Peak Shear Strength of Discrete Fiber-Reinforced Soils Computed by Machine Learning and Metaensemble Methods

Jui-Sheng Chou[1]; Kuo-Hsin Yang[2]; and Jie-Ying Lin[3]

Abstract: The accuracy of prior theoretical and empirical models for predicting the shear strength of fiber-reinforced soil (FRS) is questionable because of the difficulty of using these simplified models to describe the complex mechanism of soil-fiber interaction. This study compiled a large database of available high quality triaxial and direct shear tests on FRS documented in the literature from 1983 to 2015. The database includes information on the properties of sand, fibers, soil-fiber interface, and stress parameters. After data preprocessing, data mining technologies were employed to identify factors influencing shear strength and to predict the peak friction angle of FRS. The analysis techniques included (1) classification and regression methods, i.e., linear regression (REG) analysis, classification and regression tree (CART) analysis, a generalized linear (GENLIN) model, and chi-squared automatic interaction detection (CHAID); (2) machine learners, i.e., artificial neural network (ANN) and support vector machine (SVM) and support vector regression (SVR); and (3) metaensemble models, i.e., voting, bagging, stacking, and tiering. The analytical results indicated that fiber content, fiber aspect ratio, soil friction angle, and stress parameter had major effects on FRS shear strength. The optimal model obtained after further model training, cross-validation, and testing was the Tiering SVM-(SVR/SVR) method. The correlation coefficient (R) of the prediction values with the measured values in the database was 0.89, indicating a strong association. The mean absolute percentage error (MAPE) was 3.27%, root mean square error (RMSE) was 1.98°, and mean absolute error (MAE) was 1.07°. The overall improvement in performance measures was 9.31–79.50%, which was comparable to that of other models reported in the literature. This study contributes to the domain knowledge by developing an effective artificial intelligence (AI) model for predicting the peak friction angle of FRS. **DOI: 10.1061/(ASCE)CP.1943-5487.0000595.** © *2016 American Society of Civil Engineers.*

Author keywords: Fiber-reinforced soil (FRS); Geosynthetics; Soil-fiber interaction; Peak shear strength; Data mining; Machine learning; Metaheuristic computation.

Introduction

Natural earth materials (i.e., soils and rocks) are often combined with artificial synthetic materials (or geosynthetics), such as geotextiles, geogrids, geomembranes, and geocomposites (Gray and Ohashi 1983), to enhance the mechanical and hydraulic performance of soil in geotechnical and geoenvironmental applications. Soil reinforcement technology using geosynthetics is an attractive and cost-effective means and has been widely applied to numerous geotechnical projects in the past decades. In conventional soil reinforcement techniques, continuous planar reinforcement (e.g., geogrids and geotextiles) is oriented in a preferred direction, usually perpendicular to the loading direction, to help stabilize reinforced structures.

Reinforcement of soil by mixing it with randomly distributed fibers is a promising alternative in projects involving stabilizing thin soil veneers; repairing locally failed slope; improving the bearing capacity of soft ground; strengthening soil in footings, pavement, and earth retaining walls; enhancing soil piping resistance in waterfront structures; increasing dynamic resistance to liquefaction; reducing surficial soil erosion; or mitigating desiccation cracking of compacted clay systems. Compared with planar reinforcement, fiber reinforcement is superior because fibers are not easily damaged during construction. Additionally, a preferential plane of weakness does not develop along the soil-planar reinforcement interface because the fibers are uniformly mixed with soil.

Other advantages of using fibers for soil reinforcement include the following: (1) fibers can be easily mixed with soil (Hejazi et al. 2012; Li et al. 2014); (2) incorporating the fiber affects the soil behavior physically (not chemically), thus causing little negative effect on the environment (Chauhan et al. 2008); (3) the fiber provides an isotropic shear strength increase (peak shear strength is increased and post-peak shear strength loss is reduced) (Hejazi et al. 2012; Li et al. 2014; Maher and Gray 1990); and (4) the fiber increases soil tensile strength, rendering soil desiccation cracking less susceptible to weather influences (Hejazi et al. 2012).

Previous studies have investigated the mechanical behavior of fiber-reinforced soil (FRS) (Kumar and Tabor 2003; Maher and Gray 1990; Michalowski and Čermák 2003; Michalowski and Zhao 1996; Mortazavian and Fatemi 2015; Ranjan et al. 1994;

[1]Professor, Dept. of Civil and Construction Engineering, National Taiwan Univ. of Science and Technology, 43, Section 4, Keelung Rd., Taipei 10607, Taiwan; Eminent Scholar, Del E. Webb School of Construction, School of Sustainable Engineering and the Built Environment, Arizona State Univ., Tempe, AZ 85287 (corresponding author). E-mail: jschou@mail.ntust.edu.tw

[2]Associate Professor, Dept. of Civil and Construction Engineering, National Taiwan Univ. of Science and Technology, 43, Section 4, Keelung Rd., Taipei 10607, Taiwan. E-mail: khy@mail.ntust.edu.tw

[3]Graduate Research Assistant, Dept. of Civil and Construction Engineering, National Taiwan Univ. of Science and Technology, 43, Section 4, Keelung Rd., Taipei 10607, Taiwan. E-mail: M10205105@mail.ntust.edu.tw

Note. This manuscript was submitted on November 2, 2015; approved on February 17, 2016; published online on June 6, 2016. Discussion period open until November 6, 2016; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Computing in Civil Engineering***, © ASCE, ISSN 0887-3801.

Shao et al. 2014; Yetimoglu and Salbas 2003) and proposed theoretical and empirical models for estimating the shear strength properties of FRS (Consoli et al. 1998; Michalowski and Čermák 2003).

Although these models are very convenient, their accuracy is questionable because simplified models do not adequately describe complex mechanisms of soil-fiber interaction. Specifically, simplified models do not consider soil-fiber interaction factors and their nonlinear relationships with the shear strength properties of FRS (Najjar et al. 2013). The preceding discussion has prompted the current study in which an alternative to predict the shear strength properties of FRS was proposed using data mining techniques for constructing an artificial intelligence (AI) model, thereby providing valid and reliable prediction results.

AI-based approaches have attracted a great deal of scientific attention and have been widely used in civil engineering (Alkroosh and Nikraz 2011; Cao et al. 2015; Chou and Lin 2013; Chou et al. 2016; Chou and Pham 2013, 2015; Chou and Tsai 2012; Das 2012; Park and Kim 2011; Samui 2013; Tinoco et al. 2014). Various supervised AI techniques [e.g., artificial neural network (ANN), classification and regression tree (CART), chi-squared automatic interaction detector (CHAID), linear regression (REG), generalized linear (GENLIN) models, and support vector machine (SVM)] are typically used individually to construct single models as the benchmark models (Wu et al. 2008). The AI-based approaches have been confirmed as the potential solutions for solving real-world engineering problems (Cao et al. 2015; Chou and Lin 2013; Chou et al. 2016; Chou and Pham 2013, 2015; Chou et al. 2015; Wei et al. 2008).

To demonstrate the applicability of AI models and their meta-ensembles in geotechnical engineering, this study compiled available high quality data for triaxial and direct shear tests of FRS that were documented in the literature during 1983–2015. The database included the following: soil parameters (soil type, soil classification, mean particle size, dry unit weight, and soil friction angle), fiber parameters (fiber type, fiber dimension, gravimetric and volumetric fiber content, and specific gravity), soil-fiber interface parameter (soil-fiber interface friction angle or soil-fiber interface coefficient) and stress parameters (confining pressure and normal stress).

Data mining, which is an effective data-driven process, was performed in the compiled database by using the recommended AI techniques, such as ANN, CART, CHAID, REG, GENLIN, and SVM models. The prediction performance of the proposed model was also enhanced by incorporating an optimal ensemble method that applied a series of metaheuristic-mining technologies, i.e., voting, bagging, stacking, and tiering. The prediction accuracy of the theoretical and empirical models and AI models was evaluated and compared. By validating the effectiveness and accuracy of the proposed AI models for predicting the peak shear strength properties of FRS, this study facilitates the use of fiber reinforcement in widely varying geotechnical projects.

The rest of this paper is organized as follows. This paper first reviews the functional mechanisms of FRS and methods used to estimate its shear strength. Then the research and evaluation methodologies are presented, and the data collection process and the compiled database are described. After data preparation, this study demonstrates the streamflow of the implemented AI models. The prediction results of the models are compared with those of theoretical and empirical methods reported in the literature. Finally, the concluding remarks and recommendations are provided in the last section.

Literature Review

Functional Mechanism of FRS

Fiber reinforcement is a technique in which randomly distributed natural or synthetic fibers are mixed with soil as uniformly as possible to improve the mechanical behavior of soil. Essentially, random discrete flexible fibers provide reinforcement by mimicking the behavior of plant roots holding the soil.

The many advantages of natural fibers (e.g., reed, coconut, sisal, palm, jute, flax, bamboo, straw, and vine) are their low cost, biodegradable nature, limited environmental effect, lightness, and favorable interface characteristics (e.g., roughness and angles). Natural FRS is believed to play a key ecology–composite role in civil engineering (Hejazi et al. 2012). However, the disadvantages of natural fibers are that their quality and durability depend on the natural environment and that the hydrophilic characteristic of these fibers can reduce the reinforcing effects.

The advantages of commonly used synthetic fibers, such as polypropylene (PP), polyester (PE), polyethylene terephthalate (PET), polyvinyl alcohol (PVA), nylon, glass, and steel, are their high strength, high durability, easily controllable quantity and quality, and chemical, environmental, and climate resistance (Hejazi et al. 2012). Synthetic fibers are resistant to biodecomposition and chemical corrosion. These fibers prevent soil shrinkage and are less susceptible to weather and chemical changes. The most widely used synthetic fiber is PP.

An analysis of the mechanical behavior of FRS by Santoni et al. indicated that reed fibers are superior to glass fibers because natural fibers such as reed fibers have favorable interface characteristics (e.g., roughness and angles) (Santoni et al. 2001). Al-Refeai also indicated that PP fiber was superior to glass fiber and that the optimal length for fiber reinforcement was approximately 76 mm (Al-Refeai 1991). Because natural fibers decompose easily, Ahmad et al. used a waterproof coating containing coir, fly ash, and fibers to reinforce silty sand (Ahmad et al. 2010). They concluded that the mechanical characteristics of the waterproofed fibers were superior to those of the fibers without the waterproof coating.

Sivakumar Babu et al. also reported that coir fiber reinforcement increased the friction angle of reinforced soil (Sivakumar Babu et al. 2008). Anggraini et al. used tensile and unconfined compression tests to study the effects of adding coir fiber and lime to soft soil and showed that the reinforcing effects of fiber and lime depended considerably on the optimal moisture content (Anggraini et al. 2015). Anggraini et al. also reported that a fiber content of 1% increased the friction angle of the soil.

However, Gray and Al-Refeai argued that fiber does not affect the behavior of soil with a fiber content >2% (Gray and Al-Refeai 1986). Kumer and Tabor determined that, when nylon fiber is used, the increase in residual shear strength is greater than the increase in peak shear strength (Kumar and Tabor 2003). Consoli et al. performed triaxial tests of PP fiber–reinforced cement sand. The tests revealed that the PP fiber enhanced peak shear strength and changed the mechanical behavior of the cement sand (Consoli et al. 2004). In Michalowski and Zhao, triaxial tests of soil reinforced with steel fiber showed that steel fiber enhanced the peak shear strength of the soil (Michalowski and Zhao 1996).

Mortazavian and Fatemi further reviewed studies of fatigue behavior in short synthetic fibers and reported that the fatigue behavior of fibers is primarily affected by moisture content and temperature and that fiber fatigue increases with its aspect ratio (Mortazavian and Fatemi 2015). In addition to discrete fibers, the effect of continuous fibers on soil shear strength improvement has also been investigated (Chen et al. 2011). The test results

showed that, like discrete fibers, continuous fibers improve the shear strength of soil in both peak and residual states. The orientation of continuous fiber appears to have significant influence on the reinforcing effect.

In summary, the discrete characteristics of fibers facilitate uniform mixing of soil with fibers, provide an isotopic increase in shear strength to soil-fiber composite, and avoid or delay formation of localized shear plane. Hence, fibers enhance the stability of geotechnical structures.

