

1 **Development of genetic-based models for predicting the resilient modulus**
2 **of cohesive pavement subgrade soils**

3
4
5 **Behnam Ghorbani¹, Arul Arulrajah², Guillermo Narsilio³, Suksun Horpibulsuk⁴, and**
6 **Myint Win Bo⁵**

7
8
9 Behnam Ghorbani¹

10 Ph.D. Student, Department of Civil and Construction Engineering, Swinburne University of
11 Technology, P.O. Box 218, Hawthorn, Melbourne, VIC 3122, Australia. E-mail:
12 bghorbani@swin.edu.au

13
14 Arul Arulrajah²

15 Professor, Department of Civil and Construction Engineering, Swinburne University of
16 Technology, P.O. Box 218, Hawthorn, Melbourne, VIC 3122, Australia. E-mail:
17 aarulrajah@swin.edu.au

18
19 Guillermo A. Narsilio³

20 Associate Professor, Department of Infrastructure Engineering, The University of Melbourne,
21 Parkville, VIC 3010, Australia. E-mail: narsilio@unimelb.edu.au

22
23 Suksun Horpibulsuk⁴

24 Adjunct Professor, Department of Civil and Construction Engineering, Swinburne University
25 of Technology, Hawthorn, VIC 3122, Australia & Director, Center of Excellence in innovation
26 for sustainable infrastructure development, Suranaree University of Technology, Nakhon
27 Ratchasima 30000, Thailand. E-mail: suksun@g.sut.ac.th

28
29 Myint Win Bo⁵

30 President and CEO, Bo & Associates Inc., Mississauga, Ontario, Canada. E-mail:
31 mwbo@boassociates.ca

32
33
34
35 Corresponding Authors:

36 Prof Suksun Horpibulsuk

37 Adjunct Professor, Department of Civil and Construction Engineering, Swinburne University
38 of Technology & Director, Center of Excellence in Innovation for Sustainable Infrastructure
39 Development, Suranaree University of Technology, Nakhon Ratchasima 30000, Thailand

40 Email : suksun@g.sut.ac.th

41 Phone : 66-44-22-4322

42 Fax : 66-44-22-4607

43
44 Prof. Arul Arulrajah

45 Professor, Department of Civil and Construction Engineering, Swinburne University of
46 Technology, P.O. Box 218, Hawthorn, Melbourne, VIC 3122, Australia. E-mail:

47 aarulrajah@swin.edu.au

48
49 **ABSTRACT**

50 Accurate determination of resilient modulus (M_r) of pavement subgrade soils is an important
51 factor for successful design of pavement system. M_r is an important soil property, which is
52 complex in nature as it is dependent on several influential factors, such as soil physical
53 properties, applied stress conditions, and environmental conditions. The aim of this study is to
54 explore the potential of an evolutionary algorithm, i.e., genetic algorithm (GA), and a hybrid
55 intelligent approach combining neural network with GA (ANN-GA), for estimation of M_r of
56 cohesive pavement subgrade soils. To achieve this aim, a reliable database containing results
57 of repeated load triaxial tests (RLT) and other index properties of subgrade soils was utilized.
58 GA was employed to develop a precise equation for predicting M_r of subgrade soils. In addition,
59 GA was used as a tool for determining the optimal values of the weights and the bias of the
60 ANN-GA approach. The developed ANN-GA model was then transferred to a functional
61 relationship for further application and analyses. Several validation and verification phases
62 were conducted to examine the performance of the developed models. The results indicated that
63 both GA and ANN-GA models could accurately predict the M_r of cohesive subgrade soils, and
64 had better performance compared to the available models in the literature. Finally, a sensitivity
65 analysis was conducted to evaluate the effect of utilized parameters on the M_r .

66

67 **Keywords:** Pavement subgrade; Resilient modulus; Genetic algorithm; Optimized neural
68 network; Hybrid.

69 **Introduction**

70 The successful design of a pavement system with various material properties is affected by the
71 stiffness and strength of the pavement layers, i.e. surface, base, subbase, and subgrade. Amongst
72 pavement layers, the subgrade is considered as the foundation of the pavement system that
73 transfers applied loads to the ground. To characterize the subgrade material under different
74 environmental and loading conditions, the resilient modulus (M_r) has been introduced as a
75 fundamental material property that describes the inelastic behavior of material under traffic
76 loading (AASHTO, 2003). As pavement layers undergo repeated traffic loading, the subgrade
77 soil undergoes recoverable as well as permanent strains with each load repetition. When the
78 number of load repetitions is increased, plastic deformation decreases until it becomes all
79 recoverable (Sadrossadat et al., 2016). M_r is defined as the ratio of the applied deviator stress
80 to the recoverable strain (AASHTO, 2003).

81 M_r of subgrade soils is commonly determined through repeated load triaxial (RLT) test on
82 remoulded or undisturbed samples (AASHTO, 2003). In this test, varying combinations of
83 confining and deviator stresses are applied to the specimen in various sequences to simulate the
84 field conditions. Despite their accuracy and reliability, performing RLT tests requires
85 sophisticated equipment and skilled personnel. In order to avoid performing time consuming
86 and uneconomical RLT tests, several relationships have been proposed that formulate M_r in
87 terms of stress states, physical properties, and strength tests parameters.

88 M_r mainly depends on stress states and physical properties of the subgrade soils. Generally,
89 models obtained for calculation of resilient modulus of subgrade soils can be classified into two
90 main groups: (I) Correlations with laboratory tests and in situ test results (II) Constitutive
91 equations. Recently, P-wave and S-wave measurements have also been used for estimation of
92 resilient modulus of pavement materials (Schuettpelez et al., 2010). A summary of some of the
93 available equations for estimation of M_r of cohesive soils are summarized in **Table 1**.

94 Correlations are typically developed by relating the M_r value to the results of laboratory tests
95 such as California bearing ratio (CBR) and unconfined compressive strength (UCS); in-situ
96 tests such as cone penetration test (CPT) and dynamic cone penetrometer (DCP); soil physical
97 properties like moisture content (w) and dry density (γ_d); stress states such as confining pressure
98 (σ_3) and deviator stress (σ_d); or a combination of these parameters. Constitutive models, on the
99 other hand, are obtained by relating the M_r value to various stress invariants, i.e. bulk stress (θ),

100 octahedral shear stress (τ_{oct}), and octahedral normal stress (σ_{oct}). Unknown parameters of the
 101 constitutive models can be related to soil physical properties to include both physical properties
 102 and stress conditions.

103 Considering Kim (2004) equation in **Table 1**, regression coefficients for cohesive subgrade
 104 soils (A-6 soils) can be obtained using following equations (Hanittinan, 2007):

$$k_1 = a_1 a_3^{a_2} + a_3 \left(\frac{S_r}{100} \right)^{a_4} + a_5 q_u + a_6 PI + a_7 (LL - w) + a_8 (w_{opt} - w) + a_9 (P_{200} - a_1) \quad (1)$$

$$k_2 = b_1 \sigma_3^{b_2} + b_3 \left(\frac{S_r}{100} \right)^{b_4} + b_5 q_u^{b_6} + b_7 PI + b_8 LL \quad (2)$$

$$a_1 = a_{11} + a_{12} \left(\frac{w_{opt} - w}{w_{opt}} \right) \quad (3)$$

$$b_1 = b_{11} + b_{12} (w - w_{opt}) \quad (4)$$

105 where a_1 to a_{12} and b_1 to b_{12} are constants of Kim's model. Detailed description of the Kim's
 106 regression coefficients can be found in Hanittinan (2007). As can be seen in **Table 1**,
 107 constitutive models such as those provided by Seed et al. (1967), Witczak and Uzan (1988),
 108 and Puppala et al. (1996) have a pre-defined structure, and unknown coefficients of these
 109 equations are typically obtained by performing regression analysis. Considering Kim (2004)
 110 equation in **Table 1**, a second set of regression analysis is required to relate the obtained
 111 coefficients, i.e., k_1 and k_2 , to soil parameters as summarized in **Eq. (1-4)**. As can be seen, the
 112 procedure for finding the final model is complicated and time consuming. On the other hand,
 113 while regression based equations may perform well on the utilized datasets, they are not
 114 typically tested and validated on new datasets (Gandomi et al., 2013, Ghorbani et al., 2018).
 115 The complexity of the M_r factor as well as the importance of considering the nonlinear
 116 interaction between variables necessitates the use of more advanced techniques for prediction
 117 of M_r of subgrade soils.

