Shear Strength Prediction of Treated Soft Clay with Sugarcane Bagasse Ash Using Artificial Intelligence Methods

Rufaizal Che Mamat*, Azuin Ramlia & Sri Atmaja P. Rosyidib

aCentre of Green Technology for Sustainable Cities, Department of Civil Engineering, Politeknik Ungku Omar, Jalan Raja Musa Mahadi, 31400 Ipoh, Perak, Malaysia
bDepartment of Civil Engineering, Universitas Muhammadiyah Yogyakarta, Jalan Bravijaya, Bantul, Yogyakarta, 55183 Indonesia

*Corresponding author: rufaizal.cm@gmail.com

Received 5 August 2022, Received in revised form 25 November 2022
Accepted 14 December 2022, Available online 30 May 2023

ABSTRACT

Soil shear strength is an essential engineering characteristic used in designing and evaluating geotechnical structures. In this study, we intend to analyse and compare the performance of the Genetic Algorithm - Adaptive Network-based Fuzzy Inference System (GANFIS) and Artificial Neural Networks (ANN) in predicting the strength of soft clay. Case studies of 144 soft clay soil samples from Sarang Buaya, Semerah, Malaysia, were utilised to generate training and testing datasets for developing and validating models. RMSE and R have been employed to validate and compare the models. The GANFIS has the highest prediction capability ($RMSE=0.042$ and $R=0.850$), while the ANN has the lowest ($RMSE=0.065$ and $R=0.49$). From a comparison of the two models, it can be stated that GANFIS is the most promising technique for predicting the strength of soft clay.

Keywords: Shear strength; Soft clay; Sugarcane bagasse ash; Artificial neural networks; Adaptive network based fuzzy inference system; Genetic algorithm

INTRODUCTION

Shear strength of the soil is an important engineering parameter utilized in the design and audit of many geotechnical constructions, such as road foundations and pavements, earth dams, and retaining walls. Soil shear strength is determined by internal friction angle and unit cohesion and is affected by plastic index, liquid limit, moisture content, and clay content (R.C. Mamat et al. 2019). In the investigation of the soft clay characteristics, it was found that it is usually described as having low shear strength. The soil’s shear strength is less than 40 kPa and can be moulded with light finger pressure. Insufficient bearing capability, severe post-construction settlement, and excavation and embankment instability are common in this deposit. Thus, improving material characteristics is needed. Soil stabilization is a way to make the soil stronger, able to hold more weight, and last longer in wet and rocky conditions.

Consideration of the beneficial reuse of industrial waste products is a recent trend in soil stabilization. Stabilized soils with cement have good compressive strength but weak tensile and flexural strengths and behave in a brittle way (Rufaizal Che Mamat et al. 2020; Rufaizal Che Mamat, Ramli, Khahro et al. 2022). Several studies on synthetic fibres’ effect on stabilized soils have been based on unconfined compressive strength tests (Correia et al. 2015; George et al. 2020). In general, investigations using soft soils stabilized with a low binder content (less than 8 %) demonstrate that increasing the total of fibres increases compressive and tensile strength. In contrast, Consoli et al. (2009) found that polypropylene fibres’ reinforcing impact diminishes as the cement level increases, resulting in a decrease in the peak strength of up to 6% of cement content for a stabilized soft soil. For this reason, the polypropylene fibres are unable to mobilize the tensile strength prior to peak failure because of the low deformations obtained when the cement concentration is high. As a result, the stabilized soil becomes extremely stiff. Therefore, adding more fibres to a mix with a high cement percentage does not affect the final strength.

A high binder percentage is often necessary to provide a sufficient amount of strength when used to stabilize soft soils. There has been a lack of attention paid to the role of fibres in stabilizing soft soils with a high binder concentration, maybe because the stabilized soil becomes quite stiff with such amounts of binder.