Methods of Estimating FRS Shear Strength

Peak shear strength values obtained in laboratory tests of FRS or predicted by theoretical or empirical models were used to design geotechnical structures involving FRS. In the industry, however, the need for testing FRS specimens has discouraged the development and implementation of fiber reinforcement because geosynthetics manufacturers have to deal with the properties of composite soil materials rather than focus on the properties of fiber products, which prevents the proper characterization of the fibers' contribution and the optimization of fiber products (Zornberg 2002). As a result, the theoretical or empirical models, considering the soil and fiber separately, gain more popularity in FRS peak shear strength prediction for design.

Michalowski and Čermák (2003) and Zornberg (2002) proposed two different models for predicting the peak shear strength of FRS as a function of the fiber properties, the peak shear strength of soil, and the shear strength of the fiber-soil interface. These models are considered the best available approaches and offer the engineers a simplified and straightforward method to predict the shear strength of FRS. In addition, although FRS could fail attributable to the fiber-slip and fiber-rupture, depending on fiber properties and soil stresses, the fiber-slip is considered as the governing failure mode for FRS in the stress range for most engineering practical applications. Therefore, the present study including the theoretical or empirical models introduced next and the proposed AI models discussed subsequently focuses on the shear strength prediction for FRS in the fiber-slip range.

Michalowski and Čermák proposed a theoretical model based on the concept of energy dissipation to estimate the peak friction angle of discrete randomly distributed FRS under axisymmetric loading conditions (Michalowski and Čermák 2003). Based on the approaches used to develop the model relationships, the Michalowski and Čermák model is referred to as the energy-based model. The model calculates a *macroscopic* internal friction angle, which is equivalent to the peak friction angle of FRS (ϕ_{FRS})

$$\phi_{\text{FRS}} = 2\tan^{-1}\sqrt{\frac{X_f \eta_f M \tan\delta + 6K_p}{6 - X_f \eta_f M \tan\delta}} - \frac{\pi}{2} \tag{1a}$$

where

$$M{:}K_p \sin\theta_o \tag{1b}$$

$$K_p{:}\tan^2\left(45 + \frac{\phi}{2}\right) \tag{1c}$$

$$\theta_o{:}tan^{-1}\sqrt{\frac{K_p}{2}} \tag{1d}$$

where ϕ_{FRS} = peak friction angle of FRS; X_f = volumetric fiber content; η_f = fiber aspect ratio; ϕ = friction angle of soil; and δ = interface friction angle between the sand and fiber.

Zornberg developed a semiempirical model that considered the bilinear shear strength envelope of FRS (Zornberg 2002). In the Zornberg model, the fiber-induced tensile force on failure surface is added to the soil shear strength. The Zornberg model is renowned as a discrete model. For the cases in which fiber-slip mode dominates, the ϕ_{FRS} is calculated as

$$\phi_{\text{FRS}} = \tan^{-1}[(1 + \alpha \cdot \eta_f \cdot X_f \cdot c_{i,\varphi}) \cdot \tan\phi] \tag{2}$$

where η_f = fiber aspect ratio; X_f = volumetric fiber content; $c_{i,\phi}$ = coefficient of interaction for the friction angle component ($\tan\delta/\tan\phi$, where δ = sand-fiber interface friction angle); and α = empirical coefficient that accounts for the orientation and efficiency of mixing of the fibers. If the fibers are uniformly distributed and 100% efficient, then α is equal to 1; otherwise, $\alpha < 1.0$; ϕ = friction angle of soil.

Najjar et al. performed a statistical analysis to quantify uncertainies in models for predicting the shear strength of fiber-reinforced sand (Najjar et al. 2013). The previously discussed models for predicting sand-fiber shear strength were evaluated. Statistical results showed that the energy-based model underestimated the measured friction angle by an average of approximately 10% whereas the discrete model overestimated the friction coefficient by 6%, with associated coefficients of variation on bias values of 0.20 and 0.17, respectively. Najjar et al. reported that the discrepancy occurs because some factors (e.g., fiber-grain scale effect and sand grain size) known to affect the behavior of FRS are not included in the previously discussed models. Additionally, these models do not consider the nonlinear relationship between the shear strength properties of FRS and the fiber contents. The accuracy of these two models is also evaluated using the database compiled in this study. Their accuracy is then compared with that of the proposed AI models.

Artificial Intelligence in Geotechnical Engineering

AI technology has recently been used to solve geotechnical engineering problems in various contexts. For instance, Goh applied values for standard penetration tests and the parameters of soil and earthquakes in ANNs for investigating the conditions of soil liquefaction potential and the relationship between soil parameters and earthquakes and showed fine sand content and standard penetration test values to be the key soil parameters (Goh 1994). Samui used a multivariate regression model to analyze slope stability (Samui 2013). Park and Kim adopted ANNs to predict the strength of lightweight reinforced soil and showed that the major influencing factors were cement content, moisture content, and air content (Park and Kim 2011).

Alkroosh and Nikraz simulated the subsidence of pile foundations under loading in cone penetration tests (Alkroosh and Nikraz 2011). Experimental applications of ANN for predicting the axial settlement of pile foundations embedded in sand showed that the models accurately predicted the nonlinear behaviors of soil under loading. Tsompanakis et al. used ANN to simulate the nonlinear response of a levee under seismic loadings to reduce the engineering costs of predicting earthquakes on a massive scale (Tsompanakis et al. 2009). Benardos and Kaliampakos used ANN to obtain the advance speed and related parameters of a tunnel-boring method (Benardos and Kaliampakos 2004). The obtained data were then used to facililate completion of the engineering project.

In Goh et al., a Bayesian neural network used to analyze the undrained side resistance of drilled shafts showed that the model not only made accurate predictions, it also provided information on the characteristic prediction errors related to uncertainty among the parameters of the data (Goh et al. 2005). Alkroosh and Nikraz

used a gene-expression-programming (GEP) model to investigate the correlation between pile axial capacity and data obtained in cone penetration tests. Simulations obtained a maximal coefficient R value of 0.96 between the pile axial capacity predicted by the GEP model and the actual pile capacity (Alkroosh and Nikraz 2011).

With advancement of technology and increasing requirements for precision in geotechnical engineering, experts and scholars have attempted to optimize AI models and have proposed related optimizing technologies. For example, Armaghani et al. used a hybrid model that combined particle swarm optimization and ANN (PSO-ANN) to determine a related index for rock (Jahed Armaghani et al. 2014). The model obtained a maximal correlation value of 0.944 between the predicted and actual values of rock shear parameters. In Yu et al., an ANN model was combined with an evolutionary algorithm (Yu et al. 2007). An intelligent displacement back-analysis method was then used to explore earth-rockfill dams, demonstrating that the new method could replace the finite element analysis method in solving large-scale nonlinear engineering problems.

Chou et al. combined a smart firefly algorithm, based on swarm intelligence, and least squares support vector regression (SVR) to simulate the reinforcement tensile loads of geosynthetic-reinforced soil retaining walls. This metaheuristic optimized regression model accurately predicted the mobilize reinforcement tensile loads under working stress and under severe deformation. The analysis results revealed that the prediction accuracy of this model was superior to that of the single and composite models: the maximal coefficient R between the predicted and actual tensile loads achieved 0.99, whereas the mean absolute percentage error (MAPE) was lower than 10% (Chou et al. 2015).

In summary, AI has great potential use as a prediction tool in geotechnical engineering. The following section summarizes current prediction methods and describes the innovative technology used in this study to create metaensembles that improve accuracy in predicting FRS shear strength.

Research Method

The AI technologies used in this study included classification and regression methods, machine learning, and metaensemble models. Next, prediction techniques, validation and evaluation methods are discussed in detail.

Classification and Regression Methods

REG

The minimum square function of REG model (IBM 2009) establishes relationships between one or more explanatory variables and the dependent variables. That is, it explains the correlation among the varaibles and the prediction problems (Sykes 1993). In practice, changes in the dependent variable Y are often affected by the characteristic variable X. A general equation is as follows:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + e_i \tag{3}$$

where Y_i = dependent variable; $X_{1i}, X_{2i}, \ldots, X_{pi}$ = explanatory variables; β_0 = constant variable; and $\beta_1, \beta_2, \ldots, \beta_p$ = regression coefficients; and e_i is the error term.

CART

The CART is a simple and highly efficient prediction model based on empirical learning. If the target variable is a categorical variable, then the model is called a classification tree; if the target variable is a continuous variable, then the model is called a regression tree. The model is presented in a tree structure. Each internal (nonleaf) node represents a test of an attribute, each branch represents the test result, and each leaf (or terminal) node has a class label and class result (IBM 2009; Timofeev 2004).

The CART model performs classification through repeated operations. The tree is "pruned" to minimize total error, which optimizes the prediction accuracy of the tree by minimizing the number of branches. The CART tree is branched and split according to the Gini index (Timofeev 2004). The formulas are as follows:

$$g(t) = \sum_{j \neq i} p(j|t)p(i|t) \tag{4}$$

$$p(j|t) = \frac{p(j,t)}{p(t)} \tag{5}$$

$$p(j,t) = \frac{p(j)N_j(t)}{N_j} \tag{6}$$

$$p(t) = \sum_j p(j,t) \tag{7}$$

where i and j = categorical variables in each item; $N_j(t)$ = recorded number of Node t in Category j; and N_j = recorded number of the root node in Category j.

CHAID

The CHAID is an efficient statistical model or a decision tree proposed by Kass et al. in 1980 (IBM 2009; Kass 1980). To establish a decision tree, the CHAID primarily uses the chi-squared test for determining the optimal splits. Merging and splitting are performed continuously until the grouped results show no substantial differences or until the number of sample units included in the subgroups is too small to enable a meaningful estimation of probability. The CHAID model also uses various methods to measure different data types. For example, continuous data are examined through F tests, and categorical data are measured through the CHAID.

GENLIN

The GENLIN regression model proposed by Nelder and Wedderburn in 1972 (IBM 2009; Nelder and Wedderburn 1972) has a wider scope and is more realistic compared with the REG model. In addition to statistics with normal distributions, GENLIN model can be used for statistical analyses of various distributions, e.g., in binary-logistic and log-linear models. The function used to establish the relationship between X (input) and Y (output) in GENLIN model is defined as follows:

$$g[E(y)] = (X \cdot \beta) + O \tag{8}$$

where g = link function; $E(y)$ = expected value of y; β = regression coefficient; and O = offset variable.

Machine Learning

ANN

The ANNs are a family of information-processing models inspired by biological neural networks; the structure of an ANN is analogous to that of the human brain, in which the neurons are interconnected through synapses (Das 2012). This virtual system receives multiple inputs and uses them to make predictions (IBM 2009). The processing element, which is called a "neuron," has the following characteristics: (1) a filtering function to ensure that incomplete

data inputted to a specific node do not substantially affect the network; and (2) adaptive learning ability to adjust the connective weight between the nodes.

ANNs have multiple-input, multiple-output systems and a basic structure that includes an input layer, a hidden layer, and an output layer. When a processing element sends an output to another processing unit, the output is received as an input by the second element. This mapping relationship in the network model can be expressed with the following equation:

$$\alpha_i = \sigma\left(\sum_j \omega_{ij} o_j\right), \qquad \sigma(x) = \frac{1}{1 - e^{-x}} \tag{9}$$

where α_i refers to ANN activities; ω_{ij} = weight connecting two neurons; o_j = output or an output signal of the ANN; x = activation of ith neuron; and $\sigma(x)$ = activation function of the ANN that facilitates transformation of inputs into outputs by multiplying the inputs from the processing elements by the corresponding weights.

SVM and SVR

The SVMs are supervised learning models that were first proposed by Vapnik (1995). The SVMs use a straight line or hyperplane for classification. A SVM model is used when the target variable involves categorical data; conversely, an SVR model is used when the target variable involves continuous data (Tinoco et al. 2014). The classifier can be described as

$$f(x) = \omega \cdot \phi(x) + b \tag{10}$$

$$\omega = \sum_i^r y_i \alpha_i \phi(x_i) \tag{11}$$

where ω = weight vector representing the flatness of $f(x)$ in the high-dimensional space; b = parameter of the model; and ϕ = high-dimensional feature space representing the nonlinear mapping function.

An input data point x can be represented as $\phi(x)$ in the high dimensional space. The computational expense of $\phi(x) \cdot \phi(x_i)$ is reduced by using a kernel function (Mathur and Foody 2008). Thus, the classification decision function becomes

$$f(x) = \text{sign}\left[\sum_i^r \alpha_i y_i K(x, x_i) + b\right] \tag{12}$$

where for each of r training cases there is a vector (x_i) that represents the spectral response of the case together with a definition of class membership (y_i); α_i $(i = 1, \ldots, r)$ are Lagrange multipliers; and $K(x, x_i)$ is the kernel function.