118 To overcome the limitations of the traditional modeling techniques, artificial intelligence (AI)
 119 methods have been employed by various researchers for solving complicated engineering
 120 problems (Ghorbani et al., 2018, Chen et al., 2019, Jayawardana et al., 2019, Ghorbani and
 121 Hasanzadehshooiili, 2018). Unlike traditional modeling techniques, artificial intelligence

122 approaches are capable of determining the nonlinear relationship between variables in a model
123 effectively, without considering any prior assumptions about the problem. Kim et al. (2014)
124 and Hanittinan (2007) used ANNs, and Khoury and Maalouf (2018) employed support vector
125 machine method for prediction of resilient modulus of subgrade soils. Sadrossadat et al. (2018b)
126 used a variant of genetic programming, namely linear genetic programming (LGP) for indirect
127 estimation of M_r of cohesive subgrade soils. Amongst computation intelligence methods,
128 artificial neural networks (ANNs) are more widely used than other methods due to their inherent
129 features, which include managing complex problems with large datasets, handling problems
130 with multiple outputs, and predicting the unseen data effectively. However, ANNs have
131 disadvantages such as slow learning rate and getting stuck in local minima. Furthermore, ANNs
132 are known as black-box systems as they do not provide a distinct relationship between inputs
133 and the output (Ziaee et al., 2015).

134 To improve the prediction capability of ANNs, evolutionary algorithms such as genetic
135 algorithm (GA) and particle swarm optimization (PSO) have been applied to find the optimal
136 values of the weights for the ANN. Evolutionary algorithms can aid ANNs in converging to the
137 global minima, and hence improving the prediction performance of the network. In this regard,
138 Mousavi et al. (2017) proposed a hybrid neural network and simulated annealing for prediction
139 of the daily solar radiation. Alsarraf et al. (2019) applied the PSO-ANN technique for prediction
140 of exergetic performance of a building integrated photovoltaic/thermal system. Mosallanezhad
141 and Moayedi (2017) investigated the potential of an integrated imperialist competitive
142 algorithm ANN to estimate the pull-out resistance of screw piles. While evolutionary
143 algorithms have been found efficient for solving engineering problem, application of these
144 methods in the field of pavement geotechnics is limited to date. To date, there has been limited
145 studies on the application of evolutionary algorithms and hybrid methods for providing
146 formulations of resilient modulus of subgrade soils.

147 This paper proposes the application of GA as well as a hybrid ANN-GA approach for predicting
148 the M_r of cohesive subgrade soils. GA was employed to establish a precise equation for
149 prediction of M_r of subgrade soils. The hybrid ANN-GA model was developed by using GA to
150 determine the optimal values of weights and bias of the ANN-GA approach, which can result
151 in a more robust model. To achieve this aim, a comprehensive and reliable set of data, including
152 the results of RLT tests on cohesive subgrade soils was utilized to develop models. To evaluate
153 the performance of the proposed models, several validation and verification study phases were
154 considered. Furthermore, results were compared with available equations in the literature, to

155 verify the superiority of the proposed models. The developed ANN-GA model was also
156 converted to a tractable formula for hand calculation and pre-design purposes to reduce the time
157 and cost associated with performing RLT tests.

158 **Methodology**

159 *Genetic algorithm (GA)*

160 GA is a heuristic search and optimization algorithm introduced by Holland (1975). GA was
161 inspired by Darwin's theory of evolution and imitates the process of natural evolution and
162 selection. In addition, unlike conventional optimization methods, GA requires less information
163 about the problem and is well suited for more complex problems. In GA, a population of
164 individuals are randomly generated to solve a problem. Each solution is encoded as a fixed-
165 length binary string of 0s and 1s, known a chromosome. A chromosome has several genes and
166 total number of chromosomes indicate the population size (Holland, 1975). The superiority of
167 an individual over other solutions is evaluated using a fitness function. Solutions are evolved
168 in successive iterations (or generations) until a satisfactory criterion, i.e. the maximum number
169 of generations or a predefined fitness value is met. In the process of evolution, several operators,
170 i.e., mutation and crossover are used to modify solutions transferred to next generations (Muduli
171 and Das, 2015). **Fig. 1** summarizes steps taken in the GA for reaching the optimal solution.

172 *Hybrid ANN-GA approach*

173 ANNs are a branch of artificial intelligence techniques that aim to mimic the behavior of human
174 brain and nervous system for solving complex problems. Amongst different variants of ANN,
175 multi-layer perceptron (MLP) neural networks is the most widely used (Cybenko, 1989). A
176 MLP neural networks consists of three distinct layers, i.e. input layer, hidden layer, output layer.
177 Each layer is connected to the subsequent layer through computing elements known as nodes.
178 Except for the input nodes which are fixed nodes, each node is a neuron or processing element
179 with a nonlinear activation function. In a neuron, an input from previous layer is multiplied by
180 a weight coefficient which connects two layers. The output of each neuron is then calculated
181 by passing through a nonlinear activation function (Ziaee et al., 2015). Training of the network
182 is done by adjusting the weights of the network so that the network's predicted values matches
183 the target values in the datasets. To do so, back propagation algorithm is typically used in which
184 the calculated error is back propagated and the network weights are altered accordingly to
185 minimize the prediction error (Ghorbani and Hasanzadehshooiili, 2018).

186 One of the advantages of GA over classic search methods is its ability to perform a global search
187 and hence avoiding the risk of trapping into a local minimum (Gandomi et al., 2013). Thus, a
188 hybrid ANN-GA model uses GA as a powerful search algorithm to find the best parameters for
189 the ANN. In the ANN-GA method, GA is employed to adjust the weights of the ANN in a way
190 that the error of the GA-ANN model (i.e., fitness value) is minimized. A fitness function (i.e.,
191 RMSE) is considered to measure the fitness of each solution vector. In case of reaching the
192 defined termination criterion, the simulation procedure stops and the results are represented.
193 The termination condition is typically set to the maximum value until which the algorithm is
194 iterated to find better solutions.

195 **Database and model variables**

196 The database used for development of ANN-GA model was collected from a study conducted
197 by Hanittinan (2007), which reported on data collected from earlier studies by Kim (2004),
198 Huang (2001), and Rodgers (2006). The database was composed of the physical soil properties,
199 unconfined compressive strengths, and the results of 283 RLT tests performed on cohesive
200 subgrade soils. Soils were mainly silt, clay, and silty clay, classified as A-6 based on the
201 AASHTO soil classification code. Samples were collected from seven different locations and
202 tests were performed in several US universities (Hanittinan, 2007). RLT tests were performed
203 at a range of moisture contents from 4% below optimum moisture content to 3% above optimum
204 moisture content, based on AASHTO designation T294-94 (Hanittinan, 2007).