The use of sugarcane bagasse ash (SBA) in stabilized soft soil has been intensively researched in the last few years. The results show that SBA can somewhat replace some of the cement used in soft soil stabilization. As much as 20% of the cement in soft soil could be replaced with SBA. One metric ton of sugarcane generates 280 kilograms of bagasse, a by-product of the sugarcane industry (Sun et al. 2004). Bagasse ash is a by-product of the sugar mills'
use of sugarcane bagasse as a fuel. Bagasse disposal is a major concern since it pollutes the atmosphere. While the sugarcane industry recommended using this ash as fertilizer on the plantation, there is not enough mineral nutrition in this ash to support the plantation’s growth. Industries are still searching for solutions to the problem of how to dispose of the waste they produce (Sales & Araújo Lima 2010).

Soft soil shear strength prediction has been studied extensively. Motaghedi & Eslami, (2014) developed an analytical approach for parameter shear strength prediction incorporating failure mechanisms at the cone tip and penetrometer sleeve. McGann et al. (2015) used multiple linear regression to predict soil shear wave velocities from cone penetration test data in Christchurch. Azari et al. (2016) evaluated shear strength variation in the disturbed zone on soft soil deposits improved with vertical drains and preloading. Griffiths et al. (2016) employed linear and nonlinear 1D site response assessments for Treasure Island to show modelling issues for soft soil locations. Oliveira et al. (2017) studied constitutive models to mimic soft soil creep in its natural or chemically stabilized state. These investigations provide a well-established mathematical paradigm for accurate prediction.

In the last few decades, many prediction models of material properties have been made using machine learning or artificial intelligence (Rufaizal Che Mamat et al. 2019a, 2019b; T.-A. Nguyen et al. 2020). There are numerous studies on predicting the shear strength of the soil. Tran et al. (2022) investigated the ability of artificial neural networks (ANN) to estimate the soil’s residual strength. Kanungo et al. 2014 analyzed the ANN and regression tree (CART) techniques by comparing the ANN and CART strategies for predicting shear strength parameters. Mamat et al. (2020) used ANN to estimate the safety factor of embankment stability. Khan et al. (2016) explored the prediction of the residual strength of clay using a new prediction model, a functional network. Machine learning algorithms are generally effective for predicting the shear strength of soft soils, as established by the research mentioned above (Rufaizal Che Mamat et al. 2021).

A new generation of potential soft computing techniques, such as the Genetic Algorithm-Adaptive Network-based Fuzzy Inference System (GANFIS), has emerged due to recent advances in machine learning and optimization. Modern techniques called GANFIS were developed by integrating meta-heuristic optimization algorithms and fuzzy neural models. They have been demonstrated to be effective at predicting a variety of civil engineering issues, including road surface (H.-L. Nguyen et al. 2019), water quality (Azad et al. 2018), concrete technology (Shaban et al. 2021), and building construction (Bozorgvar & Zahrai 2019). On the other hand, ANN is a popular and effective modeling technique for shear strength. This method has not been investigated and compared with common machine-learning techniques for predicting the shear strength of improved soft soils.

In this study, we investigate and compare the prediction performance of GANFIS and ANN for the prediction of shear strength of improved soft clay with sugarcane bagasse ash in order to contribute to the body of knowledge (SBA-SC). The comparison of such artificial intelligence systems is essential for determining an appropriate prediction model for SBA-SC shear strength applicable in real-world circumstances.

