model was calibrated and validated using an experimental database consisting of 100 load tests. The results revealed that the GP model can predict the ultimate bearing capacity precisely with a high coefficient of correlation r of 98%.

Pan *et al.* (2013) introduced two different models, a GP model and an EPR model, to predict the ultimate bearing capacity of shallow foundations. The inputs of models included the footing width (denoted P_1), embedment depth (denoted P_2), footing length-to-width ratio (denoted P_3), unit weight of soil (denoted P_4), and angle of shearing resistance or friction angle of soil (denoted P_5). The single output of models was the ultimate bearing capacity (denoted q_u). It should be noted that the GP and EPR techniques are able to choose automatically the inputs that have the most significant impact on the output and thus can produce more parsimonious models than ANNs. Moreover, the models can be presented in the form of visible mathematical formulae that facilitate parameter studies and sensitivity analysis. The results demonstrated that the proposed GP and EPR models are outstanding in prediction accuracy and provide simple mathematical formulation as well. The GP model was found to have a coefficient of determination R^2 of 0.988 and 0.984 in the training and validation sets, respectively, whereas these values were found to be equal to 0.984 and 0.987, respectively, for the EPR model. The derived formulae are as follows (Pan *et al.*, 2013)

$$q_u^{\text{GP}} = \frac{7024P_2P_4 \sin(P_1) - 86.2 \log| -43.9P_5 + 97.5P_4 |}{-4.1[1.11 \exp(3.12 \cos(-0.886P_4))]^{0.132 \log(68.5P_1)}} \quad (13)$$

$$q_u^{\text{EPR}} = 67P_1(0.0535P_5 - 2.11P_1 + 6.05P_2^2)^5 \quad (14)$$

It can be seen from equations (13) and (14) above that both the GP and EPR models did not select the geometry ratio (P_3) as a significant model input, whereas the EPR model did not also select the unit weight of soil (P_4) as a significant model input. Another successful application for the use of GP and EPR for predicting the ultimate bearing capacity of shallow foundations was carried out by Tsai *et al.* (2013).

Discussion and conclusion

In geotechnical engineering, it is most likely to encounter problems that are very complex and not well understood. In this regard, AI provides several advantages over more traditional computing techniques. For most traditional mathematical models, the lack of physical understanding is usually supplemented by either simplifying the problem or incorporating several assumptions into the models. Mathematical models also rely on assuming the structure of the model in advance, which may be sub-optimal. Consequently, many mathematical models fail to simulate the complex behavior of most geotechnical engineering problems. In contrast, AI techniques are a data-driven approach in which the model can be trained on input–output data pairs to determine the structure and parameters of the model. In this case, there is no need to either simplify the problem or incorporate any assumptions. Moreover, AI models can always be updated to obtain better results by presenting new training examples as new data become available. These factors combine to make AI a powerful modeling tool in geotechnical engineering. In the field of foundation engineering, it was evident from this review that AI techniques have been applied successfully to shallow foundations. Based on the results of the studies reviewed in this paper, it can be concluded that AI techniques perform better than, or at least as well as, the traditional methods used as a basis for comparison.

Despite the success of AI techniques, they are still facing classical opposition because of some inherent shortcomings that need further attention in the future including the lack of transparency, knowledge extraction, and model uncertainty. Detailed discussion of such shortcomings is beyond the scope of this paper but have been presented in detail by Shahin (2013). For example, special attention should be paid to incorporating prior knowledge about the underlying physical process based on engineering judgment or human expertise into the learning formulation, checking of model robustness, and evaluation of model results. Improvements in these issues will greatly enhance the usefulness of AI techniques and will provide the next generation of applied AI models with the best way for advancing the field to the next level of sophistication and application. The author suggests that AI techniques for the time being might be treated as a complement to conventional computing techniques rather than as an alternative, or may be used as a quick check on solutions developed by more time-consuming and in-depth analyses.

