Multiscale and Multidisciplinary Modeling, Experiments and Design
https://doi.org/10.1007/s41939-021-00092-8

ORIGINAL PAPER

Smart computing models of California bearing ratio, unconfined compressive strength, and resistance value of activated ash-modified soft clay soil with adaptive neuro-fuzzy inference system and ensemble random forest regression techniques

Mudassir Iqbal[1] · Kennedy C. Onyelowe[2,3] · Fazal E. Jalal[4]

Received: 6 January 2021 / Accepted: 18 March 2021

Abstract

Sugeno or Takagi–Sugeno–Kang (TSK) type fuzzy inference system ANFIS proposed by Jang and ensemble random forest (ERF) regression, an extension of bootstrap aggregation of decision trees, has been employed to forecast the triple targets of strength properties of a hydrated-lime activated rice husk ash stabilized soft clay soil. This was necessitated to deal with the incessant failure being recorded on flexible pavements around the world and the efforts being made to tackle the situation in a more smart and sustainable approach. The independent variables of this model protocol were the HARHA—hydrated-lime-activated rice husk ash, w_L—liquid limit, w_P—plastic limit, I_P—plasticity index, w_{OMC}—optimum moisture content, A_C—clay activity, $\delta_{\max}$—maximum dry density, while CBR—California bearing ratio, UCS_{28}—unconfined compressive strength at 28 days curing and R—resistance value were estimated and employed as the targets (dependent variables). The natural clayey expansive soil used for this research work was investigated through preliminary experiments and classified as A-7–6 group according to AASHTO. It exhibits a very high plasticity index with high clay content, hence needed modification to be rendered as a foundation material. The soil was treated with varying percentages of HARHA, and the effect on the consistency limits, compaction, CBR, UCS, and R-value was studied. These observed values gave rise to 61 datasets. The observed datasets were deployed on the learning capacity of ANFIS and ERF regression to proposed models for the targets. The outcome of the results showed that both the models presented a close correlation between the parameters used in the model execution. Evaluation of the models was performed using a variety of statistical errors, Kendall and Spearman's rank correlations. The results of ERF regression outclasses ANFIS model yielding a 100% coefficient of determination (R) for the triple targets. The performance evaluation and validation tests show that the coefficient of determination was more than 0.94 with minimized errors. It was concluded that ERF regression and ANFIS learning techniques are viable smart approaches to forecasting engineering problems for a more sustainable design and performance evaluation.

Keywords Smart computing · Ensemble random forest (ERF) regression and adaptive neuro-fuzzy inference system (ANFIS) · Unconfined compressive strength (UCS) · California bearing ratio (CBR) and resistance value (R) · Soft clay soil (SCS) · Hydrated-lime-activated rice husk ash (HARHA) and HARHA-stabilized soft clay soil (HSSCS)

1 Introduction

The adaptive neuro-fuzzy inference system (ANFIS), which was proposed by Jang, is a Sugeno or Takagi–Sugeno–Kang (TSK) type fuzzy inference system, which incorporates the ANN principles (Sugemo 1985; Venkatesh and Bind 2020). It is a hybrid evolutionary model forecasting system, comprising the algorithms of fuzzy and ANN (Jang 1993). There are two main limitations of this type of modeling: (1) The models formulated by ANFIS are complex in the viewpoint of considered membership functions and "if–then" rules which form the final output takes, (2) they do not adapt and train to a stochastic situation (Mazari and Rodriguez 2016; Panahi et al. 2020). On the other hand, random decision forest or simply random forest (RF) is an ensemble technique of machine learning deployed for classification and