SVR is a variation of SVM. The SVR first uses a fixed mapping procedure to map the SVR input to an n-dimensional feature space. Nonlinear kernel functions are then fit to the high-dimensional feature space, in which input data are easier to separate compared with input data in the original input space. The linear model in the feature space, $f(x, \omega)$, can be expressed by Eq. (13)

$$f(x, \omega) = \sum_{j=1}^m w_j g_j(x) + b \tag{13}$$

where $g_j(x)$ = set of nonlinear transformations; and b = bias term. Moreover, estimation quality is measured by a loss function $L_\varepsilon = [y, f(x, \omega)]$ where

$$L_\varepsilon = [y, f(x, \omega)] = \begin{cases} 0 & \text{if } |y - f(x, \omega)| \le \varepsilon \\ |y - f(x, \omega)| - \varepsilon & \text{otherwise} \end{cases} \tag{14}$$

The SVR uses the ε-insensitive loss function to identify the minimal dimensional space $\|\omega\|^2$ and to reduce the model complexity. This function is introduced by including nonnegative slack variables, ξ_i and ξ_i^*, where $i = 1, \ldots, n$ is used to identify training samples from the ε-insensitive zone. The SVR can thus be formulated as a minimized version of the following function:

$$\min \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^n (\xi_i + \xi_i^*) \tag{15}$$

$$\text{subject to} \begin{cases} y_i - f(x_i, \omega) \le \varepsilon + \xi_i^* \\ f(x_i, \omega) - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0, i = 1, \ldots, n \end{cases}$$

This optimization problem can be transformed into a dual problem, which is solved by

$$f(x) = \sum_{i=1}^{n_{SV}} (\alpha_i - \alpha_i^*) K(x_i, x) \quad \text{subject to } 0 \le \alpha_i^* \le C, 0 \le \alpha_i \le C \tag{16}$$

where n_{SV} = number of support vectors. The kernel function is

$$K(x, x_i) = \sum_{i=1}^m g_i(x) g_i(x_i) \tag{17}$$

During training, kernel functions (i.e., linear, radial basic, polynomial, or sigmoid function) are used to identify support vectors along the function surface. This study used the radial basis function (RBF) [i.e., $K(x, x_k) = \exp(-\|x - x_k\|^2/2\sigma^2)$] for the SVM or SVR as the kernel function.

Metaensemble Models

To obtain improved prediction performance compared with the previously discussed baseline models, this study proposes the following metaensemble models.

Voting

Fig. 1(a) shows that the voting model sets the mean input obtained from several models as the predictive value. Optimal heterogeneous models are then combined into ensemble models for making predictions. The method generally enhances prediction accuracy (Prodromidis and Stolfo 1999).

Bagging

Fig. 1(b) depicts the bagging process, in which samples randomly drawn from the bootstrapped replica in the Group K data set are combined to form homogenous prediction models to obtain a mean value of voting output (Austin et al. 2013).

Stacking

Stacking enables multistage predictions. Stage 1 is established first, comprising the baseline models. The prediction results obtained from these models ($Y_{i=1\sim n}$) then serve as input X of Stage 2 for making further predictions (Y_{pred}), as shown in Fig. 1(c) (Wolpert 1992).

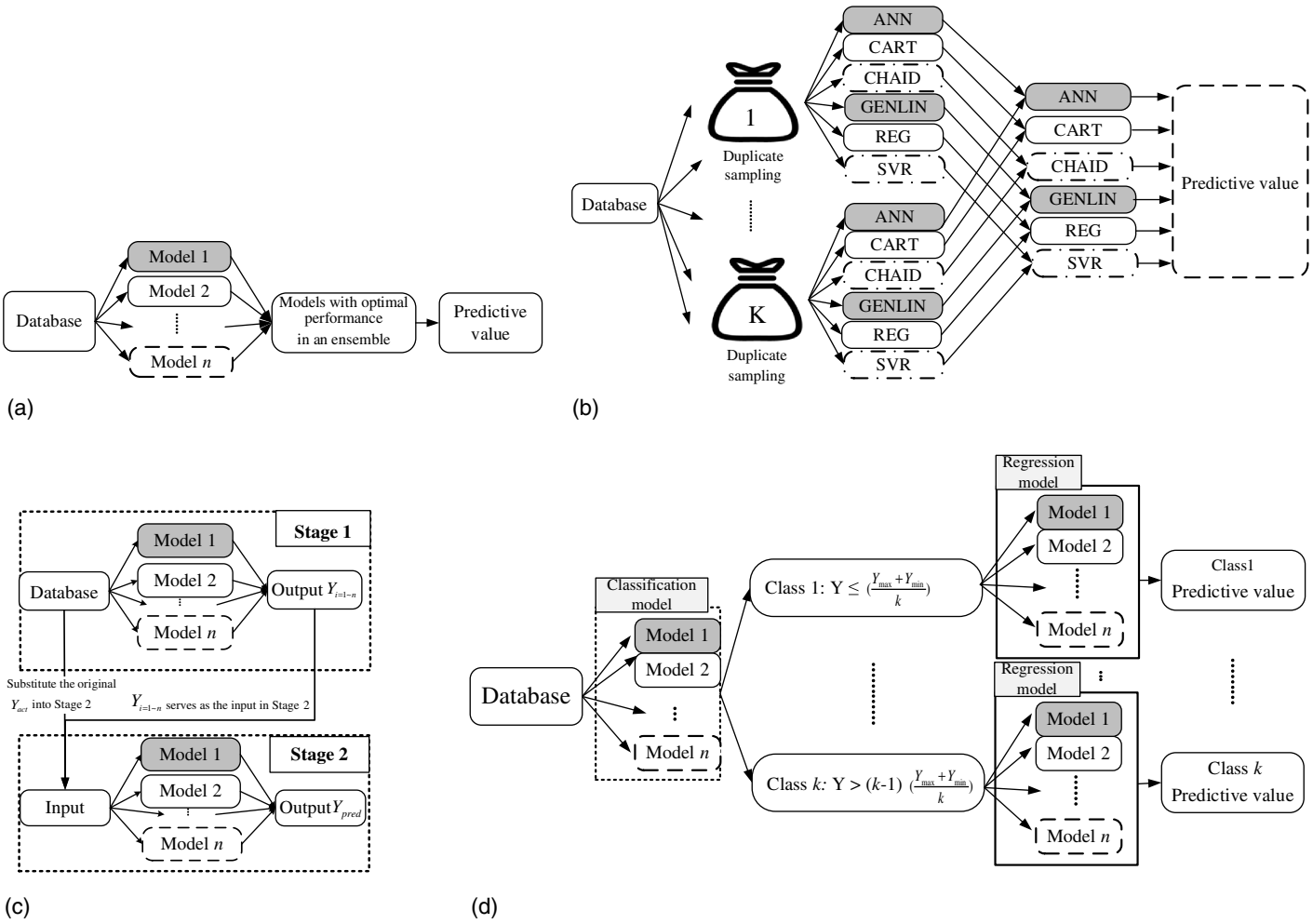

Fig. 1. Metaensemble methods: (a) voting; (b) bagging; (c) stacking; (d) tiering

Tiering

Tiering involves analyzing data in two tiers. The first tier uses classification models to make predictions; the second tier evenly divides the data into k classes according to the standard value (T value). The data can be divided into 2, 3, ... or k classes, Fig. 1(d) shows that the regression models are then used to make predictions. The data are divided according to the T value (Chou and Tsai 2012): the general equation is $T = Y_{\max} + Y_{\min}/k$. For example, if the total data are divided into three classes, then $T = Y_{\max} + Y_{\min}/3$, where $Y_{\text{act}} \leq T$ belongs to Class 1, $T < Y_{\text{act}} \leq 2T$ belongs to Class 2, and $2T < Y_{\text{act}}$ belongs to Class 3.

Stratified Cross-Validation

The stratified cross-validation procedure is applied when comparing prediction accuracy among two or more models. This method divides randomly drawn data samples into training samples and testing samples by splitting the samples into k mutually exclusive subsets. Each time, k-1 subsets serve as the training samples, and the remaining subsets serve as test samples. This model validation process is repeated k times to reduce errors during random sampling. Kohavi demonstrated that a k of 10 indicates analytical validity, computational efficiency, and optimal deviation (Kohavi 1995). Therefore, the performance of each proposed model was evaluated in terms of average prediction error in the 10 groups.

Evaluation of Prediction Accuracy

This study used four commonly used statistical methods to compare the error rate between actual and predicted values. The four methods are the correlation coefficient (R), MAPE, root mean square error (RMSE), and mean absolute error (MAE). The R is the correlation between two items: an R that is close to 1 indicates that the model has a high goodness of fit. Similarly, low MAPE, RMSE, and MAE values indicate a low rate of error in the model predictions. To evaluate the overall prediction performance of the models, averaged normalization was used to obtain a synthesis index (SI) for all evaluation principles. The evaluation methods are explained as follows.

Correlation Coefficient

The statistical index that shows the linear correlation between two variables is called R. The R values, which are between -1 and $+1$, are calculated as follows:

$$R = \frac{n\sum_{i=1}^{n} y_i \times p_i - \sum_{i=1}^{n} y_i \times \sum_{i=1}^{n} p_i}{\sqrt{n \times \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2} \times \sqrt{n \times \sum_{i=1}^{n} p_i^2 - (\sum_{i=1}^{n} p_i)^2}} \tag{18}$$

where p_i = predicted value; y_i = actual value; and n = number of samples.

The R is interpreted as follows: If $R > 0$, then the two variables are positively related; if $R < 0$, then the two variables are negatively related.

MAPE

The MAPE is the index typically used to evaluate the accuracy of prediction models. The closer MAPE is to 0, the better the prediction results achieved by the model. When the MAPE is lower than 10%, the model is highly accurate in its efficacy

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{p_i - y_i}{y_i}\right| \tag{19}$$

RMSE

The RMSE represents the dispersion of errors by a prediction model, i.e., the prediction accuracy of the model

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(p_i - y_i)^2}{n}} \tag{20}$$

MAE

The MAE is the mean absolute difference between the prediction and the actual value; therefore, the unit of MAE is the same as that of the measurement

$$\text{MAE} = \frac{\sum_{i=1}^{n}|p_i - y_i|}{n} \tag{21}$$

SI

The performance measures (discussed in the previous sections) were synthesized by using the following equation to normalize the average of the error measurements and calculate the SI. The SI ranges between 0 and 1; an SI close to zero indicates an accurate model

$$\text{SI} = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{P_i - P_{\min,i}}{P_{\max,i} - P_{\min,i}}\right) \tag{22}$$

where P_i = ith averaged performance measurement; $P_{\min,i}$ = minimum value of the ith performance measure; $P_{\max,i}$ = maximum value of the ith performance measure; and m = number of evaluation methods.

Database

This study compiled high-quality data for triaxial and direct shear tests of FRS reported in the literature during 1983-2015. Statistical analyses of the compiled data showed that 63.3% of the studies examined the behavior and failure mechanisms of FRS (Consoli et al. 2004, 2005, 2007, 2009b; Freitag 1986; Gray and Al-Refeai 1986; Gray and Ohashi 1983; Hejazi et al. 2012; Ibraim and Fourmont 2007; Li and Zornberg 2013; Loehr et al. 2005; Maher and Gray 1990; Michalowski and Zhao 1996; Nataraj and McManis 1997; Ranjan et al. 1994; Shao et al. 2014; Sivakumar Babu et al. 2008; Yetimoglu and Salbas 2003; Zornberg 2002), 20% developed models for predicting shear strength properties of FRS (Consoli et al. 2009b; Maher and Gray 1990; Najjar et al. 2013; Ranjan et al. 1996; Sadek et al. 2010; Santoni et al. 2001), and the remainder evaluated whether adding cement improved the shear stregnth of FRS (13.3%) (Ahmad et al. 2010; Chauhan et al. 2008; Consoli et al. 1998, 2009b; Kaniraj and Havanagi 2001) and documented practical applications of FRS on geotechnical engineering projects (3.3%) (Santoni et al. 2001).