205 Over the past decades, many studies have examined the factors affecting the M_r of subgrade
206 soil. Results of laboratory studies indicate that the M_r of pavement materials is highly influenced
207 by the stress state parameters, i.e., σ_3 and σ_d . For cohesive subgrade soils, increasing the σ_d at
208 a constant σ_3 decreases the M_r value (Seed et al., 1967, Huang, 2001, Kim, 2004). It is known
209 that increasing the σ_3 increases the M_r of soils (Kim et al., 2014, Sadrossadat et al., 2016). In
210 addition to stress state parameters, several studies have been performed to study the influence
211 of soil physical properties, compaction characteristics, and environmental factors on the M_r .
212 These factors should be considered in the model development for providing more
213 comprehensive and robust models. Moisture content and degree of saturation (S_r) describe the
214 changes in the soil environment and seasonal variations. Soil physical properties such as
215 percentage of soil particles passing through #200 sieve (P_{200}), liquid limit (LL), and plasticity
216 index (PI) are commonly used for classification and identifying the soil type. Besides, UCS has
217 been identified as a static strength parameter which is positively correlated with the M_r (Kim,

218 2004, Hanittinan, 2007). There are several equations in the literature which relate the M_r to
219 strength parameters such as UCS and CBR. On the other hand, compaction characteristics of
220 soils is important factor in design of pavement layers. Herein, soil optimum moisture content
221 (w_{opt}) which is defined as the soil moisture content at its maximum dry density is incorporated
222 in the model development procedure.

223 In addition to aforementioned factors, viscous nature of cohesive soils can cause creep
224 deformations under the applied loads. The developed time dependent shear strains under
225 sustained or repeated loads can reduce the M_r of cohesive soils (Viyanant et al., 2007). In
226 saturated soil conditions, application of load cycles generates excess pore water pressure in
227 cohesive soils which can cause creep deformations (Holzer et al., 1973).

228 With an aim to incorporate most of the influential parameters and considering the available
229 database, the model for prediction of M_r of cohesive subgrade soils is expressed as a function
230 of following parameters:

231

$$M_r = f(LL, PI, P_{200}, w_{opt}, S_r, w, q_u, \sigma_3, \sigma_d) \quad (5)$$

232 To have a better understanding of the model variables, key descriptive statistics of the variables
233 are summarized in **Table 2**.

234 **Data pre-processing**

235 Artificial intelligence techniques utilize data to determine the optimum model that best
236 describes the relationship between input and output parameters. One key issue in finding this
237 relationship is known as overfitting, which needs to be avoided to have a model with better
238 generalization. Overfitting occurs when the model has relatively small error on the trained
239 dataset, while when a new set of data is introduced to the model, error value becomes very high.
240 To avoid overfitting, the dataset needs to be divided into two subsets: 1. Training data and 2.
241 Testing data. The training is used to construct the models and testing data is used to evaluate
242 its performance on unseen data. Accordingly, the testing dataset is used to find a model with
243 better generalization. It is suggested by many researchers that a range of 15-30 % of the data
244 should be used for testing the model performance (Gandomi et al., 2013). It is worth noting that
245 while the results of the model on training data represents the ability of the model to learn the
246 behavior of variables in the database, the test performance indices show the performance on the

247 model on unseen data which indicates its generalization capability. In this study, 80% and 20%
 248 of the datasets are used for training and testing the model, respectively.

249 The number of datasets in database have a significant effect on the model performance in
 250 training stage of the network. The models with larger number of datasets are more reliable and
 251 safer for further analysis. It is suggested that the minimum ratio of datasets in a database to the
 252 variables of the model should be more than five (Gandomi et al., 2013). In this study, this ratio
 253 is equal to 31, which is much higher than the required value.

254 **Model performance assessment**

255 In general, several important criteria need to be checked to evaluate the accuracy and
 256 generalization of the developed ANN-GA model. Correlation coefficient (R), mean absolute
 257 error (MAE), round mean squared error (RMSE) are among essential statistical criteria that
 258 indicate the overall performance of the developed model (Ghorbani et al., 2018):

$$R^2 = \frac{\left(\sum_{i=1}^n (M_{rm} - \bar{M}_{rm})(M_{rp} - \bar{M}_{rp})\right)^2}{\sum_{i=1}^n (M_{rm} - \bar{M}_{rm})^2 \sum_{i=1}^n (M_{rp} - \bar{M}_{rp})^2} \quad (6)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (M_{rp} - M_{rm})^2}{n}} \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |M_{rp} - M_{rm}| \quad (8)$$

259 in which M_{rm} and M_{rp} are the measured and predicted values of the i^{th} output, \bar{M}_{rm} and \bar{M}_{rp}
 260 are the average values of the measured and predicted results, and n is the number of samples.

261 Smith (1986) suggests that there exists a strong correlation between the predicted and measured
 262 values if the correlation coefficient $R^2 \geq 0.64$. Additionally, error values, i.e., MAE and RMSE,
 263 should be minimum for both training and testing data.

264 **Development of GA model**

265 In this section, GA is employed to find a precise equation for prediction of M_r of subgrade soils.
 266 In this regard, a code is written in MATLAB environment (Mathworks, 2017). The input and
 267 output values are normalized before model development to increase the capability of algorithm

268 in finding the relationship between input variables and the output. The database is normalized
 269 to lie between 0 and 1 using following equation:

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (9)$$

270 where X_{\min} , X_{\max} , and X_n are the minimum, maximum, and normalized values of the variable X ,
 271 respectively. The de-normalized value of the output can be calculated as:

$$X = X_n(X_{\max} - X_{\min}) + X_{\min} \quad (10)$$

272 General form of the considered equation for prediction of M_r of subgrade soils is given in **Eq.**
 273 **(11)**:

$$M_r = w_1 LL^{a_1} + w_2 PI^{a_2} + w_3 P_{\#200}^{a_3} + w_4 W_{opt}^{a_4} + w_5 W_c^{a_5} + w_6 S_r^{a_6} + w_7 q_u^{a_7} + w_8 \sigma_3^{a_8} + w_9 \sigma_d^{a_9} + b \quad (11)$$

274 where w_1 to w_9 and a_1 to a_9 are the coefficients of the equation and b is the bias value. GA is
 275 employed as a tool for finding the optimal values of unknown coefficients. RMSE function is
 276 considered as the fitness function to evaluate solutions in each iteration.

277 Several parameters such as population size (N_{pop}), crossover probability, mutation rate, and
 278 maximum number of iterations affect the prediction capability of the GA approach. These
 279 values are typically obtained by trial and error method or using recommended values of other
 280 researchers. The combination of parameters considered to find the optimal model is
 281 summarized in **Table 3**. Several preliminary runs were performed to come up with a parameter
 282 setting to provide a robust model with high generalization capability. Crossover and mutation
 283 rates were selected based on some previously suggested values (Momeni et al., 2014,
 284 Khandelwal and Armaghani, 2016, Rostami et al., 2018). Proper determination of population
 285 size is dependent on the size and the complexity of the investigated problem. To investigate the
 286 effect of population size on the performance of the GA model, a parametric study was done as
 287 shown in **Fig. 2**. It was observed that model with $N_{pop} = 400$ had the lowest RMSE value.
 288 Considering different values for parameters resulted in $2 \times 5 \times 2 \times 2 = 40$ combinations of
 289 parameters. Also, 5 replications of each parameter combination was tested and evaluated. In
 290 total, 200 runs with different combinations of parameters were conducted. The best values of
 291 population size, mutation rate, and crossover probability for the developed GA model were 400,

292 30%, and 70%, respectively. The optimal coefficients of the developed equation based on GA
 293 optimization are summarized in **Table 4**.