<table>
<thead>
<tr>
<th>Properties</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liquid limit (%)</td>
<td>45</td>
</tr>
<tr>
<td>Plastic limit (%)</td>
<td>24</td>
</tr>
<tr>
<td>Plasticity index (%)</td>
<td>21</td>
</tr>
<tr>
<td>Linear shrinkage (%)</td>
<td>6</td>
</tr>
<tr>
<td>Moisture content (%)</td>
<td>8</td>
</tr>
<tr>
<td>Optimum moisture content (%)</td>
<td>15</td>
</tr>
<tr>
<td>Maximum dry density (kg/m³)</td>
<td>18.5</td>
</tr>
<tr>
<td>Specific gravity</td>
<td>2.64</td>
</tr>
</tbody>
</table>

**TABLE 1. Physical properties of soft clay**

**MATERIALS AND METHODS**

In this study, soft clay soil samples from a palm oil plantation in Sarang Buaya, Semerah, Malaysia, were used as a case study. A total of 144 samples were obtained from the site projects and used to make the datasets for modelling. Table 1 summarises the physical properties of soft clay, while Figure 1 depicts the particle size distribution curve. According to the AASHTO (1986) Soil Classification System, the soil is classified as A-7-5, and according to the Unified Soil Classification System (USCS), it is classified as CL (ASTM 1992). It contains no more than 2% clay by weight.

The bagasse ash was obtained from a small hawker of sugarcane water in Taman Cempaka, Ipoh, Perak. It had a specific gravity of 1.88. The air-dried bagasse was burned in an incinerator that was made in the area, and the ash from the bagasse was put through a BS No. 200 sieve. Table 2 shows the oxide composition of SBA that can be found using X-ray fluorescence analysis.
The moisture-density relationship and unconfined compression were all tested on samples that had been air-dried for one day. Due to it being simple to achieve in the field, the British Standard Light compaction effort was used. Soft clay and SBA-stabilized soil samples were prepared and evaluated in accordance with BS 1377 (1990) and BS 1924 (1990). The unconfined compressive strength (UCS) tests were carried out on cylinder specimens (38 mm dia.). Before adding water to the relevant optimum moisture content (OMC), the requisite amounts of bagasse ash by dry weight of soil and soil were measured and blended in the dry condition in the production of all specimens. In order to avoid moisture loss, the SBA mixture was compacted into a removable mould; then, specimens were removed and covered in plastic sheets. In the case of UCS, the samples were air-cured for 7, 14, and 28 days, respectively.

In this prediction problem, the shear strength is the output variable, while the input variables are clay content, moisture content, plastic index, plastic limit, liquid limit, and consistency index. The variable data was separated into two parts: the training dataset (70 %) and the validation dataset (30 %). Different data division procedures were utilised to achieve the greatest fit for each model, and the statistical data values used for each model are provided in Table 3. The training dataset was then used to train models, while the validation dataset was used to test them.

<table>
<thead>
<tr>
<th>Statistical index</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GANFIS</td>
<td>ANN</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.122</td>
<td>0.122</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.325</td>
<td>0.277</td>
</tr>
<tr>
<td>Mean</td>
<td>0.214</td>
<td>0.245</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.052</td>
<td>0.048</td>
</tr>
</tbody>
</table>

The moisture-density relationship and unconfined compression were all tested on samples that had been air-dried for one day. Due to it being simple to achieve in the field, the British Standard Light compaction effort was used. Soft clay and SBA-stabilized soil samples were prepared and evaluated in accordance with BS 1377 (1990) and BS 1924 (1990). The unconfined compressive strength (UCS) tests were carried out on cylinder specimens (38 mm dia.). Before adding water to the relevant optimum moisture content (OMC), the requisite amounts of bagasse ash by dry weight of soil and soil were measured and blended in the dry condition in the production of all specimens. In order to avoid moisture loss, the SBA mixture was compacted into a removable mould; then, specimens were removed and covered in plastic sheets. In the case of UCS, the samples were air-cured for 7, 14, and 28 days, respectively.

In this prediction problem, the shear strength is the output variable, while the input variables are clay content, moisture content, plastic index, plastic limit, liquid limit, and consistency index. The variable data was separated into two parts: the training dataset (70 %) and the validation dataset (30 %). Different data division procedures were utilised to achieve the greatest fit for each model, and the statistical data values used for each model are provided in Table 3. The training dataset was then used to train models, while the validation dataset was used to test them.