This study analyzed the following parameters, which are known to affect FRS shear strength. Ibriam and Fourmont performed direct shear tests to study the mechanical response of fiber-reinforced fine sand (Ibraim and Fourmont 2007). Factors including confining pressure, types of fibers and their physical properties (e.g., density, length, aspect ratio, orientation of alignment), and soil properties are considered in their study. Consoli et al. (2009b) used triaxial tests to investigate how addition of cement affects fiber-reinforced sand. The control factors were the type of the soil, the specific gravity of the soil, the coefficient of uniformity for the soil, the length of the fiber, the width of the fiber, and the specific gravity of the fiber.

Yetimoglu and Salbas also performed direct shear tests to investigate the reinforcing effect of discrete fibers (Yetimoglu and Salbas 2003). The test factors included the strength, size, and density of the fiber and the physical properties of the soil. Sivakumar Babu et al. considered factors including fiber length, diameter, fiber content, soil type, dry unit weight, specific gravity, and friction angle of soil in their study. Table 1 shows the parameters considered in this study after integrating the influential factors according to the literature (Ahmad et al. 2010; Al-Refeai 1998, 1991; Chen 2007; Consoli et al. 1998, 2004, 2007, 2009a; Gray and Al-Refeai 1986; Gregory 2006; Ibraim and Fourmont 2007; Li and Zornberg 2013; Maher and Gray 1990; Michalowski and Čermák 2003; Michalowski and Zhao 1996; Nataraj and McManis 1997; Sadek et al. 2010; Shao et al. 2014; Sivakumar Babu et al. 2008; Yetimoglu and Salbas 2003): soil parameters (soil type, soil classification, mean particle size, dry unit weight, soil friction angle), fiber parameters (fiber type, fiber dimension, gravimetric and volumetric fiber content, and specific gravity), soil-fiber interface parameter (soil-fiber interface friction angle or soil-fiber interface coefficient) and stress parameters (confining pressure or normal stress).

Data reported in previous FRS-related studies were compiled using two different methods, depending on the type of reported documents. Method 1 was directly recognizing the parameter values in the relevant tables and in-text descriptions; method 2 was calculating the friction angle values of FRS from the reported test results (stress-strain curves in triaxial tests or stress-displacement curves in direct shear tests) using Mohr-Coulomb failure criteria. For FRS specimens that showed ductile (strain-hardening) behavior (i.e., the stress-strain curve did not have a definite failure), the peak shear strength was defined at an axial strain of 15% for calculating the friction angle of FRS.

Of the 316 total data entries collected from 20 publications, 16 entries were excluded from the analysis because they did not have a complete set of reliable input parameters. Therefore, the total number of data entries used in this study was 300. Table 2 presents the detailed information of compiled database of tests conducted on fiber-reinforced sand.

Establishing the AI Models

The multifunctional data-mining analyzer of the IBM SPSS Modeler was used to ensure accurate prediction models. Although the software provides an easily used platform for developing basic AI models, advanced techniques such as the metaensemble models proposed in this study are quite customized based on the novelty of the experienced users. A k-fold cross-validation procedure was used to classify the data into training and test subsets. The following section explains the processes for constructing the baseline and metaensemble models. Table 3 shows the default parameters for these models.

Baseline Model Streams

When the data in the training subset were trained, the prediction model (i.e., intelligent golden brick) obtained from the training

Table 1. Summary of Parameters Considered in This Study

Type of variable	Category	Parameter	Minimum	Maximum
Input variable	Fiber	Fiber type	Polypropylene/polyester/ polyamide/steel/glass/coir/palm	
		Fiber length L_f (mm)	6	51
		Fiber diameter D_f (mm)	0.01	1.25
		Volumetric content X_f (%)	0.17	5.53
		Gravimetric content W_f (%)	0.10	6.55
		Fiber specific gravity $G_{s,f}$	0.58	7.85
	Soil	Classification	SP/SW	
		Soil type	Fine sand/medium sand/medium dune/medium mortar/coarse sand	
		Mean particle size D_{50} (mm)	0.09	1.45
		Friction angle ϕ (°)	26.4	43.0
		Cohesion c (kPa)	0	6.9
		Dry unit weight γ_d (kN/m^3)	13.00	18.39
	Soil-fiber interface	Soil-fiber interface friction angle δ (°)	16	40
		Coefficient of interaction $c_{i,\phi}$	0.37	1.33
	Stress condition	Confining or normal pressure σ_3 or σ_n (kPa)	20	600
Output variable	Shear strength property	Friction angle of FRS ϕ_{FRS} (°)	31.7	67.4

was placed in the data of the test subset. Fig. 2 shows that the test results were then exported and evaluated by using various methods to measure the prediction error for the model.

Modeling Stream of Voting Method

The intelligent golden brick prediction model obtained after training was combined with the bagging method ($K = 10$). The modified results were then added to the test subset, and the test results were exported. The error-measuring methods of the model's predictions were also employed for the assessment (Fig. 3).

Modeling Stream of Bagging Method

First, data in the training subset were randomly drawn from a bootstrapped replica and used to train the prediction models (i.e., ANN, CART, CHAID, GENLIN, REG, and SVR). The homogenous intelligent golden bricks were then combined to export the test results. Fig. 4 shows the error-measuring methods that were then used to assess the prediction accuracy of the model.

Modeling Stream of Stacking Method

This method was analyzed in two stages. In Stage 1, data in the training subset were substituted into the regression model (Step 2) to build a prediction model or intelligent golden brick. The golden brick was then used in the training subset for testing (Step 3); the prediction results obtained in Stage 1 were then used as inputs for Stage 2. In Stage 2 (Step 4), the regression model was again used to build prediction models. When the intelligent golden brick built in Stage 1 was placed in the test subset (Step 5), Stage 2 was stacked on the golden brick built in Step 4 for testing (Step 6). Finally, the metaheuristic prediction results were exported. Fig. 5 shows the modeling stream used to measure the prediction accuracy of the stacking method.

Modeling Stream of Tiering Method

First, the data in the database were recoded into specified k classes (Class 1, 2, ..., k). The data were then randomly divided into training and test subsets. Classification-regression models were then established for each of the k classes. In Tier 1, the classification model (SVM) was used to build and train the prediction model (i.e., intelligent golden brick of the classification type). In Tier 2, the regression model was then used to establish the prediction model (i.e., intelligent golden brick of the regression type). Finally, the golden bricks obtained by training were placed in the test subset to perform tests and export the results. Fig. 6 shows the tiering stream used to measure the prediction accuracy of the models.

The tiering method divided the data into k classes according to their attributes. The results were not exported until a weighted average of the prediction performance, based on the actual distribution of the data, had been calculated. The following sample was demonstrated using R values in two-class tiering: the prediction performance value R of the first class was 0.93, and that of the second class was 0.69. The error evaluation weighting was calculated using the following steps: [(250 entries of data in the first class/300 entries of data in total) × R (0.93)] + [(50 entries of data in the second class/300 entries of data in total) × R (0.69)]. The overall R value of two-class tiering (0.89) was then obtained. Table 4 shows the results obtained by other error evaluation methods, which were performed similarly.

Evaluation of Prediction Models

This section discusses and evaluates the prediction accuracy of the models. Data were categorized into three groups so that the prediction models proposed in this study (i.e., baseline and metaensemble models) could be compared with theoretical and empirical models proposed in previous studies. Group 1 includes the whole data set of FRS including with all parameters presented in Table 1. Thus, Group 1 consists of 300 samples with 15 inputs.

Consisting of inputs used in the energy-based model [Eq. (1*a*)] proposed by Michalowski and Cermak (2003), Group 2 considers four inputs including: fiber aspect ratio, η_f ($= L_f/D_f$); soil friction angle, ϕ; fiber-soil interface friction angle, δ; and volumetric fiber content. X_f. Similarly, Group 3 only includes factors used in the discrete model [Eq. (2)] proposed by Zornberg (2002): soil friction angle, ϕ; fiber aspect ratio, η_f; volumetric fiber content, X_f; and soil-fiber interface friction coefficient, $c_{i,\phi}$. In the discrete model, the empirical coefficient (α) values were assumed to be 1.0 and 0.5 to represent the maximum and medium values, respectively, for fiber uniformity and efficiency.

The theoretical and empirical models [in Eqs. (1*a*) and (2)] were developed exclusively for FRS, in which the failure mode is

Table 2. Database Compiled from Tests of Fiber-Reinforced Sand

		Fiber parameters						Soil and interface parameters						Stress parameter	Shear strength property
References	Data points	Fiber type	L_f (mm)	D_f (mm)	X_f (%)	W_f (%)	$G_{s,f}$	Classification	Soil type	D_{50} (mm)	γ_d (kN/m^3)	ϕ (degrees)	$c_{i,\phi}$	σ_3 or σ_n (kPa)	ϕ_{FRS} (degrees)
Shao et al. (2014)	15	Nylon	12	0.1	0.2–1.75	0.1–0.9	0.91	SP	Fine sand	0.31	17.5	29.8	0.55	50–250	29.8–37.1
Li and Zornberg (2013)	24	PP	25–50	0.2–0.7	0.18–0.70	0.1–0.4	0.91	SP	Medium sand	0.7	15.54–15.91	31.6–35.2	0.53–0.56	60–210	31.6–48.4
Ahmad et al. (2010)	18	Palm	15–45	0.4	0.25–0.50	0.25–0.50	1.46	SP	Medium sand	0.68	14.50	30.5	1.08–1.33	50–250	30.5–37.8
Sadek et al. (2010)	18	Nylon	7–-27	0.18	0.5–1.5	0.4–1.2	1.3	SP	Fine/medium sand	0.39–1.45	16.11–16.18	30.5–37.9	0.47–0.55	100–200	30.5–43.4
Consoli et al. (2009b)	14	PP	12–50	0.02–0.1	0.82	0.50	0.91	SP	Fine sand	0.09	14.69	37.0	0.80	100–400	37.0–47.8
Sivakumar Babu et al. (2008)	8	Coir	15	0.25	1.34–2.01	0.5–1.5	1.08	SP	Fine sand	0.43	14.38	43.0	0.9	100–150	43.4–56.3
Consoli et al. (2007)	12	PP	6–24	0.02	0.85	0.51	0.91	SP	Fine sand	0.16	15.08	26.4	0.81	100–400	26.4–37.0
Chen (2007)	12	PP	51	0.75	0.67–0.73	0.40	0.91	SP	Fine sand	0.18	14.95–16.25	29.7–33.1	0.82	140–415	29.7–41.5
Ibraim and Fourmont (2007)	33	PP	35	0.1	0.44–1.44	0.30–1.00	0.91	SP	Fine sand	0.32	13.00–14.45	30.9–35.9	0.79–0.81	55–508	30.9–40.2
Gregory (2006)	15	PP	50	0.24	0.3–3.52	0.17–2.07	0.91	SP	Fine sand	0.25	15.53	33.2	0.50	69–276	33.2–54.8
Consoli et al. (2004)	14	PE/PP/glass	6–36	0.01–0.18	0.30–0.87	0.51	0.91–2.62	SP	Fine sand	0.16	15.3	36.3–37.5	0.50–0.82	60–100	36.3–49.8
Michalowski and Čermák (2003)	24	Nylon/steel	25.4	0.3	0.5–2	0.38–2.40	1.28–7.85	SP	Fine/coarse sand	0.22–0.89	15.66–16.45	36.5–38.1	0.37–0.64	50–400	38.1–45.6
Yetimoglu and Salbas (2003)	15	Nylon	20	0.05	0.18–0.89	0.1–0.5	0.93	SP	Fine sand	0.30	16.3	42.0	0.55	103–321	40.4–42.3
Al-Refeai (1998)	8	PP	25–51	0.39–0.44	0.38	0.20	0.90	SP	Fine sand	0.30	16.57	38.0	0.80	50–200	38.0–48.8
Consoli et al. (1998)	6	Glass	12.8	0.5	0.2	0.3	2.62	SW	Medium sand	0.5	17.5	35.2	0.52	60–100	35.0–46.0
Nataraj and McManis (1997)	4	PP	25	0.4	0.17–0.51	0.1–0.3	0.91	SP	Fine sand	0.17	15.2	33.5	0.52	342	33.5–38.8
Michalowski and Zhao (1996)	14	Nylon/steel	25.4	0.3–0.64	0.5–1.25	0.40–6.55	1.28–7.85	SP	Coarse sand	0.89	15.66	36.5	0.39–0.69	50–300	36.5–37.8
Al-Refeai (1991)	6	PP/glass	25	0.3–0.4	0.3–0.91	0.5	0.9–2.68	SP//SW	Fine/medium sand	0.18–0.78	15.83–16.07	36.5–41.6	0.43–0.52	200	36.5–58.6
Maher and Gray (1990)	30	Glass	18–37	0.3	0.45–3.43	0.74–5.97	2.7	SP	Medium sand	0.41–0.60	16.05–18.39	30.3–36.5	0.52–0.66	50–300	30.3–67.4
Gray and Al-Refeai (1986)	26	Reed/glass	13–38	0.3–1.25	0.3–2.99	0.21–6.00	0.58–2.7	SP	Medium dune	0.41	16.05	34.3–34.4	0.56–0.85	200	33.3–59.4

Note: PE = polyester; PP = polypropylene; SP = poorly-graded sand; SW = well-graded sand.