294 Considering one of the testing samples (i.e. $LL = 31$, $PI = 12$, $P_{200} = 56$ %, $w_{opt} = 13.4$ %, $w =$
 295 11.4 %, $S_r = 66.44$ %, $q_u = 696.13$ kPa, $\sigma_3 = 41.37$ kPa, $\sigma_d = 40.76$ kPa, $M_r = 112.81$ MPa), the
 296 normalized values of LL , PI , P_{200} , w_{opt} , w , S_r , q_u , σ_3 , and σ_d are equal to 0.617, 0.82, 0, 0.476,
 297 0.337, 0.412, 1, 0.999, 0.494, and 0.602, respectively. By substituting the obtained coefficient
 298 values in **Eq. (11)**, M_r is calculated equal to 0.6076. Using **Eq. (10)**, the de-normalized value
 299 of M_r is calculated equal to 113.7 (MPa).

300 **Hybrid ANN-GA model development**

301 In order to develop the ANN-GA model for prediction of resilient modulus of subgrade soils, a
 302 code is written in MATLAB environment (Mathworks, 2017). The procedure for modeling of
 303 M_r using ANN-GA is illustrated in **Fig. 3**. The model includes nine input variables and M_r is
 304 the only output of the model.

305 In the first phase of constructing the model, the dataset should be normalized between -1 and 1
 306 to facilitate model development using following equation (Mathworks, 2017):

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} - \frac{X_{\max} - X}{X_{\max} - X_{\min}} \quad (12)$$

307 Considering **Eq. (13)**, the de-normalized value of the output can be calculated as:

$$X = 0.5(X_n + 1)(X_{\max} - X_{\min}) + X_{\min} \quad (13)$$

308 In the hybrid ANN-GA model development, there are several parameters in both the ANN and
 309 GA that need to set. The precision of the ANN models increases with increasing the number of
 310 nodes in the hidden layer. However, increasing the nodes would result in a more complicated
 311 model with many different parameters. Cybenko (1989) indicated that a single hidden layer
 312 neural network would provide satisfactory results for approximation of nonlinear problems.
 313 Thus, in this study, only one hidden layer consisting of three nodes was considered to develop
 314 a simple model that can be represented as a tractable formulation rather than a complicated
 315 black-box. The activation function for the hidden layer nodes was Tangent-sigmoid as follows
 316 (Mathworks, 2017):

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (14)$$

317 The crossover and mutation rates were chosen equal to the values adopted for developing the
 318 GA model. In addition, a parametric study was performed by developing several ANN-GA
 319 models to find the optimal value of N_{pop} . Increasing population size usually increases the chance
 320 of obtaining better results, however, it decreases the speed of the model development. The
 321 population size was varied from 50 to 400 and maximum number of generations was set to 500.
 322 **Fig. 4** shows the effect of population size on the model performance based on the RMSE criteria.
 323 As evident, the model with $N_{pop} = 400$ had the lowest RMSE value. It is also evident that after
 324 about 300 generations, there is no significant change in the RMSE value. The hybrid ANN-GA
 325 model was run several times with $3 \times 5 \times 2 \times 2 = 60$ different combinations of the parameters
 326 and 5 replications for each combination which resulted in a total of 300 runs. The parameter
 327 settings during the ANN-GA model development are summarized in Table 5.

328 The best ANN-GA model for prediction of resilient modulus of cohesive subgrade soils has the
 329 following parameters: Population size = 400; Mutation rate = 10%; Crossover rate = 70%. In
 330 order to transform the optimal ANN-GA model into a tractable formulation for further analysis,
 331 following function is used (Ziaee et al., 2015):

$$h = f_{HO} \left(bias_h + \sum_{k=1}^h V_k f_{IH} \left(bias_{hk} + \sum_{i=1}^m w_{ik} x_i \right) \right) \quad (15)$$

332 where $bias_h$ is the hidden layer bias; V_k is the weight connection between neuron k of the hidden
 333 layer and the single output neuron; $bias_{hk}$ would be the bias at neuron k of the hidden layer (k
 334 $= 1, h$); w_{ik} denotes the weight connection between the input variable ($i=1, m$) and neuron k of
 335 the hidden layer; x_i would be the i^{th} input parameter; f_{HO} is the transfer function between the
 336 hidden layer and the output layer; and f_{IH} is the transfer function between the input and hidden
 337 layer.

338 After de-normalization of the output, the optimal ANN-GA model for prediction of M_r of
 339 cohesive subgrade with nine inputs ($LL, PI, P_{200}, w_{opt}, w, S_r, q_u, \sigma_3, \sigma_d$) can be expressed as:

$$(M_{r_i})_{ANN-GA} = 8377 (M_{r_i})_n + 119 \quad (16)$$

340 where,

$$(M_{r_i})_n = \sum_{k=1}^3 V_k \tanh(A_k) + bias_n \quad (17)$$

$$A_k = w_{1k}LL_n + w_{2k}PI_n + w_{3k}P_{200n} + w_{4k}W_{optn} + w_{5k}W_{cn} + w_{6k}S_{rn} + w_{7k}q_{un} + w_{8k}\sigma_{3n} + w_{9k}\sigma_{dn} + bias_k \quad (18)$$

341 where LL_n , PI_n , P_{200n} , W_{optn} , W_{cn} , S_{rn} , q_{un} , σ_{3n} , and σ_{dn} are the normalized input values obtained
 342 from **Eq. (12)**, and k is the number of hidden layer nodes (i.e., 3). The obtained values of bias
 343 and weights of the optimal model for input-hidden and hidden-output layers are summarized in
 344 **Table 6** and **Table 7**, respectively.

345 Considering the same testing sample used for calculating the output of the GA model ($LL = 31$,
 346 $PI = 12$, $P_{200} = 56$ %, $W_{opt} = 13.4$ %, $W_c = 11.4$ %, $S_r = 66.44$ %, $q_u = 696.13$ kPa, $\sigma_3 = 41.37$
 347 kPa, $\sigma_d = 40.76$ kPa, $M_r = 112.81$ MPa), the procedure for calculating the M_r is described as:

348 **Step 1:** Normalization of the input dataset. Using **Eq. (12)**, the normalized values of LL , PI ,
 349 P_{200} , W_{opt} , W_c , S_r , q_u , σ_3 , and σ_d are 0.235, 0.64, -1, -0.048, -0.325, -0.176, 1, 0.99, -0.011,
 350 respectively.

351 **Step 2:** Calculation of hidden nodes parameters. Using the values summarized in **Table 6**, the
 352 values of A_1 to A_3 are calculated as: $A_1 = 1.522$; $A_2 = -1.73$; $A_3 = -2.89$.

353 **Step 3:** Prediction of the M_{rn} . The output of each hidden layer neuron is calculated by passing
 354 through an activation function (i.e., tansig). M_{rn} can then be calculated as the summation of
 355 output of each neuron which is multiplied to the hidden-output weights summarized in **Table**
 356 **7**.

357 **Step 4:** De-normalization of the M_r . Using **Eq. (13)**, M_r is calculated in the range of datasets.
 358 The output of the ANN-GA model is obtained equal to 110.46 MPa.