<table>
<thead>
<tr>
<th>Chemical</th>
<th>Concentration (% by weight)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO₂</td>
<td>42</td>
</tr>
<tr>
<td>CaO</td>
<td>3</td>
</tr>
<tr>
<td>Fe₂O₃</td>
<td>3</td>
</tr>
<tr>
<td>Al₂O₃</td>
<td>7</td>
</tr>
<tr>
<td>K₂O</td>
<td>9</td>
</tr>
<tr>
<td>MgO</td>
<td>0.2</td>
</tr>
<tr>
<td>TiO₂</td>
<td>1</td>
</tr>
<tr>
<td>SO₃</td>
<td>0.04</td>
</tr>
<tr>
<td>LOI</td>
<td>18</td>
</tr>
</tbody>
</table>

The moisture-density relationship and unconfined compression were all tested on samples that had been air-dried for one day. Due to it being simple to achieve in the field, the British Standard Light compaction effort was used. Soft clay and SBA-stabilized soil samples were prepared and evaluated in accordance with BS 1377 (1990) and BS 1924 (1990). The unconfined compressive strength (UCS) tests were carried out on cylinder specimens (38 mm dia.). Before adding water to the relevant optimum moisture content (OMC), the requisite amounts of bagasse ash by dry weight of soil and soil were measured and blended in the dry condition in the production of all specimens. In order to avoid moisture loss, the SBA mixture was compacted into a removable mould; then, specimens were removed and covered in plastic sheets. In the case of UCS, the samples were air-cured for 7, 14, and 28 days, respectively.

In this prediction problem, the shear strength is the output variable, while the input variables are clay content, moisture content, plastic index, plastic limit, liquid limit, and consistency index. The variable data was separated into two parts: the training dataset (70 %) and the validation dataset (30 %). Different data division procedures were utilised to achieve the greatest fit for each model, and the statistical data values used for each model are provided in Table 3. The training dataset was then used to train models, while the validation dataset was used to test them.

<table>
<thead>
<tr>
<th>Statistical index</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GANFIS</td>
<td>ANN</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.122</td>
<td>0.122</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.325</td>
<td>0.277</td>
</tr>
<tr>
<td>Mean</td>
<td>0.214</td>
<td>0.245</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.052</td>
<td>0.048</td>
</tr>
</tbody>
</table>
The adaptive neuro-fuzzy inference system (ANFIS) is a neuro-fuzzy system that uses an ANN and a fuzzy system to build a powerful and successful prediction model in various domains. The ANFIS structure is made up of five levels, as seen in Figure 2. Layer 1 is the input for the subsequent layers. The input layer in this investigation is made up of 144 samples. This is an adaptive step in Layer 2. Each controlling factor’s membership value is calculated using membership functions (\(M_k\)). As stated, the Gaussian function was employed as the membership function (Equation 1). There are two antecedent parameters to tweak for each \(\mu M_k(X)\): \(M_k\) and \(\sigma_k\).

The output value is generated as a sum of \(\mu M_k(X)\); \(M_k\) and \(\sigma_k\).

\[
\mu M_k(X) = \exp \left( - \frac{(X-M_k)^2}{2\sigma_k^2} \right)
\]

Equation 2 was used in layer 3 to calculate preliminary weights.

\[
w_i = \mu M_k(X1) \times \mu M_k(X2) \times \mu M_k(X_m)
\]

In the layer 4:

\[
\bar{w}_i = \frac{w_i}{\text{sum}(w_i)}
\]

Then, layer 5:

\[
f_i = \bar{w}_i(a_0 + \sum (a_i x_i))
\]

The Genetic Algorithm (GA) is an optimization algorithm and search engine based on genetic and natural selection principles. GA begins with the formation of a population of solutions (individuals), each of which is represented by a chromosome. People from the population are employed to produce new individuals. This is done with the hope that the new population will outperform the old one. Individuals are chosen to make new individuals - offspring - are chosen to depend on their level of adaptation. The higher the adaptation, the more likely the individuals are employed to reproduce. This operation is continued until the specified requirements are met. This algorithm employs natural genetic processes such as selection, crossover, and mutation.