Table 3. Default Parameters of Baseline Models

Model	Parameters	Values/options
ANN	Alpha	0.9
	Initial eta	0.3
	High eta	0.1
	Low eta	0.01
	Eta decay	30
	Hidden layers	Three (20,15,10)
	Persistence	200
CART	Levels below root	5
	Mode	Simple
	Maximum surrogates	5
	Minimum change in impurity	0.0001
	Impurity measure for categorical targets	Gini
	Minimum records in parent branch (%)	2
	Minimum records in child branch (%)	1
CHAID	Mode	Simple
	Alpha for splitting	0.05
	Alpha for merging	0.05
	Chi-square method	Pearson
	Minimum records in parent branch (%)	2
	Minimum records in child branch (%)	1
	Epsilon for convergence	0.001
	Maximum iterations for convergence	100
	Allow splitting of merged categories	False
	Use Bonferroni adjustment	True
GENLIN	Distribution	Normal
	Singularity tolerance	1×10^{-7}
	Value order for categorical inputs	Ascending
	Scale parameter method	Maximum likelihood estimate
	Covariance matrix	Model-based estimator
	Confidence interval level (%)	95
SVM/SVR	Stopping criteria	1.0×10^{-3}
	Regularization parameter (C)	10
	Regression precision (epsilon)	0.1
	Kernel type	RBF
	RBF gamma	0.1
REG	Mode	Simple
	Only use complete records	True
	Singularly tolerance	1.0×10^{-4}
	Stepping criteria	Use probability of F
	Probability option(s) entry	0.05
	Probability option(s) removal	0.1
	F value option(s) entry	3.84
	F value option(s) removal	2.71

governed by slippage of the fibers in the soil matrix (typical failure mode in the stress range for most engineering practical applications). Therefore, cases in which fiber rupture could have been the dominant failure mode [e.g., very large confining pressure ($\sigma_3 > 400$ kPa) and fiber aspect ratio (η_f) > 200] were excluded before assessing the model performance. For that reason, 85 samples with that criteria were removed from Groups 2 and 3, resulting in a total number of 215 samples including data from 148 reinforced sand tests and 67 unreinforced sand tests.

To summarize, Group 1 consists of 300 samples of FRS with all 15 inputs. Group 2 involves 215 samples with four inputs (i.e., η_f, ϕ, δ, and X_f), and Group 3 includes 215 samples with four inputs (i.e., ϕ, η_f, X_f, and $c_{i,\phi}$). Model performance was assessed by a stratified 10-fold cross-validation approach as discussed in the "Stratified Cross-Validation" subsection. The average prediction results for 10 test folds can then be used to appraise model performance. To do so, randomly selected data were divided into 10 distinct folds. Each fold was used in turn as a test set with the remaining folds used as a training set so as to ensure that all data instances were applied in both training and testing phases.

Model Results: Group 1

Baseline Prediction Models

The baseline models used to predict the friction angles of FRS were the ANN, CART, CHAID, GENLIN, REG, and SVR models. For Group 1, the SVR model had the highest prediction accuracy. Specifically, the SVR model had a maximal R value of 0.90 ($R = 1$ denoted a perfect positive correlation), a MAPE of 5.61%. The SI value for the SVR model was also superior to that of other models (Table 5).

Metaensemble Models

When the voting method was used, the CART+GENLIN+SVR ensemble model had the best performance (R value of 0.92 and MAPE of 5.01%). In the bagging method, the SVR model had the best stability and prediction accuracy (R value of 0.94 and a MAPE of 4.09%). The CART model used in the stacking method also obtained an R value of approximately 0.90 and a relatively unsatisfactory MAPE of 5.56%. The best prediction performance by the tiering method was obtained when the samples were divided into two classes. The best metaensemble model was the tiering SVM-(SVR/SVR) model, which had an R value of 0.89 and a low MAPE of only 3.27%.

Model Results: Group 2

Baseline Prediction Models

In Group 2, the ANN model showed the optimal performance among all of the baseline models, with an R value of 0.89 and a MAPE of 6.58%. The prediction results of other models are shown in Table 5.

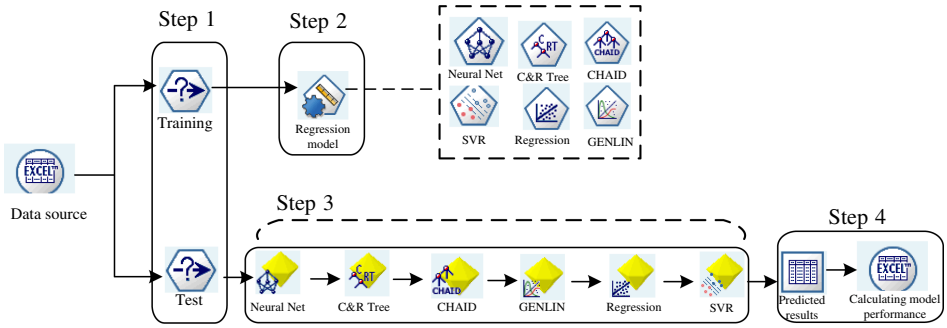

Fig. 2. Baseline model stream

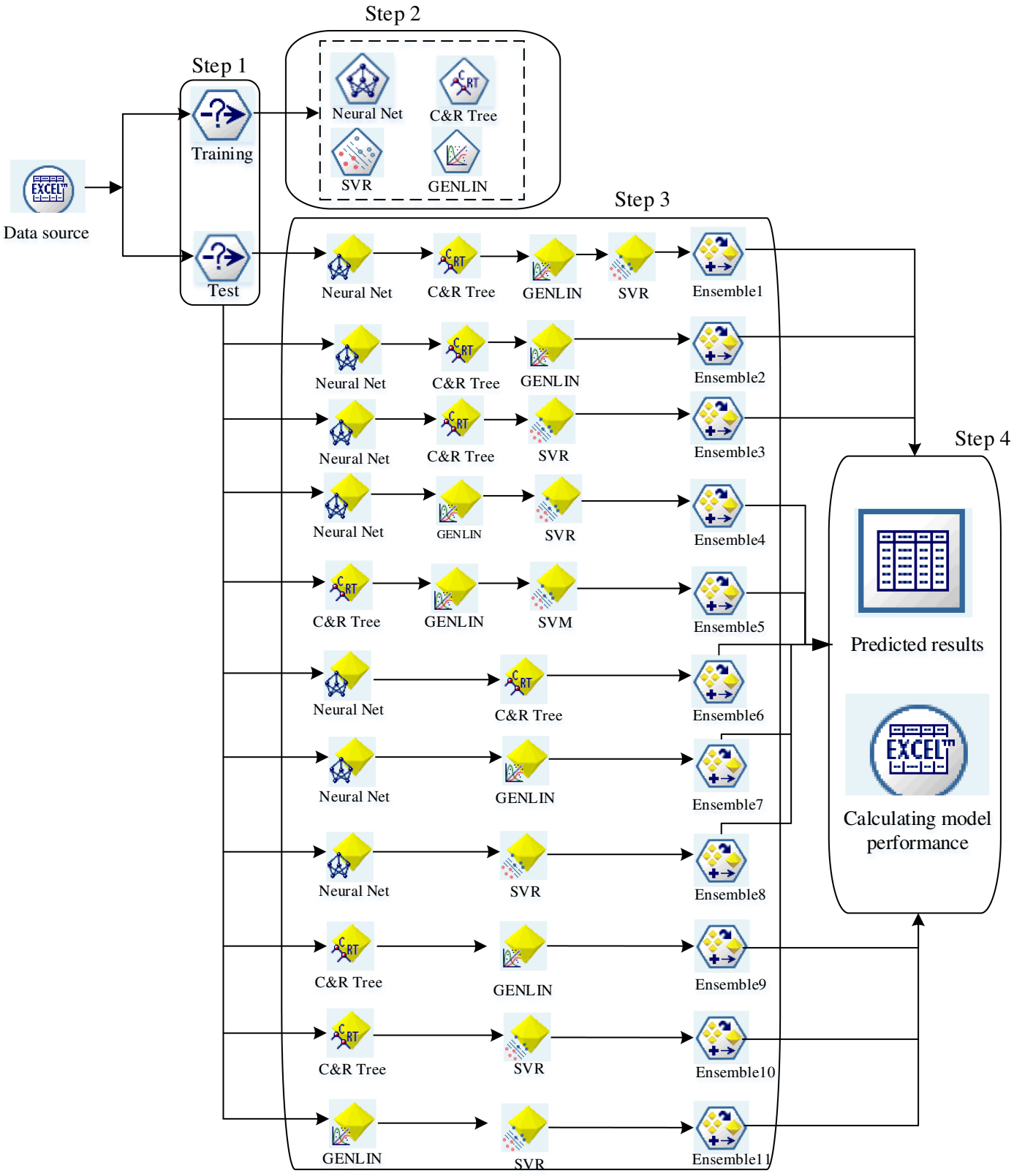

Fig. 3. Voting model stream

Metaensemble Models

In the voting method, the best model was the ANN+GENLIN+SVR ensemble model, which had an R value of 0.91 and a MAPE of 5.65%. When the bagging method was used, the CHAID model had the highest prediction accuracy (with an R value of 0.85 and a MAPE of 6.82%). The stacking method using REG obtained an R value of 0.91 and a relatively satisfactory MAPE of 5.15%. When the tiering method was used, the samples divided into two classes using the SVM-(CART/CHAID) model had the highest prediction accuracy (with an R value of 0.86 and a MAPE of 3.88%).

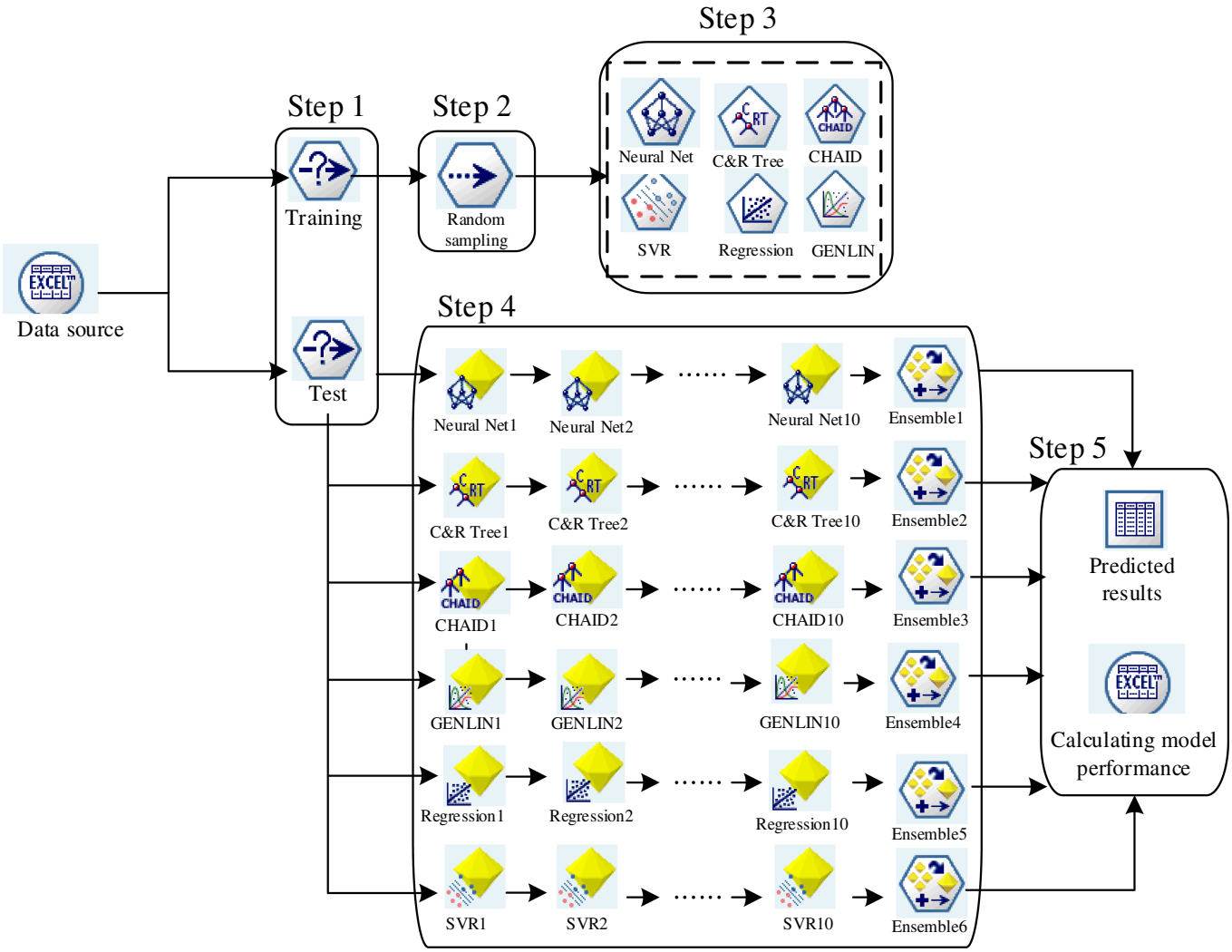

Fig. 4. Bagging model stream

Model Results: Group 3

Baseline Prediction Models

The best baseline model was the ANN model, which had an R value of 0.89 and a MAPE of 6.60%. Table 5 compares the prediction accuracy of the other models.