359 **Performance analysis of ANN-GA and GA models**

360 In order to represent the capability of the obtained ANN-GA and GA models, the predicted
 361 versus measured values of M_r for training and testing datasets are depicted in **Fig. 5 (a)** and **(c)**,
 362 respectively. In addition, histograms of the errors obtained by each method for training and
 363 testing datasets is illustrated in **Fig. 5 (b)** and **(d)** to have a general view of the frequency of the
 364 errors in different intervals.

365 As evident in these figures, the ANN-GA model has R^2 values of 0.97 for both training and
366 testing datasets, and RMSE values are 5.5 and 5.2 for training and testing datasets, respectively.
367 On the other hand, the R^2 of GA model for both training and testing datasets is equal to 0.87.
368 The RMSE value for training and testing datasets are 11.1 and 11.3 for training and testing data,
369 respectively. Considering the high R^2 values and low RMSE values, it can be concluded that
370 the developed models are capable of predicting the M_r of subgrade soils with a high degree of
371 accuracy. In addition, close values of R^2 and RMSE values for training and testing datasets
372 indicate that overfitting is avoided (Ghorbani et al., 2018, Sadrossadat et al., 2018a). This means
373 that the proposed models would have a satisfactory performance on unseen data and thus has
374 suitable generalization performance. It should be noted that while the functional structure of
375 the GA model is simpler, the ANN-GA model outperforms GA model with a high degree of
376 accuracy.

377 **Comparative study**

378 To further examine the performance of the proposed models, their performance is compared
379 with the LGP model (Sadrossadat et al., 2018b) and model proposed by Kim (2004). The Kim
380 model was chosen for the purpose of comparison as it outperformed other existing regression
381 based models in the literature (Kim, 2004). In this regard, testing datasets are used to evaluate
382 the prediction capability of all methods. Results of comparative study is represented in **Fig. 6**.
383 Residual error (RE) is calculated as the difference between the measured and predicted values
384 of each model. As evident in **Fig. 6**, the ANN-GA approach outperforms other three methods
385 in terms of all indicators. The R^2 values for ANN-GA, GA, LGP, and Kim's model are 0.97,
386 0.87, 0.83, and 0.56, respectively. The max $|RE|$ of the ANN-GA model is about 11, while this
387 value for GA, LGP, and Kim's model are around 29, 33, and 62, respectively. Furthermore, it
388 is evident that AI-based methods perform notably better than the regression based model
389 proposed by Kim (2004).

390 While R, MAE and RMSE together give an overall perspective of the performance of each
391 model, other criteria need to be satisfied on the testing datasets to ensure the external validation
392 of a prediction model. In this regard, a ranking index (RI) is used in this study that incorporates
393 some other statistical criteria to compare the performance of the models (Ghorbani et al., 2018).
394 This statistical procedure is based on the following four criteria:

- 395 1. The equation of best fine line of predicted (M_{rp}) versus measured (M_{rm}) resilient
 396 modulus (M_{r-fit}/M_{rm}), along with corresponding coefficient of determination (R^2). It is
 397 worth noting that model with (M_{r-fit}/M_{rm}) and R^2 closer to 1 has the best performance.
- 398 2. The arithmetic mean (μ) and standard deviation (σ) of M_{rp}/M_{rm} . It is suggested that a
 399 model with μ closer to 1 and σ closer to 0 has a better performance in prediction of the
 400 M_r .
- 401 3. The 50% cumulative probability ($P_{50\%}$) of M_{rp}/M_{rm} . To calculate the $P_{50\%}$, the values of
 402 M_{rp}/M_{rm} are arranged in an ascending order, and the cumulative probability is calculated
 403 using following equation:

$$P = \frac{i}{n+1} \quad (19)$$

404 The value of the $P_{50\%}$ for the optimal model should be close to one.

- 405 4. The coefficient of efficiency (E). This parameter evaluates how well each model
 406 describes the variance of the datasets. E can be calculated using following equations:

$$E = \frac{E_1 - E_2}{E_1} \quad (20)$$

$$E_1 = \sum_{i=1}^n (M_{rm} - \bar{M}_{rp})^2 \quad (21)$$

$$E_2 = \sum_{i=1}^n (M_{rp} - M_{rm})^2 \quad (22)$$

407 The overall performance of the each model model can be measured in terms of rank index (RI)
 408 which is the sum of the ranks of each sub criteria:

$$RI = R_1 + R_2 + R_3 + R_4 \quad (23)$$

409 where R_1 , R_2 , R_3 , and R_4 are the ranks from each of the explained sub criteria. The model with
 410 the lowest RI has the best performance for prediction of M_r of cohesive subgrade soils. The best
 411 fit line of M_{rp}/M_{rm} for each investigated model is illustrated in **Fig. 7**. As evident, the ANN-GA
 412 model has the best performance among all models, followed by GA, LGP and Kim model.
 413 Therefore, ANN-GA model is ranked 1 (i.e. $R_1 = 1$) based on the first criterion. While AI-based

414 models are highly capable of predicting the M_r , Kim's model tends to underestimate the M_r
 415 values. The results of the first criterion are summarized in the column corresponding to R_I in
 416 **Table 8**.

417 Considering the cumulative probability criteria, the plots of M_{rp}/M_{rm} versus cumulative
 418 probability (%) are illustrated in **Fig. 8**. As evident from this figure, more than 95% of the
 419 predicted values by the ANN-GA model have the M_{rp}/M_{rm} values in a range of 0.8-1.2. The $P_{50\%}$
 420 for the ANN-GA, ANN, LGP, and Kim's model are 1, 1, 1, and 0.8. A similar trend can be
 421 found in all other criteria as summarized in **Table 8**. It can be concluded that ANN-GA
 422 approach is highly capable of predicting the M_r of subgrade soils considering several validation
 423 criteria, followed by GA model and the LGP model. Generally, AI-based methods perform
 424 notably better compared to the regression based method by Kim (2004).

425 **Sensitivity analysis**

426 To measure the relative importance of each input variable and its contribution to the final model,
 427 a sensitivity analysis (SA) is performed. SA aims to measure the strength of the relationship
 428 between model inputs and the output variable. In this study, Cosine amplitude method (CAM)
 429 is used as an indicator of significance of each input variable (Majdi and Rezaei, 2013).
 430 Considering a set of n data samples in the common X-space, a data array X can be defined as:

$$X = \{X_1, X_2, \dots, X_m\} \quad (24)$$

431 Each of the elements, x_i , in the data array X is a vector of lengths of m, that is:

$$X_i = \{X_{i1}, X_{i2}, \dots, X_{im}\}_i \quad (25)$$

432 Each data sample can therefore be regarded as a point in m-dimensional space, where each point
 433 requires m coordinates for full description. Each element of relation, r_{ij} is the result of a pairwise
 434 comparison of two data samples, x_i and x_j . The strength of the relation between these two data
 435 pairs is in a 0 to 1 scale and is expressed by the following equation (Sadrossadat et al., 2016):

$$r_{ij} = \frac{\sum_{k=1}^m x_{ik} \cdot x_{jk}}{\sqrt{\sum_{k=1}^m x_{ik}^2 \cdot \sum_{k=1}^m x_{jk}^2}} \quad (26)$$

436 **Fig. 9** describes the relative importance of each input variables in the developed models. The
 437 closer the r_{ij} to 1, the more impact corresponding variable has on the M_r value. As evident, the
 438 r_{ij} values for all input variables of both methods are between 0.75 and 0.92, which indicates the

439 significance of all input variables for model development. In other words, all input variables
440 are approximately equally important for prediction of M_r of subgrade soils, and their importance
441 for model development cannot be neglected. These results are in agreement with those of
442 similar studies in the literature (Sadrossadat et al., 2016, Kim, 2004, Sadrossadat et al., 2018b).