The root mean square error (RMSE) and correlation coefficient (R) are used to assess a model’s accuracy. These
three indicators are frequently used in model validation. The following are the formulas:

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \]  

(6)

\[ R = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}} \]  

(7)

\[ R^2 = \left( \frac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}} \right)^2 \]  

(8)

where \( y \) and \( \bar{y} \) are the measured and mean values of the soil shear strength, and \( \hat{y} \) and \( \bar{y} \) are model output values.

### RESULTS ANALYSIS

Two models, GANFIS and ANN, were trained and developed using training datasets to predict the strength of SBA-SC. For the GANFIS, an initial FIS model with initial parameters was generated, followed by the creation of a FIS structure based on a number of membership functions. The GA is then used to identify the most appropriate antecedent and consequent parameters for training the ANFIS. Five hundred iterations were used to evaluate the model’s performance using the fitness function (RMSE), as shown in Figure 3. Initial GA learning parameters were set to 0.2, 0.4, 0.4 (Table 4), and 0.2 for crossover percentage, mutation percentage, gamma, and mutation rate, respectively. In order to generate the final GANFIS, the stopping criteria or RMSE is employed.

<table>
<thead>
<tr>
<th>Performance index</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0.5841</td>
<td>0.6122</td>
<td>0.4623</td>
<td>0.2762</td>
<td>0.5545</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0348</td>
<td>0.0345</td>
<td>0.0366</td>
<td>0.1844</td>
<td>0.0379</td>
</tr>
</tbody>
</table>

Table 4. GANFIS validation with varying gamma values

Concerning the ANN, the artificial network is composed of four input neurons, six hidden neurons, and one output neuron as present in Table 5. A trial-and-error procedure is utilized to identify the initial values of the model’s parameters. Validation was conducted to evaluate the efficacy of the models for predicting the strength of soils based on various metrics, including RMSE and R. This challenge used both training and validation datasets. While the training dataset was used to verify the models’ fit with the data, the validation dataset was used to validate the prediction ability of the models.

FIGURE 3. RMSE analysis
TABLE 5. ANN validation with varying number of hidden layers

<table>
<thead>
<tr>
<th>Performance index</th>
<th>Number of hidden layers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>R</td>
<td>0.4100</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0535</td>
</tr>
</tbody>
</table>

The results of the validation and comparison of the models using RMSE, R, and R2 criteria are depicted in Figures 4, 5, 6, and 7. According to the validation results using RMSE criteria (Figs. 4 and 5), the RMSE values of the models range from 0.014 to 0.019 for the training dataset, which is smaller than the standard deviation of the training dataset used for the respective models as present in Table 3, indicating that all models have good performance; however, the GANFIS has the highest RMSE value compared with ANN, indicating that the GANFIS has the best goodness of fit with the data.

Figure 4 shows that the R and R² values of the two models for the training dataset range from 0.9972 to 0.9981 and 0.9943 to 0.9961, respectively, indicating that all two models have a decent match with the data; however, the GANFIS (R=0.9981, R²=0.9961) has the best fit, followed by the ANN (R=0.9972, R²=0.9943). In this study, the GANFIS (R=0.9984, R²=0.9968) outperforms the ANN (R=0.9983, R²=0.9967) in predicting the strength of SBA-SC based on the validation dataset depicted in Figure 7.