✉ Kennedy C. Onyelowe
kennedychibuzor@kiu.ac.ug; konyelowe@gmail.com

Extended author information available on the last page of the article

regression by constructing multitudes of decision trees at the time of training and outputting the outcome that is the mode or means prediction in the class of the individual trees. It is a flexible algorithmic technique that produces good results most of the time without hyper-parameter tuning (Sharma 2020). To deal with the complexity of geo-construction behavior of problematic soils, binder materials, and the blend of soils and cementing materials are simplified approaches to design (Cabalar et al. 2011). Certain empirical and semi-empirical approaches are based on the available data alone to determine the structure, validity, and applications of the model. The techniques, known as Adaptive Neuro-Fuzzy Inference System (ANFIS) and Random Forest (RF), seem to be suited with success to model complex problems in materials, geotechnical and geo-environmental engineering where the relationship between the model variables is unknown. An ANFIS model brings together the elemental and individual representation of fuzzy logic (FL) system with the learning ability of Artificial Neural Networks (ANNs), while RF builds multiple decision trees and merges them into one, to get a more accurate and reliable prediction. This adds additional randomness to the model while the trees are being grown. With the advancement of recent artificial intelligence (AI) techniques, ANFIS and RF algorithms have been recently deployed in the diverse fields of engineering to model the behavior of systems thereby recording successful results. In geotechnical and geo-environmental systems particularly, ANFIS was applied by Cabalar et al. (2011) in conducting an overview of its applications in geotechnical engineering, by Erdirencelebi and Yalpir (2011) in the prediction of anaerobic digestion effluent quality, and by Jokar and Mirasi (2017) in the modeling of unsaturated soils shear strength. Also, this evolving field of ANFIS has been applied by Karaboga and Kaya (2019) as a comprehensive review; Mohammed et al. (2020) to predict shallow foundation settlement quantification; Panahi et al. (2020) to predict spatial landslide susceptibility by applying various metaheuristic algorithms; Venkatesh and Bind (2020) to model the shear strength characteristics of the soil. It is observed that no recent works have been carried out with ANFIS on the mixture experiments of HARHA based soil stabilization. The above exercise which deployed the use of ANFIS showed relatively higher correlations that can be useful in the design and monitoring of those systems. However, ANN and FL were also explored in their functionalities to predict various systems in geo-environmental engineering applications. These included; prediction of UCS of the treated expansive soil by backpropagation algorithms of ANN (Salahudeen et al. 2020), modelling of swelling potential of quicklime-activated rice husk ash treated soft soil by FL method (Alaneme et al. 2020a), modeling volume change of hydrated lime-activated rice husk ash-modified soft soil by ANN method (Alaneme et al. 2020b), comparative