443 **Conclusions**

444 In this study, potential of two intelligent methods, i.e. GA and ANN-GA, was evaluated for
445 prediction of M_r of cohesive subgrade soils. GA was used to develop an equation for prediction
446 of M_r of subgrade soils. In addition, a GA was utilized to enhance the predictive capability of
447 the ANN by adjusting the weights and bias.

448 A comprehensive database was utilized for the development of the models. Nine input
449 parameters including liquid limit (LL), plastic index (PI), percentage of soil particles passing
450 through #200 sieve (P_{200}), optimum moisture content (w_{opt}), degree of saturation (S_r), moisture
451 content (w), unconfined compressive strength (q_u), confining stress (σ_3), and deviator stress (σ_d)
452 were used for model development. The prediction performance of the developed models were
453 compared with existing prediction equations in the literature. A ranking index was used to
454 evaluate the external capability of the proposed model.

455 Results of this study indicates that both GA and ANN-GA methods can be employed as efficient
456 tools in predicting the M_r of cohesive soils. The R^2 of the predicted and measured values for
457 the ANN-GA model and GA model was 0.97 and 0.87 for both training and testing datasets,
458 which was superior to the available prediction equations. One of the main objectives of the
459 present study was to express the fact that the ANN-GA model can be expressed as explicit
460 formula that can be used for manual calculation purposes. Besides, considering the prediction
461 capability of the developed GA model, it can be concluded that the evolutionary algorithms can
462 be regarded as efficient tools for providing precise and simple equations. The results obtained
463 by ANN-GA and GA models outperformed other existing equations in the literature in terms of
464 precision and accuracy considering several validation criteria.

465 The obtained results of sensitivity analysis indicated the importance of all input variables for
466 prediction of M_r of cohesive subgrade soils. It should be noted that the capability of the AI-
467 based methods is mostly limited to the range, number and statistical features of the database
468 used for model developments. To resolve this, developed models can be enhanced by increasing
469 the number of datasets. Nonetheless, the models provided in this study could be used for

470 estimation of M_r of subgrade soils without conducting any tests. The proposed models are
471 expected to be useful for preliminary design stages or when the testing is not feasible.

472 **Acknowledgements**

473 This research was supported under Australian Research Council's Linkage Projects funding
474 scheme (project number LP170100072).

475

476

477

478

479

480 **References**

481 AASHTO 2003. AASHTO T307, Determining the Resilient Modulus of Soils and Aggregate Materials,
482 Standard Specifications for Transportation Materials and Methods of Sampling and Testing,
483 Washington, D. C.

484 Alsarraf, J., Moayed, H., Rashid, A. S. A., Muazu, M. A. & Shahsavar, A. 2019. Application of PSO-ANN
485 modelling for predicting the exergetic performance of a building integrated
486 photovoltaic/thermal system. *Engineering with Computers*, 1-14.

487 Carmichael III, R. & Stuart, E. 1985. Predicting resilient modulus: A study to determine the mechanical
488 properties of subgrade soils (abridgment). *Transportation Research Record*.

489 Chen, R.-P., Zhang, P., Kang, X., Zhong, Z.-Q., Liu, Y. & Wu, H.-N. 2019. Prediction of maximum surface
490 settlement caused by earth pressure balance (EPB) shield tunneling with ANN methods. *Soils
491 and Foundations*.

492 Cybenko, G. 1989. Approximations by superpositions of a sigmoidal function. *Mathematics of Control,
493 Signals and Systems*, 2, 183-192.

494 Gandomi, A. H., Yang, X.-S., Talatahari, S. & Alavi, A. H. 2013. Metaheuristic algorithms in modeling and
495 optimization. *Metaheuristic applications in structures and infrastructures*, 1-24.

496 Ghorbani, A. & Hasanzadehshooili, H. 2018. Prediction of UCS and CBR of microsilica-lime stabilized
497 sulfate silty sand using ANN and EPR models; application to the deep soil mixing. *Soils and
498 foundations*, 58, 34-49.

499 Ghorbani, B., Sadrossadat, E., Bazaz, J. B. & Oskoei, P. R. 2018. Numerical ANFIS-based formulation
500 for prediction of the ultimate axial load bearing capacity of piles through CPT data.
501 *Geotechnical and Geological Engineering*, 36, 2057-2076.

502 Hanittinan, W. 2007. Resilient modulus prediction using neural network algorithm. The Ohio State
503 University.

504 Holland, J. H. 1975. *Adaptation in natural and artificial systems* Ann Arbor. The University of Michigan
505 Press, 1, 975.

506 Holzer, T. L., Höeg, K. & Arulanandan, K. 1973. Excess pore pressures during undrained clay creep.
507 *Canadian Geotechnical Journal*, 10, 12-24.

508 Huang, J. 2001. Degradation of resilient modulus of saturated clay due to pore water pressure buildup
509 under cyclic loading. The Ohio State University.

510 Jayawardana, P., Thambiratnam, D. P., Perera, N., Chan, T. & De Silva, G. S. 2019. Use of artificial neural
511 network to evaluate the vibration mitigation performance of geofoam-filled trenches. *Soils
512 and Foundations*, 59, 874-887.

513 Khandelwal, M. & Armaghani, D. J. 2016. Prediction of drillability of rocks with strength properties
514 using a hybrid GA-ANN technique. *Geotechnical and Geological Engineering*, 34, 605-620.

515 Khoury, N. & Maalouf, M. 2018. Prediction of Resilient Modulus from Post-compaction Moisture
516 Content and Physical Properties Using Support Vector Regression. *Geotechnical and Geological*
517 *Engineering*, 36, 2881-2892.

518 Kim, D.-G. 2004. Development of a constitutive model for resilient modulus of cohesive soils. The Ohio
519 State University.

520 Kim, S.-H., Yang, J. & Jeong, J.-H. 2014. Prediction of subgrade resilient modulus using artificial neural
521 network. *KSCCE Journal of Civil Engineering*, 18, 1372-1379.

522 Majdi, A. & Rezaei, M. 2013. Prediction of unconfined compressive strength of rock surrounding a
523 roadway using artificial neural network. *Neural Computing and Applications*, 23, 381-389.

524 Mathworks, I. 2017. MATLAB and statistics toolbox release 2017b. Natick (Massachusetts, United
525 States).

526 Mohammad, L. N., Herath, A., Abu-Farsakh, M. Y., Gaspard, K. & Gudishala, R. 2007. Prediction of
527 resilient modulus of cohesive subgrade soils from dynamic cone penetrometer test
528 parameters. *Journal of Materials in Civil Engineering*, 19, 986-992.

529 Mohammad, L. N., Titi, H. H. & Herath, A. 2002. Effect of moisture content and dry unit weight on the
530 resilient modulus of subgrade soils predicted by cone penetration test. Louisiana
531 Transportation Research Center.

532 Momeni, E., Nazir, R., Armaghani, D. J. & Maizir, H. 2014. Prediction of pile bearing capacity using a
533 hybrid genetic algorithm-based ANN. *Measurement*, 57, 122-131.

534 Mosallanezhad, M. & Moayedi, H. 2017. Developing hybrid artificial neural network model for
535 predicting uplift resistance of screw piles. *Arabian Journal of Geosciences*, 10, 479.