In the same way, the RMSE values for GANFIS and ANN are 0.026 and 0.022, respectively, which are smaller than the standard deviation of the testing dataset used for each model. The results of this study show that these models do a good predicting the strength of SBA-SC, but the GANFIS model performs better than the ANNs models.
DISCUSSION

Auditing and designing geotechnical buildings and constructions require evaluating SBA-SC shear strength. Shear strength testing is time-consuming and requires expensive laboratory equipment. Consequently, predicting shear strength using advanced artificial intelligence approaches is a useful solution for rapidly determining and economically efficient experimentation. Few researchers have utilized artificial intelligence approaches to forecast the features of soft soil. Moreover, the capacity to estimate the shear strength of SBA-SC using these techniques is still restricted, necessitating the development of more sophisticated methods with enhanced predictive capability. In this study, two advanced artificial intelligence methods, GANFIS and ANN, were employed to improve the shear strength prediction of SBA-SC.

Based on the examination of model validation findings, it can be seen that, of the two models, the GANFIS has the acceptable capability for predicting the strength of SBA-SC. In contrast, the ANN performs somewhat less well in this study. The GANFIS, however, has the highest performance, followed by the ANN. The reasonableness of the generated results can be inferred from the GANFIS usage of GA optimization techniques, which can aid in reducing the RMSE of prediction.

Despite the fact that artificial intelligence techniques such as GANFIS and ANN are advanced solutions for prediction issues, their effectiveness is highly dependent on the quality of the input data. In geotechnical problems, the use of variables determined by many experiments on multiple samples of the same soils might result in biased results, which can impact the effectiveness of the employed models. In this investigation, these two models demonstrated adequate predictive capacity; nevertheless, their performance might be enhanced by supplying more data to make the models more regressive and by employing over-sampling or under-sampling techniques to address imbalanced data sets. In addition, the usage of different input combinations may result in varied model predictions, which must be considered in future research.

In reality, soil is an extremely complex substance whose qualities are difficult to anticipate. Nevertheless, laboratory evaluation of their qualities is not always correct due to
numerous influencing factors (such as testing settings, equipment, tester expertise, etc.). Advanced artificial intelligence models predicted the shear strength of SBA-SC in this work with an average error rate of 2.6%, which is acceptable for geotechnical challenges. Consequently, these artificial intelligence algorithms may also be utilised to predict additional SBA-SC features.

CONCLUSION

This study examined and compared the effectiveness of the two artificial intelligence systems, GANFIS and ANN, in predicting the strength of SBA-SC. A palm oil plantation in Sarang Buaya, Semerah, Malaysia, provided the soft clay soil data. GANFIS are relatively new fuzzy inference systems that have been studied infrequently for forecasting the strength of SBA-SC. In contrast, ANN is a popular and effective artificial intelligence in soil strength prediction. The result indicates that the prediction quality of SBA-SC strength is highly influenced by the approach employed. GANFIS has the highest prediction performance of the two models; hence, we conclude that GANFIS is a valuable tool for predicting the strength of SBA-SC.

The primary benefit of GANFIS is that the model was developed and then automatically optimized by meta-heuristic optimization methods, GA. Therefore, they may ensure that the parameters of the model inference procedures for predicting SBA-SC strength are optimum. The GANFIS model performed more effectively than the ANN model. This is due to GA’s robust global search capability and rapid convergence (Rufaizal Che Mamat, Ramli, Yazid, et al. 2022). Consequently, the GA model determined the optimal parameters more efficiently than the ANN model.

The limitation of this study is the investigation of just meta-heuristic optimization algorithms (GA). Fuzzy clustering and newer machine learning algorithms, including the Bagging framework, ensembles, and advanced decision trees, should be considered for optimizing the ANFIS. As input variables such as LL, PI, W, and CC are accessible, the results of this study are useful for geotechnical engineers in predicting the strength of SBA-SC for auditing geotechnical structures and constructions in practice. It will also aid in reducing building costs by lowering the price of laboratory experiments.

ACKNOWLEDGEMENT

The authors are appreciative of the Centre of Green Technology for Sustainable Cities research group at Politeknik Ungku Omar, Malaysia, for their unwavering support of this study.

DECLARATION OF COMPETING INTEREST

None

REFERENCES