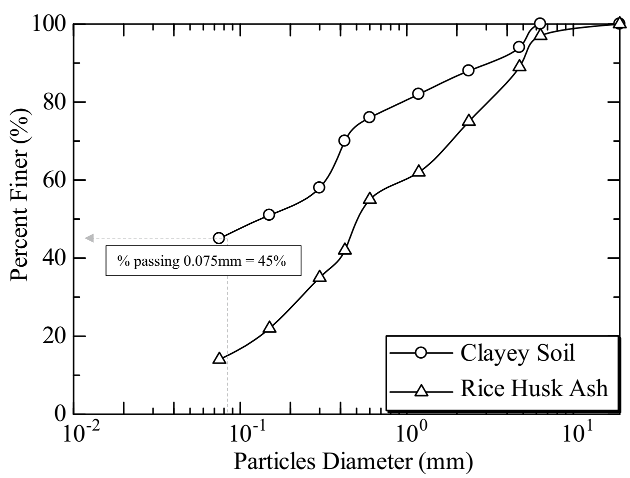

Fig. 1 Particle size distribution curve of clayey soil and rice husk ash

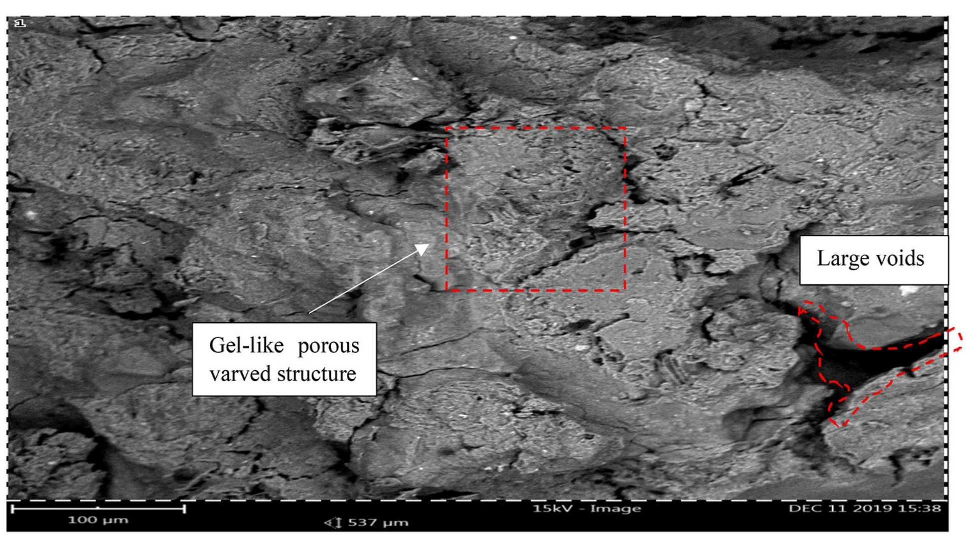

Fig. 2 Morphology of silica-rich RHA

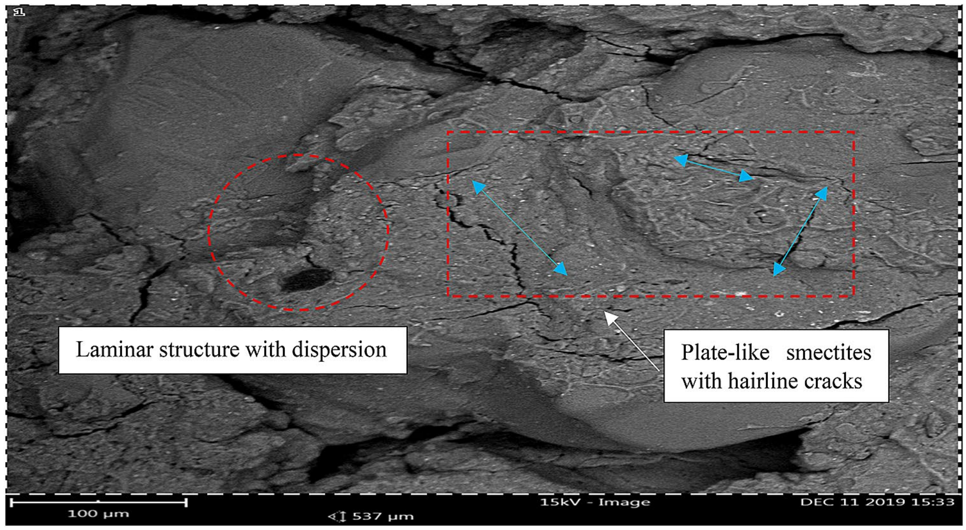

Fig. 3 Morphology of expansive soil

modeling of strength properties of hydrated lime-activated rice husk ash-modified soft soil by ANN and FL methods (Alaneme et al. 2021), the prediction of UCS and CBR of treated sulfate silty sand with applications to deep soil mixing using ANN (Ghorbani and Hasanzadehshoiili 2018) and the prediction of pavement roughness using ANN (Mazari and Rodriguez 2016). These show that ANN and FL in their individual applications have been pronounced in the field of soil properties modification for sustainable construction while however, the use of ANFIS in this field has not been well utilized to simulate the behavior of soil in blends of cementitious materials. Other fields in engineering have also found the application of ANFIS very useful. For instance, in the field of electronic engineering, mobile learning was predicted with the use of ANFIS (Al-Hmouz et al. 2012) and in the field of agriculture, a model for the prediction of moisture diffusivity and specific energy consumption of potato, garlic, and cantaloupe drying under convective hot air dryer was developed with great success. ANFIS was applied to simulate the response of the model footing subjected to vertical centered and eccentric loads. The results of their study encourage the use of ANFIS in supporting the optimization of model testing program. Gokceoglu et al. (2004) constructed a Neuro-Fuzzy model for the prediction of deformation modulus of rock masses. Kayadelen et al. (2009) studied the friction angle using soft computing methods, as a primary requirement for more reliable design of geotechnical structures (i.e., foundations, roads, embankments, and excavation, slopes, and liner systems for the solid waste). They also developed two ANFIS models, which were found to be able to learn the complex relationship between the basic soil properties (e.g., percentage of fine grains,