536 Mousavi, S. M., Mostafavi, E. S. & Jiao, P. 2017. Next generation prediction model for daily solar
537 radiation on horizontal surface using a hybrid neural network and simulated annealing
538 method. *Energy conversion and management*, 153, 671-682.

539 Muduli, P. K. & Das, S. K. 2015. Model uncertainty of SPT-based method for evaluation of seismic soil
540 liquefaction potential using multi-gene genetic programming. *Soils and Foundations*, 55, 258-
541 275.

542 Puppala, A. J., Mohammad, L. N. & Allen, A. 1996. Engineering behavior of lime-treated Louisiana
543 subgrade soil. *Transportation Research Record*, 1546, 24-31.

544 Rani, D., Jain, S. K., Srivastava, D. K. & Perumal, M. 2013. Genetic algorithms and their applications to
545 water resources systems. *Metaheuristics in Water, Geotechnical and Transport Engineering*,
546 43.

547 Rodgers, G. A. 2006. Resilient modulus predictions using engineering properties and neural networks.
548 Ohio State University.

549 Rostami, M. F., Sadrossadat, E., Ghorbani, B. & Kazemi, S. M. 2018. New empirical formulations for
550 indirect estimation of peak-confined compressive strength and strain of circular RC columns
551 using LGP method. *Engineering with Computers*, 34, 865-880.

552 Sadrossadat, E., Ghorbani, B., Zohourian, B., Kaboutari, M. & Rahimzadeh Oskooei, P. 2018a. Predictive
553 modelling of the MR of subgrade cohesive soils incorporating CPT-related parameters through
554 a soft-computing approach. *Road Materials and Pavement Design*, 1-19.

555 Sadrossadat, E., Heidariapanah, A. & Ghorbani, B. 2018b. Towards application of linear genetic
556 programming for indirect estimation of the resilient modulus of pavements subgrade soils.
557 *Road Materials and Pavement Design*, 19, 139-153.

558 Sadrossadat, E., Heidariapanah, A. & Osouli, S. 2016. Prediction of the resilient modulus of flexible
559 pavement subgrade soils using adaptive neuro-fuzzy inference systems. *Construction and*
560 *Building Materials*, 123, 235-247.

561 Schuettpelz, C., Fratta, D. & Edil, T. 2010. Mechanistic corrections for determining the resilient modulus
562 of base course materials based on elastic wave measurements. *Journal of geotechnical and*
563 *geoenvironmental engineering*, 136, 1086-1094.

564 Seed, H., Mitry, F., Monismith, C. & Chan, C. 1967. Prediction of flexible pavement deflections from
565 laboratory repeated-load tests. NCHRP report.
566 Smith, G. N. 1986. Probability and statistics in civil engineering. Collins Professional and Technical
567 Books, 244.
568 Viyanant, C., Rathje, E. M. & Rauch, A. F. 2007. Creep of compacted recycled asphalt pavement.
569 Canadian Geotechnical Journal, 44, 687-697.
570 Witczak, M. & Uzan, J. 1988. The universal airport pavement design system. Report I of V: granular
571 material characterization, Department of Civil Engineering, University of Maryland, College
572 Park, MD.
573 Ziaee, S. A., Sadrossadat, E., Alavi, A. H. & Shadmehri, D. M. 2015. Explicit formulation of bearing
574 capacity of shallow foundations on rock masses using artificial neural networks: application
575 and supplementary studies. Environmental earth sciences, 73, 3417-3431.

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

LIST OF TABLES

593 **Table 1.** Available prediction equations for M_r of subgrade soils
 594 **Table 2.** Descriptive statistics of the variables in the database
 595 **Table 3.** Parameter settings for development of GA model
 596 **Table 4.** Optimal coefficients of the developed GA model
 597 **Table 5.** Parameter settings for development of ANN-GA model
 598 **Table 6.** Values of weights and bias for input-hidden layer
 599 **Table 7.** Values of the weights and bias of the hidden-output layer
 600 **Table 8.** Results of statistical analysis of models

601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612

Table 1
 Available prediction equations for M_r of subgrade soils

| Reference | Equation | Affecting variables | Applicability |
|-----------|----------|---------------------|---------------|
|-----------|----------|---------------------|---------------|

| | | | |
|------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------|----------------------------------|
| Seed et al. (1967) | $M_r = k_1 \left(\frac{\theta}{P_a} \right)^{k_2}$ | θ, P_a, k_2 | All soil types |
| Witczak and Uzan (1988) | $M_r = k_1 P_a \left(\frac{\theta}{P_a} \right)^{k_2} \left(\frac{\tau_{oct}}{P_a} \right)^{k_3}$ | $\theta, P_a, \tau_{oct}, k_2, k_3$ | All soil types |
| Puppala et al. (1996) | $M_R = k_1 P_a \left(\frac{\sigma_3}{P_a} \right)^{k_2} \left(\frac{\sigma_d}{P_a} \right)^{k_3}$ | $\sigma_3, \sigma_d, P_a, k_2, k_3$ | All soil types |
| Kim (2004) | $\frac{M_R}{P_a} = k_1 \left[\frac{P_a \sigma_{oct}}{\tau_{oct}^2} \right]^{k_2}$ | $\sigma_{oct}, \tau_{oct}, P_a, k_1, k_2$ | A-4 and A-6 soils |
| Carmichael and Stuart (1985) | $M_r = 37.431 - 0.456PI - 0.617w - 0.142P_{200} + 0.179\sigma_3 - 0.324\sigma_d + 3.642CH + 17.09MH$ | $PI, w, P_{200}, \sigma_3, \sigma_d, MH, CH$ | Cohesive subgrade soils |
| AASHTO (2003) | $M_r (MPa) = 17.6(CBR)^{0.62}$ | CBR | All soil types |
| Mohammad et al. (2007) | $M_R = \frac{165.5}{DCPI^{1.147}} + 0.0966 \left(\frac{\gamma_d}{w} \right)$ | $DCPI, \gamma_d, w$ | A-4, A-6, A-7-5, and A-7-6 soils |
| | $M_R = \frac{151.8}{DCPI^{1.096}}$ | $DCPI$ | |
| Mohammad et al. (2002) | $\frac{M_R}{\sigma_3^{0.55}} = \frac{1}{\sigma_1} \left(31.79q_c + 74.81 \frac{f_s}{w} \right) + 4.08 \frac{\gamma_d}{\gamma_w}$ | $\sigma_3, \sigma_d, q_c, f_s, \gamma_d, \gamma_w$ | Cohesive subgrade soils |
| Sadrossadat et al. (2018b) | $M_R = \sigma_3 + W_{opt} + (3P_{\#200} + q_u + \sigma_3 - 2\sigma_d - (5LL \times (W_c + S_r + 8\sigma_3 + \sigma_d - 8(PI - W_{opt} + 1.5)^4)) / q_u) / W$ | $w_{opt}, P_{200}, q_u, \sigma_3, \sigma_d, LL, w, S_r$ | A-6 soils |

613 θ : bulk stress ($[\sigma_1 + 2\sigma_3]$); P_a : atmospheric pressure (=101 kPa); σ_{oct} : octahedral normal stress ($[\sigma_1 +$
614 $2\sigma_3]/3$); τ_{oct} : octahedral shear stress ($[2^{0.5}(\sigma_1 - \sigma_3)]/3$); σ_1 : major principal stress ($[\sigma_3 + \sigma_d]$); σ_3 : minor
615 principal stress; σ_d : deviatoric stress ($[\sigma_1 - \sigma_3]$); PI : plasticity index; w : moisture content; P_{200} :
616 percentage of soil particles passing through #200 sieve; CBR : California bearing ratio; $DCPI$: dynamic
617 cone penetration index; γ_d : dry density; q_c : cone tip resistance; f_s : sleeve friction resistance; q_u :
618 unconfined compressive strength; S_r : degree of saturation; k_1, k_2, k_3 : regression coefficients; $CH=1$ for
619 CH soil; $MH = 1$ for MH soil.