Metaensemble Models

When the voting method was used, the best model was the ANN+CART+SVR ensemble model, which had an R value of 0.92 and a MAPE of 5.46%. When the bagging method was used, the best model was the CART model, which had an R value of 0.86 and a relatively unsatisfactory MAPE of 6.50%. When the stacking method was used, the best model was the GENLIN model, which had a maximal R value of 0.92 and a MAPE of 5.07%. When the tiering method was used, the SVM-(^{na}/CART/CHAID/^{na}) model divided the samples into four classes and had an R value of 0.87 and a MAPE of only 3.65%.

Comparison of Predicted Results

Table 6 compares the prediction results obtained by the best baseline and metaensemble models (Table 5) in each of the three groups with those calculated by Eqs. (1) and (2). The comparisons indicated that the Tiering SVM-(SVR/SVR) model in Group 1 has the best performance, exhibiting an R of 0.89 and a MAPE of only 3.27%. Groups 2 and 3 had very similar results because only one factor differed between them. Moreover, the two different factors ($c_{i,\phi}$ in Group 2 and δ in Group 3) are related, i.e., one can be used to calculate the other. The relationship between these two factors is as follows: $c_{i,\phi} = \tan\delta / \tan\phi$.

The prediction accuracy of theoretical and empirical models proposed in previous studies was also evaluated. The R values calculated for the Michalowski and Cermak model [Eq. (1a)] and for the Zornberg model [Eq. (2)] were 0.66 and 0.75-0.79, respectively, and the MAPE values were between 10.6% and 9.1–9.8%. In the discrete model [Eq. (2)], $\alpha = 1$ generates better prediction results compared with $\alpha = 0.5$, demonstrating that the high efficiency of fiber to soil shear strength improvement. Other studies have also recommended a setting of $\alpha = 1$ in the Zornberg model (Li and Zornberg 2013; Najjar et al. 2013; Zornberg 2002). Apparently, the discrete model outperforms the energy-based model [Eq. (1a)], regardless of the assumed α value.

Comparisons of the theoretical and empirical models with the AI models in Table 6 indicate that prediction accuracy is greatly improved by the metaensemble models even for large confining pressure and fiber aspect ratio that are beyond the estimation limit of theoretical and empirical equations. Table 6 shows that all the metaensemble models had higher prediction accuracy compared

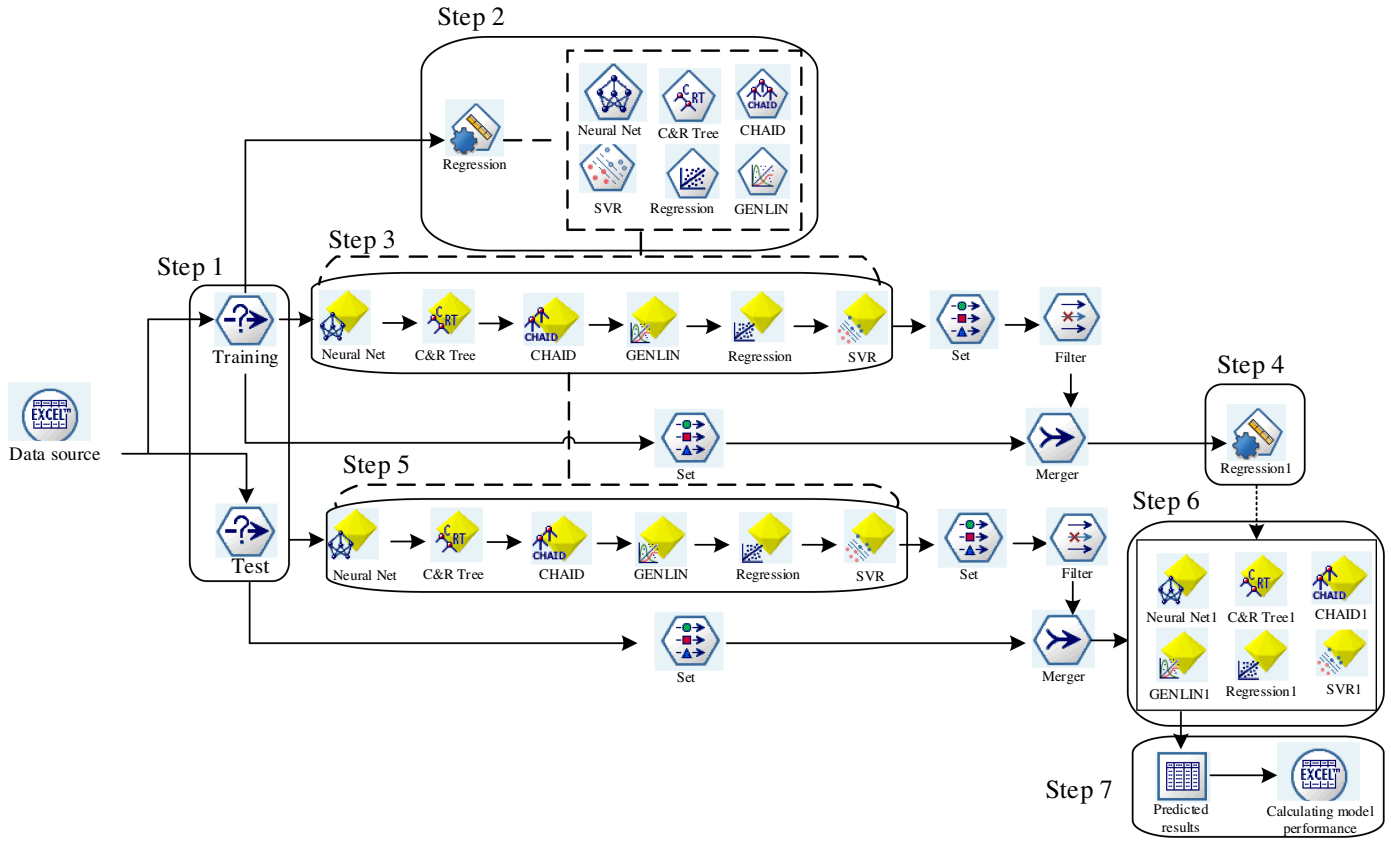

Fig. 5. Stacking model stream

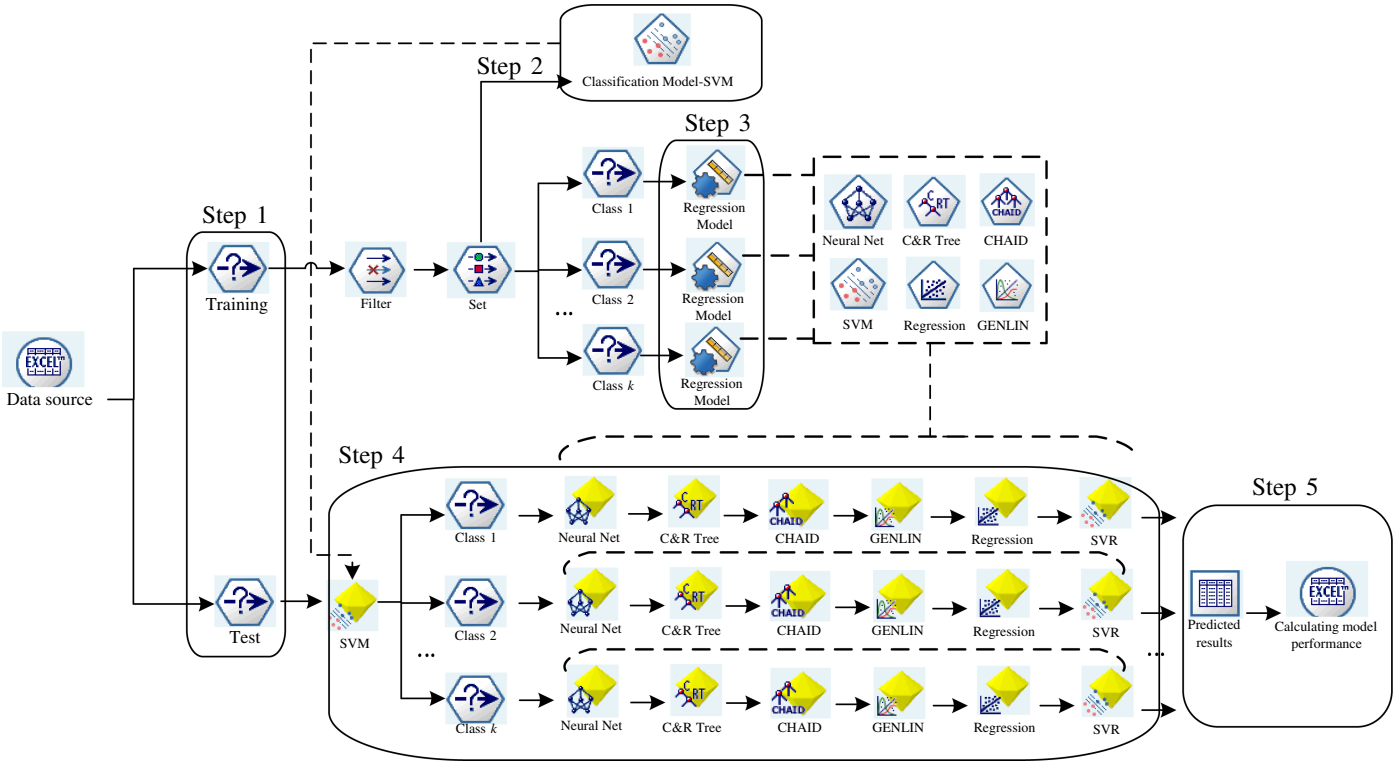

Fig. 6. Tiering model stream

Table 4. Data Distribution in Tiering Method

Class number	2 classes	3 classes	4 classes	5 classes
1	250	21	0	0
2	50	275	250	118
3	—	4	50	164
4	—	—	0	18
5	—	—	—	0
Total		300		

with the theoretical and empirical models. Specifically, the following coefficients were significantly increased: the R values increased 9.31–35.63%, RMSE increased 58.40–73.43%, MAE increased 70.04–79.50%, and MAPE increased 57.36–69.15%.