References

- Adarsh, S., Dhanya, R., Krishna, G., Merlin, R. and Tina, J. 2012. Prediction of ultimate bearing capacity of cohesionless soils using soft computing techniques, *ISRN Artif. Intell.*, **2012**, (628496), 10.
- Alavi, A. H. and Gandomi, A. H. 2011. A robust data mining approach for formulation of geotechnical engineering systems, *Eng. Comput.*, **28**, (3), 242–274.
- Arnold, M. A. 1999. Artificial neural networks applied to the prediction of settlements of shallow foundations in granular soils, MEngSc Thesis, James Cook University, Queensland.
- Behera, R. N., Patra, C. R., Sivakugan, N. and Das, B. M. 2013a. Prediction of ultimate bearing capacity of eccentrically inclined loaded strip footing by ANN, part I, *Int. J. Geotech. Eng.*, **7**, (1), 36–44.
- Behera, R. N., Patra, C. R., Sivakugan, N. and Das, B. M. 2013b. Prediction of ultimate bearing capacity of eccentrically inclined loaded strip footing by ANN: part II, *Int. J. Geotech. Eng.*, **7**, (2), 165–172.
- Bowles, J. E. 1997. Foundation analysis and design, New York, NY, McGraw-Hill.
- Brown, M. and Harris, C. J. 1995. A perspective and critique of adaptive neurofuzzy systems used for modelling and control applications, *Int. J. Neural Syst.*, **6**, (2), 197–220.
- Chen, Y., Azzam, R., Fernandez, T. M. and Li, L. 2009. Studies on construction pre-control of a connection aisle between two neighbouring tunnels in Shanghai by means of 3D FEM, neural networks and fuzzy logic, *Geotech. Geol. Eng.*, **27**, (1), 155–167.
- Cramer, N. L. 1985. A representation for the adaptive generation of simple sequential programs, in Proceedings of the International Conference on Genetic Algorithms and their Applications, Pittsburgh, PA, USA, 183–187.
- Elshorbagy, A., Corzo, G., Srinivasulu, S. and Solomatine, D. P. 2010. Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology-part I: concepts and methodology, *Hydrol. Earth Syst. Sci.*, **14**, 1931–1941.
- Fausett, L. V. 1994. Fundamentals neural networks: architecture, algorithms, and applications, Englewood Cliffs, NJ, Prentice-Hall.
- Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems, *Complex Syst.*, **13**, (2), 87–129.
- Flood, I. 2008. Towards the next generation of artificial neural networks for civil engineering, *Adv. Eng. Informat.*, **22**, (1), 4–14.
- Gandomi, A. H. and Alavi, A. H. 2011. Multi-stage genetic programming: a new strategy to nonlinear system modeling, *Inform. Sci.*, **181**, (23), 5227–5239.
- Gandomi, A. H. and Alviri, A. H. 2012. A new multi-gene genetic programming approach to non-linear system modeling, part II: geotechnical and earthquake engineering problems, *Neural Comput. Appl.*, **21**, (1), 189–201.
- Gardner, M. W. and Dorling, S. R. 1998. Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences, *Atmos. Environ.*, **32**, (14/15), 2627–2636.
- Giustolisi, O. and Savic, D. A. 2006. A symbolic data-driven technique based on evolutionary polynomial regression, *J. Hydroinform.*, **8**, (3), 207–222.
- Goldberg, D. E. 1989. Genetic algorithms in search optimization and machine learning, Boston, MA, Addison-Wesley Longman Publishing Co., Inc.
- Holland, J. H. 1975. Adaptation in natural and artificial systems, University of Michigan Press, Michigan, MA, USA.
- Huang, C. C. and Meng, F. Y. 1997. Ultimate bearing capacity and settlement of footings on reinforced sandy ground, *J. Geotech. Geoenviron. Eng.*, **123**, (1), 30–36.
- Huang, C. C. and Tatsuoka, F. 1990. Bearing capacity of reinforced sandy ground, *Geotext. Geomembranes*, **9**, (1), 51–82.
- Kalinli, A., Cemal Acar, M. and 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.
- Kayadelen, C. 2011. Soil liquefaction modeling by genetic expression programming and neuro-fuzzy, *Expert Syst. Appl.*, **38**, (4), 4080–4087.
- Koza, J. R. 1992. Genetic programming: on the programming of computers by natural selection, Cambridge, MA, MIT Press.
- Kuo, Y. L., Jaksa, M. B., Lyamin, A. V. and Kaggwa, W. S. 2009. ANN-based model for predicting the bearing capacity of strip footing on multi-layered cohesive soil, *Comput. Geotech.*, **36**, (3), 503–516.
- Loukidis, D., Chakraborty, T. and Salgado, R. 2008. Bearing capacity of strip footing on purely frictional soil under eccentric and included loads, *Can. Geotech. J.*, **45**, (6), 768–787.
- Maier, H. R. and Dandy, G. C. 2000. Applications of artificial neural networks to forecasting of surface water quality variables: issues, applications and challenges, in Artificial neural networks in hydrology, (eds. R. S. Govindaraju and A. R. Rao), 287–309, Dordrecht, Netherlands, Kluwer.
- McCulloch, W. S. and Pitts, W. 1943. A logical calculus of ideas imminent in nervous activity, *Bull. Math. Biophys.*, **5**, 115–133.
- Meyerhof, G. G. 1965. Shallow foundations, *J. Soil Mech. & Found. Div.*, **91**, (SM2), 12–31.