620

621

622

623

624

Table 2
Descriptive statistics of the variables in the database

| Parameter | <i>LL</i> | <i>PI</i> | <i>P</i> ₂₀₀ (%) | <i>w</i> _{opt} (%) | <i>S</i> _r (%) | <i>w</i> (%) | <i>q</i> _u (kPa) | σ_3 (kPa) | σ_d (kPa) | <i>M</i> _r (MPa) |
|--------------------|-----------|-----------|--------------------------------|--------------------------------|---------------------------|-----------------|--------------------------------|---------------------|---------------------|--------------------------------|
| Mean | 30.83 | 11.17 | 77.74 | 14.01 | 81.13 | 13.90 | 324.87 | 20.95 | 41.45 | 65.34 |
| Median | 32.00 | 11.00 | 84.00 | 14.00 | 85.07 | 14.00 | 316.50 | 20.69 | 41.37 | 64.83 |
| Mode | 31.00 | 11.00 | 56.00 | 13.40 | 66.44 | 14.00 | 696.13 | 0.00 | 41.37 | 104.00 |
| Standard deviation | 4.31 | 1.72 | 15.48 | 2.03 | 14.23 | 2.84 | 178.70 | 16.36 | 18.12 | 31.22 |
| Kurtosis | 0.65 | 1.43 | -1.40 | 0.97 | -0.24 | -0.61 | -0.29 | -1.41 | -1.05 | -0.14 |
| Skewness | -1.10 | -1.37 | -0.15 | -0.85 | -0.77 | -0.39 | 0.76 | -0.03 | -0.02 | 0.29 |
| Range | 16.20 | 6.10 | 44.00 | 8.40 | 57.08 | 11.47 | 607.18 | 41.40 | 60.21 | 167.54 |
| Minimum | 21.00 | 7.00 | 56.00 | 9.40 | 42.92 | 7.53 | 88.95 | 0.00 | 11.00 | 11.90 |
| Maximum | 37.20 | 13.10 | 100.00 | 17.80 | 100.00 | 19.00 | 696.13 | 41.40 | 71.21 | 179.44 |

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

Table 3
Parameter settings for development of GA model

| Parameter | Setting |
|----------------------|------------------------|
| Number of generation | 500, 1000 |
| Population size | 50, 100, 200, 300, 400 |
| Mutation rate (%) | 10, 30 |
| Crossover rate (%) | 70, 95 |
| Fitness function | RMSE |

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

Table 4
Optimal coefficients of the developed GA model

| Coefficients | Coefficient number | | | | | | | | | bias (<i>b</i>) |
|--------------|--------------------|--------|--------|-------|--------|--------|-------|-------|--------|----------------------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| w_i | 0.103 | -0.204 | -0.131 | 0.310 | -0.085 | -0.273 | 0.351 | 0.168 | -0.292 | 0.726 |
| a_i | 5.999 | 0.023 | 0.961 | 3.096 | 0.045 | 0.862 | 0.640 | 0.723 | 0.148 | |

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

Table 5

Parameter settings for development of ANN-GA model

| Parameter | Setting |
|-------------------------|------------------------|
| Number of generation | 100, 300, 500 |
| Population size | 50, 100, 200, 300, 400 |
| Mutation rate (%) | 10, 30 |
| Crossover rate (%) | 70, 95 |
| Number of hidden layers | 1 |
| Number of hidden nodes | 3 |
| Activation function | Tangent-sigmoid |
| Fitness function | RMSE |

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

Table 6
 Values of weights and bias for input-hidden layer

| Weight (w_{ik}) | Neuron (k) Between Input and Hidden Layer | | |
|---------------------|-----------------------------------------------|--------|--------|
| | 1 | 2 | 3 |
| W_{1k} | 0.519 | 0.171 | -1.909 |
| W_{2k} | 0.003 | -0.196 | -1.203 |
| W_{3k} | -0.252 | 0.125 | 0.640 |
| W_{4k} | 0.778 | 0.522 | -0.867 |
| W_{5k} | -0.541 | -1.093 | 1.409 |
| W_{6k} | -0.245 | 0.138 | 0.960 |
| W_{7k} | 0.141 | 0.103 | 0.921 |
| W_{8k} | 0.035 | -0.363 | -0.105 |
| W_{9k} | -0.148 | -0.067 | 0.703 |
| $bias_k$ | 0.787 | -1.563 | -1.256 |

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

Table 7

Values of the weights and bias of the hidden-output layer

| Weight | Neuron (k) Between Hidden and Output Layer | | | $bias_h$ |
|--------|------------------------------------------------|---------|--------|----------|
| | 1 | 2 | 3 | |
| W_k | 2.6376 | -1.9416 | 1.6468 | -2.4069 |

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

Table 8

Results of statistical analysis of models

| Method | Overall performance | | | Arithmetic calculation of M_{rp}/M_{rm} | | | Cumulative probability | | Coefficient of efficiency | | Overall rank | |
|----------------------------|---------------------|-------|-------|-------------------------------------------|------|-------|-----------------------------|-------|---------------------------|-------|--------------|------------|
| | M_{fit}/M_{rm} | R^2 | R_1 | Mean | SD | R_2 | M_{rp}/M_{rm} at P_{50} | R_3 | E | R_4 | RI | Final rank |
| ANN-GA | 1.01 | 0.97 | 1 | 1.01 | 0.09 | 1 | 1 | 1 | 0.97 | 1 | 4 | 1 |
| GA | 1.01 | 0.87 | 2 | 1.04 | 0.30 | 2 | 1 | 1 | 0.85 | 2 | 8 | 2 |
| Sadrossadat et al. (2018b) | 0.98 | 0.83 | 3 | 1.06 | 0.33 | 3 | 1 | 1 | 0.82 | 3 | 11 | 3 |
| Kim (2004) | 0.78 | 0.56 | 4 | 0.77 | 0.26 | 4 | 0.8 | 4 | 0.45 | 4 | 16 | 4 |

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

LIST OF FIGURES

755 **Fig. 1.** Flowchart of GA for obtaining the optimal solution (Rani et al., 2013)

756 **Fig. 2.** Effect of the population size on development of the GA model

757 **Fig. 3.** Hybrid ANN-GA model development procedure

758 **Fig. 4.** Effect of the population size on development of the ANN-GA model

759 **Fig. 5. (a) and (c)** Measured versus predicted M_r values of the GA and ANN-GA models; **(b)**

760 and **(d)** Histogram of errors obtained for training and testing data in GA and ANN-GA

761 models

762 **Fig. 6.** Measured values of M_r and those obtained by (a) ANN-GA (b) GA (c) LGP

763 (Sadrossadat et al., 2018b) (d) Kim (Kim, 2004) method

764 **Fig. 7.** Best fit line of M_{rp}/M_{rm} for (a) ANN-GA model (b) GA model (c) LGP model (d) Kim

765 model

766 **Fig. 8.** Plots of M_{rp}/M_{rm} versus cumulative probability (%) for different models

767 **Fig. 9.** Relative importance of the input variable in the developed models

768



