Fig. 7 further shows the results of performance tests of the Michalowski and Cermak model, the Zornberg model and the Tiering SVM-(SVR/SVR) model in Group 1. The comparison clearly shows that the data points predicted by the tiering SVM-(SVR/SVR)

Table 5. Prediction Performance Using the Baseline and Metaensemble Models through 10-Fold Cross-Validation

Method	Model	R	RMSE (°)	MAE (°)	MAPE (%)	SI_{Group}	$RANK_{Group}$	$SI_{Overall}$	$RANK_{Overall}$
Group 1									
Single	SVR	0.90	3.44	1.90	5.61	0.84	8	0.58	18
Voting	CART+GENLIN+SVR	0.92	2.91	1.85	5.01	0.62	5	0.43	10
Bagging	SVR	0.94	2.46	1.38	4.09	0.28	3	0.20	4
Stacking	CART	0.90	2.54	2.05	5.56	0.74	7	0.49	15
Tiering (2)	**SVM-(SVR/SVR)**	**0.89**	**1.98**	**1.07**	**3.27**	0.22	1	0.12	**1**
Tiering (3)	SVM-(REG/CART/GENLIN)	0.89	3.00	1.91	4.40	0.71	6	0.47	13
Tiering (4)	**SVM-(na/SVR/SVR/na)**	**0.89**	**1.98**	**1.41**	**3.27**	0.30	4	0.18	**3**
Tiering (5)	**SVM-(na/REG/CHAID/GENLIN/na)**	**0.86**	**1.92**	**1.17**	**2.79**	0.28	2	0.14	**2**
Group 2									
Single	ANN	0.89	3.56	2.47	6.58	0.73	7	0.76	21
Voting	ANN+GENLIN+SVR	0.91	3.18	2.10	5.65	0.49	5	0.57	17
Bagging	CHAID	0.85	3.80	2.45	6.82	0.87	8	0.87	24
Stacking	REG	0.91	3.02	1.92	5.15	0.38	4	0.48	14
Tiering (2)	SVM-(CART/CHAID)	0.86	2.28	1.27	3.88	0.13	1	0.27	7
Tiering (3)	SVM-(REG/CART/GENLIN)	0.85	3.43	2.25	5.14	0.63	6	0.69	19
Tiering (4)	SVM-(na/CART/SVM/na)	0.83	2.46	1.31	3.96	0.22	2	0.35	8
Tiering (5)	SVM-(na/CART/CHAID/SVR/na)	0.77	2.28	1.49	3.47	0.30	3	0.41	9
Group 3									
Single	ANN	0.89	3.57	2.48	6.60	0.75	7	0.77	22
Voting	ANN+CART+SVR	0.92	3.20	2.05	5.46	0.48	5	0.54	16
Bagging	CART	0.86	3.91	2.24	6.50	0.81	8	0.81	23
Stacking	GENLIN	0.92	2.94	1.88	5.07	0.36	3	0.44	11
Tiering (2)	SVM-(CART/CHAID)	0.86	2.24	1.23	3.83	0.14	2	0.25	6
Tiering (3)	SVM-(REG/CART/GENLIN)	0.85	3.45	2.26	5.18	0.65	6	0.69	20
Tiering (4)	SVM-(na/CART/CHAID/na)	0.87	2.18	1.19	3.65	0.09	1	0.21	5
Tiering (5)	SVM-(na/CART/CHAID/CART/na)	0.78	2.39	1.62	3.73	0.37	4	0.46	12

Note: The methods that had the best prediction performance are indicated by bold font. na = no data are available for that class; $RANK_{Group}$ = ranking among the group; $RANK_{Overall}$ = ranking among the overall models; SI_{Group} = synthesis index within the group; $SI_{Overall}$ = synthesis index within the overall models.

Table 6. Performance Comparison of Energy-Based Model, Discrete Model, and Best Metaensemble Models

	10-fold cross-validation performance				Improvement			
Model	R	RMSE (°)	MAE (°)	MAPE (%)	R (%)	RMSE (%)	MAE (%)	MAPE (%)
Michalowski and Cermak (2003)	0.66	7.45	5.22	10.60	—	—	—	—
Zornberg (2002), $\alpha = 0.5$	0.75	6.52	4.47	9.10	—	—	—	—
Zornberg (2002), $\alpha = 1$	0.79	5.48	4.24	9.80	—	—	—	—
Best model in Group 1; Tiering SVM-(SVR/SVR)	0.89	1.98	1.07	3.27	35.63	73.43	79.50	69.15
					17.97	69.62	76.04	64.07
					13.12	63.87	74.76	66.63
Best model in Group 2; Tiering SVM-(CART/CHAID)	0.86	2.28	1.27	3.88	31.06	69.40	75.66	63.40
					13.99	65.02	71.56	57.36
					9.31	58.40	70.04	60.41
Best model in Group 3; Tiering SVM-(na/CART/CHAID/na)	0.87	2.18	1.19	3.65	31.06	69.40	75.66	63.40
					15.40	66.53	73.39	59.89
					10.66	60.20	71.97	62.76

Note: na = no data are available for that class.

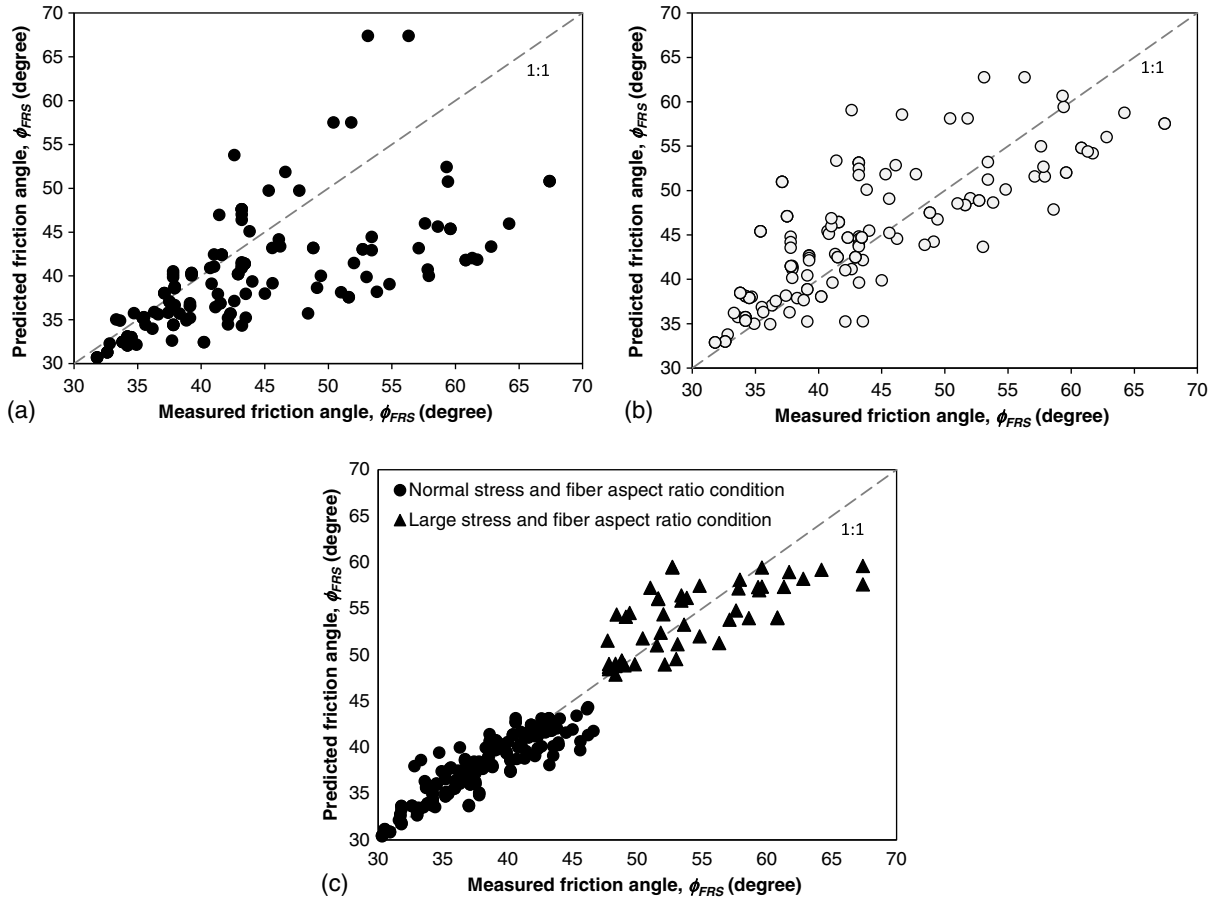

Fig. 7. Performance comparison: (a) Michalowski and Čermák model (energy-based model); (b) Zornberg model using $\alpha = 1$ (discrete model); (c) Tiering SVM-(SVR/SVR) in Group 1

model follows 1:1 line closely, but the data points are relatively scattered in the Michalowski-Čermák and Zornberg models.

Conclusions and Recommendation

This study compiled high quality data for triaxial and direct shear tests of FRS reported in the literature during 1983–2015. The resulting database included information on the properties of sand, fibers, soil-fiber interface, and stress parameters. Data-mining technology and the baseline models (i.e., ANN, CART, CHAID, REG, GENLIN, SVR/SVM) were used to explore hidden relationships among the variables, thereby building a model for predicting the friction angle (i.e., shear strength parameter) of FRS.

The baseline models were then used to design a series of metaensemble models (i.e., voting, bagging, stacking, and tiering). The prediction results obtained by the theoretical and empirical models were then compared with those obtained by the AI models proposed in this study. The comparison results showed that the metaensemble models had greatly improved accuracy in predicting FRS shear strength. The results showed that the prediction performance levels of all of the metaensemble models are more favorable than those of the baseline models.

For the baseline single models, analytical results of Group 1, with considering 300 samples and all 15 inputs, confirmed that the SVR is superior to the others in predicting shear strength of FRS with the average MAPE of 5.61%. Moreover, the comparison results show that the ANN model outperformed the others, with the average MAPE of 6.58 and 6.60% for Group 2 and Group 3, respectively.

Regarding the voting method, the ensemble model of CART+GENLIN+SVR showed the best prediction performance (with a MAPE of 5.01%) among the others in Group 1; whereas the ensemble models of ANN+GENLIN+SVR and ANN+CART+SVR obtained the highest accuracy compared with the other models in Group 2 and Group 3 (with a MAPE of 5.65 and 5.46%, respectively). For the bagging method, the SVR model, CHAID model, and CART model yielded the highest prediction performance compared with the others in Groups 1, 2, and 3, respectively. For the stacking method, the CART model, REG model, and GENLIN model were superior to the others in Groups 1, 2, and 3, respectively.

Particularly, the optimal metaensemble model was the Tiering SVM-SVR/SVR model, which had an R up to 0.89, a MAPE of only 3.27%, and highly accurate predictions. The comparison results also revealed that all metaensemble models had higher prediction accuracy compared with the theoretical and empirical models, regardless of normal confining pressure and fiber aspect ratio conditions. The values of the following coefficients increased significantly: the R values increased 9.31–35.63%, RMSE increased 58.40–73.43%, MAE increased 70.04–79.50%, and MAPE increased 57.36–69.15%.

This study demonstrated that AI models provide an effective alternative for FRS shear strength prediction and, thereby, facilitate assessment of the stability of FRS structures in design and promoting the application of fiber reinforcement into a wide range of geotechnical projects. Compared with the conventional theoretical and empirical models, the AI models could produce more accurate predictions, resulting in more cost-effective and reliable designs. In addition, the prediction results from AI models can aid assessment

of experimental designs during preplanning, providing researchers with a basis for verification. The AI models can also be used to plan and design teams as a feedback, serving as information aids in decision making.

Although the developed AI model demonstrated high prediction accuracy, its applicability is limited to FRS with parameter values within the ranges in the compiled database in this study (shown in Tables 1 and 2). In particular, the developed model is only applicable when the FRS is granular rather than fine grained and cohesive.

For further optimization of the prediction models and enhancement of the prediction efficacy of the AI models, nature-inspired metaheuristic optimization algorithms can be devised to fine-tune their default parameters. Moreover, an information exchange platform, combining the prediction results with engineering practices, can be developed for use to support decision making by engineers and design teams.

References

Ahmad, F., Bateni, F., and Azmi, M. (2010). "Performance evaluation of silty sand reinforced with fibres." *Geotext. Geomembr.*, 28(1), 93–99.

Alkroosh, I., and Nikraz, H. (2011). "Simulating pile load-settlement behavior from CPT data using intelligent computing." *Open Eng.*, 1(3), 295–305.

Al-Refeai, A.-S. (1998). "Dynamic and static characterization of polypropylene fiber-reinforced dune sand." *Geosynth. Int.*, 5(5), 443–458.

Al-Refeai, T. O. (1991). "Behavior of granular soils reinforced with discrete randomly oriented inclusions." *Geotext. Geomembr.*, 10(4), 319–333.

Anggraini, V., Asadi, A., Huat, B. B. K., and Nahazanan, H. (2015). "Effects of coir fibers on tensile and compressive strength of lime treated soft soil." *Meas*, 59, 372–381.