- Padmini, D., Ilamparuthi, K. and Sudheer, K. P. 2008. Ultimate bearing capacity prediction of shallow foundations on cohesionless soils using neurofuzzy models, *Comput. Geotech.*, **35**, (1), 33–46.
- Pan, C.-P., Tsai, H.-C. and Lin, Y.-H. 2013. Improving semi-empirical equations of ultimate bearing capacity of shallow foundations using soft computing polynomials, *Eng. Appl. Artif. Intell.*, **26**, (1), 478–487.
- Patra, C. R., Behera, R. N., Sivakugan, N. and Das, B. M. 2012a. Ultimate bearing capacity of shallow foundation under eccentrically inclined load: part II, *Int. J. Geotech. Eng.*, **6**, (4), 507–514.
- Patra, C. R., Behera, R. N., Sivakugan, N. and Das, B. M. 2012b. Ultimate bearing capacity of shallow strip foundation under eccentrically inclined load: part I, *Int. J. Geotech. Eng.*, **6**, (3), 343–352.
- Provenzano, P., Ferlisi, S. and Musso, A. 2004. Interpretation of a model footing response through an adaptive neural fuzzy inference system, *Comput. Geotech.*, **31**, (3), 251–266.
- Rezania, M., Faramarzi, A. and Javadi, A. 2011. An evolutionary based approach for assessment of earthquake-induced soil liquefaction and lateral displacement, *Eng. Appl. Artif. Intell.*, **24**, (1), 142–153.
- Rezania, M. and Javadi, A. 2007. A new genetic programming model for predicting settlement of shallow foundations, *Can. Geotech. J.*, **44**, (12), 1462–1472.
- Rumelhart, D. E., Hinton, G. E. and Williams, R. J. 1986. Learning internal representation by error propagation, in *Parallel distributed processing: explorations in the microstructure of cognition*, (eds. D. E. Rumelhart and J. L. McClelland), Cambridge, MA, MIT Press, pp. 318–362.
- Savic, D. A., Giutolisi, O., Berardi, L., Shepherd, W., Djordjevic, S. and Saul, A. 2006. Modelling sewer failure by evolutionary computing, *Proc. Inst. Eng. Water Manag.*, **159**, (WM2), 111–118.
- Schmertmann, J. H. 1970. Static cone to compute static settlement over sand, *J. Soil Mech. & Found. Div.*, **96**, (SM3), 7302–1043.
- Schmertmann, J. H., Hartman, J. P. and Brown, P. B. 1978. Improved strain influence factor diagrams, *J. Geotech. Eng.*, **104**, (GT8), 1011–1043.
- Schultze, E. and Sherif, G. 1973. Prediction of settlements from evaluated settlement observations for sand, *Proc. 8th Int. Conf. Soil Mech. & Found. Eng.*, **1**, 225–230.
- Shahin, M. A. 2013. Artificial intelligence in geotechnical engineering: applications, modeling aspects, and future directions, in *Metaheuristics in water, geotechnical and transport engineering*, (eds. X. Yang, A. H. Gandomi, S. Talatahari, and A. H. Alavi), 169–204, London, Elsevier Inc.
- Shahin, M. A. 2014. Use of evolutionary computing for modelling some complex problems in geotechnical engineering, *Geomech. Geoeng.*