Austin, P. C., Tu, J. V., Ho, J. E., Levy, D., and Lee, D. S. (2013). "Using methods from the data-mining and machine-learning literature for disease classification and prediction: A case study examining classification of heart failure subtypes." *J. Clin. Epidemiol.*, 66(4), 398–407.

Benardos, A. G., and Kaliampakos, D. C. (2004). "Modelling TBM performance with artificial neural networks." *Tunnelling Underground Space Technol.*, 19(6), 597–605.

Cao, M. S., et al. (2015). "Neural network ensemble-based parameter sensitivity analysis in civil engineering systems." *Neural Comput. Appl.*, 8–1.

Chauhan, M. S., Mittal, S., and Mohanty, B. (2008). "Performance evaluation of silty sand subgrade reinforced with fly ash and fibre." *Geotext. Geomembr.*, 26(5), 429–435.

Chen, C.-W. (2007). "A constitutive model for fiber-reinforced soils." Ph.D. dissertation, Univ. of Missouri, Columbia, MO.

Chen, R.-H., Chi, P.-C., Wu, T.-C., and Ho, C.-C. (2011). "Shear strength of continuous-filament reinforced sand." *J. Geo-Eng.*, 6(2), 99–107.

Chou, J.-S., and Lin, C. (2013). "Predicting disputes in public-private partnership projects: Classification and ensemble models." *J. Comput. Civ. Eng.*, 10.1061/(ASCE)CP.1943-5487.0000197, 51–60.

Chou, J.-S., Ngo, N.-T., and Pham, A.-D. (2016). "Shear strength prediction in reinforced concrete deep beams using nature-inspired metaheuristic support vector regression." *J. Comput. Civ. Eng.*, 10.1061/(ASCE)CP.1943-5487.0000466, 04015002.

Chou, J.-S., and Pham, A.-D. (2013). "Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength." *Constr. Build. Mater.*, 49, 554–563.

Chou, J.-S., and Pham, A.-D. (2015). "Smart artificial firefly colony algorithm-based support vector regression for enhanced forecasting in civil engineering." *Comput.-Aided Civ. Infrastruct. Eng.*, 30(9), 715–732.

Chou, J.-S., and Tsai, C.-F. (2012). "Concrete compressive strength analysis using a combined classification and regression technique." *Autom. Constr.*, 24, 52–60.

Chou, J.-S., Yang, K.-H., Pampang, J. P., and Pham, A.-D. (2015). "Evolutionary metaheuristic intelligence to simulate tensile loads in reinforcement for geosynthetic-reinforced soil structures." *Comput. Geotech.*, 66, 1–15.

Consoli, N., Casagrande, M., and Coop, M. (2005). "Effect of fiber reinforcement on the isotropic compression behavior of a sand." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)1090-0241(2005)131:11(1434), 1434–1436.

Consoli, N., Casagrande, M., and Coop, M. (2007). "Performance of a fibre-reinforced sand at large shear strains." *Géotechnique*, 57(9), 751–756.

Consoli, N., Festugato, L., and Heineck, K. (2009a). "Strain-hardening behaviour of fibre-reinforced sand in view of filament geometry." *Geosynth. Int.*, 16(2), 109–115.

Consoli, N., Montardo, J., Donato, M., and Prietto, P. (2004). "Effect of material properties on the behaviour of sand–cement–fibre composites." *Proc. Inst. Civ. Eng. Ground Improv.*, 8(2), 77–90.

Consoli, N., Prietto, P., and Ulbrich, L. (1998). "Influence of fiber and cement addition on behavior of sandy soil." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)1090-0241(1998)124:12(1211), 1211–1214.

Consoli, N. C., Vendruscolo, M. A., Fonini, A., and Rosa, F. D. (2009b). "Fiber reinforcement effects on sand considering a wide cementation range." *Geotext. Geomembr.*, 27(3), 196–203.

Das, S. K. (2012). "Artificial neural networks in geotechnical engineering: modeling and application issues." Chapter 10, *Metaheuristics water, geotechnical transportation engineering*, Elsevier, Oxford, U.K., 231–270.

Freitag, D. (1986). "Soil randomly reinforced with fibers." *J. Geotech. Engrg.*, 10.1061/(ASCE)0733-9410(1986)112:8(823), 823–826.

Goh, A. (1994). "Seismic liquefaction potential assessed by neural networks." *J. Geotech. Engrg.*, 10.1061/(ASCE)0733-9410(1994)120:9(1467), 1467–1480.

Goh, A. T., Kulhawy, F. H., and Chua, C. (2005). "Bayesian neural network analysis of undrained side resistance of drilled shafts." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)1090-0241(2005)131:1(84), 84–93.

Gray, D., and Al-Refeai, T. (1986). "Behavior of fabric-versus fiber-reinforced sand." *J. Geotech. Engrg.*, 10.1061/(ASCE)0733-9410(1986)112:8(804), 804–820.

Gray, D., and Ohashi, H. (1983). "Mechanics of fiber reinforcement in sand." *J. Geotech. Engrg.*, 10.1061/(ASCE)0733-9410(1983)109:3(335), 335–353.

Gregory, G. H. (2006). "Shear strength, creep and stability of fiber-reinforced soil slopes." Ph.D. dissertation, Oklahoma State Univ., Stillwater, OK.

Hejazi, S. M., Sheikhzadeh, M., Abtahi, S. M., and Zadhoush, A. (2012). "A simple review of soil reinforcement by using natural and synthetic fibers." *Constr. Build. Mater.*, 30, 100–116.

IBM. (2009). *Clementine 12.0 algorithms guide*, Integral Solutions, Chicago.

Ibraim, E., and Fourmont, S. (2007). "Behaviour of sand reinforced with fibres." *Soil stress-strain behavior: Measurement, modeling and analysis*, Vol. 146, Springer, Dordrecht, Netherlands, 807–818.

Jahed Armaghani, D., Hajihassani, M., Yazdani Bejarbaneh, B., Marto, A., and Tonnizam Mohamad, E. (2014). "Indirect measure of shale shear strength parameters by means of rock index tests through an optimized artificial neural network." *Meas.*, 55, 487–498.

Kaniraj, S., and Havanagi, V. (2001). "Behavior of cement-stabilized fiber-reinforced fly ash-soil mixtures." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)1090-0241(2001)127:7(574), 574–584.

Kass, G. V. (1980). "An exploratory technique for investigating large quantities of categorical data." *Appl. Stat.*, 29(2), 119–127.

Kohavi, R. (1995). "*A study of cross-validation and bootstrap for accuracy estimation and model selection*, Stanford Univ., Stanford, CA.

Kumar, S., and Tabor, E. (2003). "Strength characteristics of silty clay reinforced with randomly oriented nylon fibers." *Electron. J. Geotech. Eng.*, 8(2), 1–10.

Li, C., and Zornberg, J. (2013). "Mobilization of reinforcement forces in fiber-reinforced soil." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)GT.1943-5606.0000745, 107–115.

Li, J., Tang, C., Wang, D., Pei, X., and Shi, B. (2014). "Effect of discrete fibre reinforcement on soil tensile strength." *J. Rock Mech. Geotech. Eng.*, 6(2), 133–137.

Loehr, J., Romero, R., and Ang, E. (2005). "Development of a strain-based model to predict strength of geosynthetic fiber-reinforced soil." *Geosynth. Res. Dev. Prog.*, 1–7.

Maher, M., and Gray, D. (1990). "Static response of sands reinforced with randomly distributed fibers." *J. Geotech. Engrg.*, 10.1061/(ASCE)0733-9410(1990)116:11(1661), 1661–1677.

Mathur, A., and Foody, G. M. (2008). "Multiclass and binary SVM classification: Implications for training and classification users." *IEEE Geosci. Remote Sens. Lett.*, 5(2), 241–245.

Michalowski, R., and Čermák, J. (2003). "Triaxial compression of sand reinforced with fibers." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)1090-0241(2003)129:2(125), 125–136.

Michalowski, R., and Zhao, A. (1996). "Failure of fiber-reinforced granular soils." *J. Geotech. Engrg.*, 10.1061/(ASCE)0733-9410(1996)122:3(226), 226–234.

Mortazavian, S., and Fatemi, A. (2015). "Fatigue behavior and modeling of short fiber reinforced polymer composites: A literature review." *Int. J. Fatigue*, 70, 297–321.

Najjar, S., Sadek, S., and Alcovero, A. (2013). "Quantification of model uncertainty in shear strength predictions for fiber-reinforced sand." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)GT.1943-5606.0000742, 116–133.

Nataraj, M. S., and McManis, K. L. (1997). "Strength and deformation properties of soils reinforced with fibrillated fibers." *Geosynth. Int.*, 4(1), 65–79.

Nelder, J., and Wedderburn, R. (1972). "Generalized linear models." *J. R. Stat. Soc. A*, 135(3), 370–384.

Park, H. I., and Kim, Y. T. (2011). "Prediction of strength of reinforced lightweight soil using an artificial neural network." *Eng. Comput.*, 28(5), 600–615.

Prodromidis, A. L., and Stolfo, S. J. (1999). "A comparative evaluation of meta-learning strategies over large and distributed data sets." *Proc., Workshop on Meta-Learning, 16th Int. Conf. Machine Learning*, CiteseerX, State College, PA, 18–27.

Ranjan, G., Vasan, R., and Charan, H. (1996). "Probabilistic analysis of randomly distributed fiber-reinforced soil." *J. Geotech. Engrg.*, 10.1061/(ASCE)0733-9410(1996)122:6(419), 419–426.

Ranjan, G., Vasan, R. M., and Charan, H. D. (1994). "Behaviour of plastic-fibre-reinforced sand." *Geotext. Geomembr.*, 13(8), 555–565.

Sadek, S., Najjar, S., and Freiha, F. (2010). "Shear strength of fiber-reinforced sands." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)GT.1943-5606.0000235, 490–499.

Samui, P. (2013). "Slope stability analysis using multivariate adaptive regression spline." *Metaheuristics in water, geotechnical and transport engineering*, X.-S. Yang, A. H. Gandomi, S. Talatahari, and A. H. Alavi, eds., Elsevier, Oxford, U.K., 327–342.

Santoni, R., Tingle, J., and Webster, S. (2001). "Engineering properties of sand-fiber mixtures for road construction." *J. Geotech. Geoenviron. Eng.*, 10.1061/(ASCE)1090-0241(2001)127:3(258), 258–268.

Shao, W., Cetin, B., Li, Y., Li, J., and Li, L. (2014). "Experimental investigation of mechanical properties of sands reinforced with discrete randomly distributed fiber." *Geotech. Geol. Eng.*, 32(4), 901–910.

Sivakumar Babu, G. L., Vasudevan, A. K., and Haldar, S. (2008). "Numerical simulation of fiber-reinforced sand behavior." *Geotext. Geomembr.*, 26(2), 181–188.

Sykes, A. O. (1993). *An introduction to regression analysis*, Law School, Univ. of Chicago, Chicago.

Timofeev, R. (2004). *Classification and regression trees (cart) theory and applications*, Humboldt Univ., Berlin.

Tinoco, J., Gomes Correia, A., and Cortez, P. (2014). "Support vector machines applied to uniaxial compressive strength prediction of jet grouting columns." *Comput. Geotech.*, 55, 132–140.

Tsompanakis, Y., Lagaros, N. D., Psarropoulos, P. N., and Georgopoulos, E. C. (2009). "Simulating the seismic response of embankments via artificial neural networks." *Adv. Eng. Software*, 40(8), 640–651.

Vapnik, V. N. (1995). *The nature of statistical learning theory*, Springer, New York.

Wei, X.-K., Li, Y.-H., Li, Y.-F., and Zhang, D.-F. (2008). "Enclosing machine learning: concepts and algorithms." *Neural Comput. Appl.*, 17(3), 237–243.

Wolpert, D. H. (1992). "Stacked generalization." *Neural Netw.*, 5(2), 241–259.

Wu, X., et al. (2008). "Top 10 algorithms in data mining." *Knowl. Inf. Syst.*, 14(1), 1–37.

Yetimoglu, T., and Salbas, O. (2003). "A study on shear strength of sands reinforced with randomly distributed discrete fibers." *Geotext. Geomembr.*, 21(2), 103–110.

Yu, Y., Zhang, B., and Yuan, H. (2007). "An intelligent displacement back-analysis method for earth-rockfill dams." *Comput. Geotech.*, 34(6), 423–434.

Zornberg, J. (2002). "Discrete framework for limit equilibrium analysis of fibre-reinforced soil." *Géotech.*, 52(8), 593–604.