- Shahin, M. A., Jaksa, M. B. and Maier, H. R. 2001. Artificial neural network applications in geotechnical engineering, *Aust. Geomech.*, **36**, (1), 49–62.
- Shahin, M. A., Jaksa, M. B. and Maier, H. R. 2002a. Artificial neural network-based settlement prediction formula for shallow foundations on granular soils, *Aust. Geomech.*, **37**, (4), 45–52.
- Shahin, M. A., Jaksa, M. B. and Maier, H. R. 2009. Recent advances and future challenges for artificial neural systems in geotechnical engineering applications, *J. Adv. Artif. Neural Syst.*, **2009**, (308239), 9.
- Shahin, M. A., Maier, H. R. and Jaksa, M. B. 2002b. Predicting settlement of shallow foundations using neural networks, *J. Geotech. Geoenviron. Eng.*, **128**, (9), 785–793.
- Shahin, M. A., Maier, H. R. and Jaksa, M. B. 2003. Settlement prediction of shallow foundations on granular soils using B-spline neurofuzzy models, *Comput. Geotech.*, **30**, (8), 637–647.
- Shahnazari, H., Shahin, M. A. and Tutunchian, M. A. 2014. Evolutionary-based approaches for settlement prediction of shallow foundations on cohesionless soils, *Int. J. Civ. Eng.*, **12**, (1), 55–64.
- Shahnazari, H. and Tutunchian, M. A. 2012. Prediction of ultimate bearing capacity of shallow foundations on cohesionless soils: an evolutionary approach, *KSCE J. Civ. Eng.*, **16**, (6), 950–957.
- Sivakugan, N., Eckersley, J. D. and Li, H. 1998. Settlement predictions using neural networks, *Aust. Civ. Eng. Trans.*, **CE40**, 49–52.
- Soleimanbeigi, A. and Hataf, N. 2005. Predicting ultimate bearing capacity of shallow foundations on reinforced cohesionless soils using artificial neural networks, *Geosynth. Int.*, **12**, (6), 321–332.
- Soleimanbeigi, A. and Hataf, N. 2006. Prediction of settlement of shallow foundations on reinforced soils using neural networks, *Geosynth. Int.*, **13**, (4), 161–170.
- Solomatine, D. P. and Ostfeld, A. 2008. Data-driven modelling: some past experience and new approaches, *J. Hydroinform.*, **10**, (1), 3–22.
- Teodorescu, L. and Sherwood, D. 2008. High energy physics event selection with gene expression programming, *Comput. Phys. Commun.*, **178**, (6), 409–419.
- Terzaghi, K. and Peck, R. 1967. *Soil Mechanics in Engineering Practice*, New York, John Wiley & Sons.
- Tsai, H.-C., Tyan, Y.-Y., Wu, Y.-W. and Lin, Y.-H. 2013. Determining ultimate bearing capacity of shallow foundations using a genetic programming system, *Neural Comput. Appl.*, **23**, (7–8), 2073–2084.
- Zhao, A., Riccutti, A. and Gupta, V. 1998. Design of geosynthetics reinforced foundations by the slip-line method, in *Geohorizon: state of the art in geosynthetics technology*, (ed. S. S. Sarkar), 283–293, Rotterdam, Netherlands, Balkema.
- Zurada, J. M. 1992. *Introduction to artificial neural systems*, St. Paul, MN, West Publishing Company.