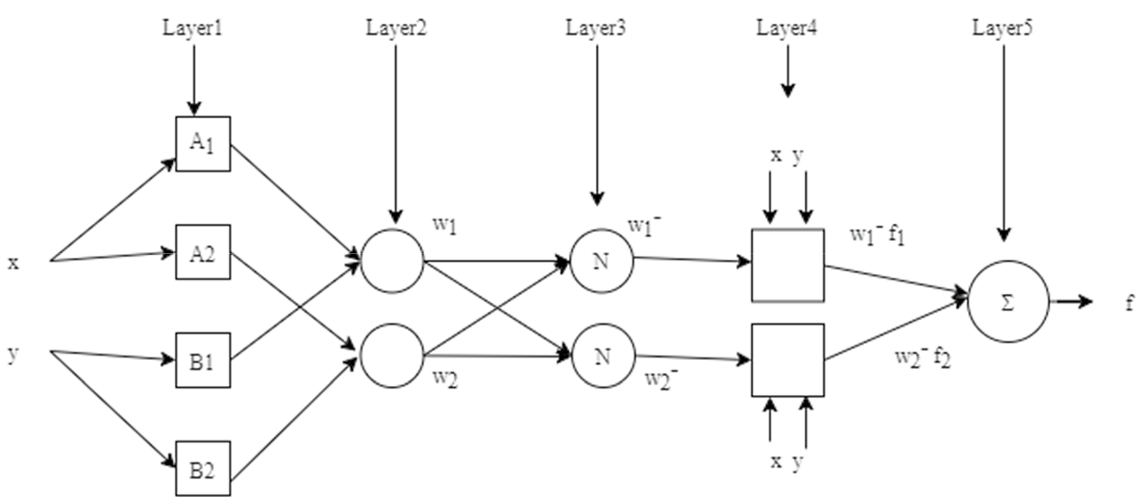

Fig. 4 Typical architecture of two-input ANFIS model

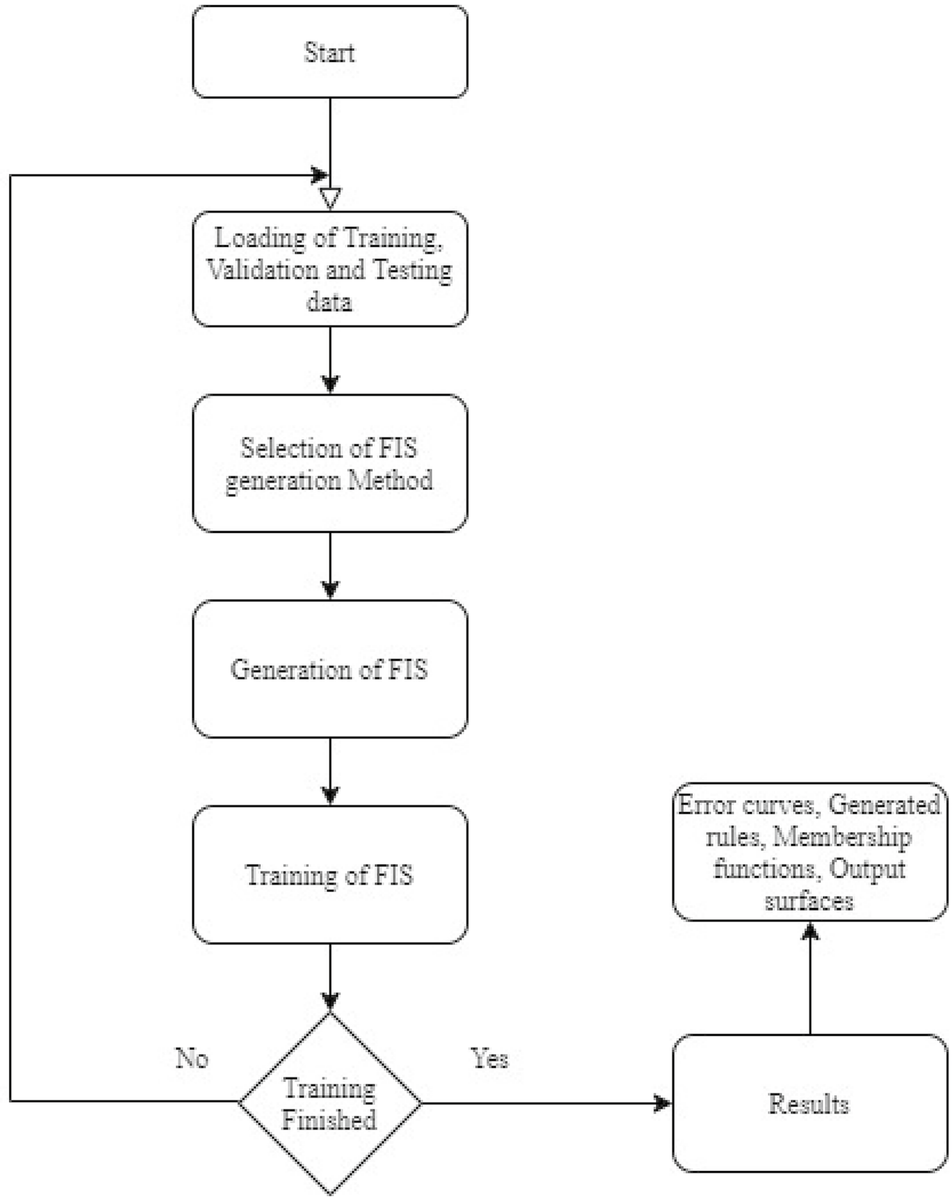

Fig. 5 Flow diagram of ANFIS modelling

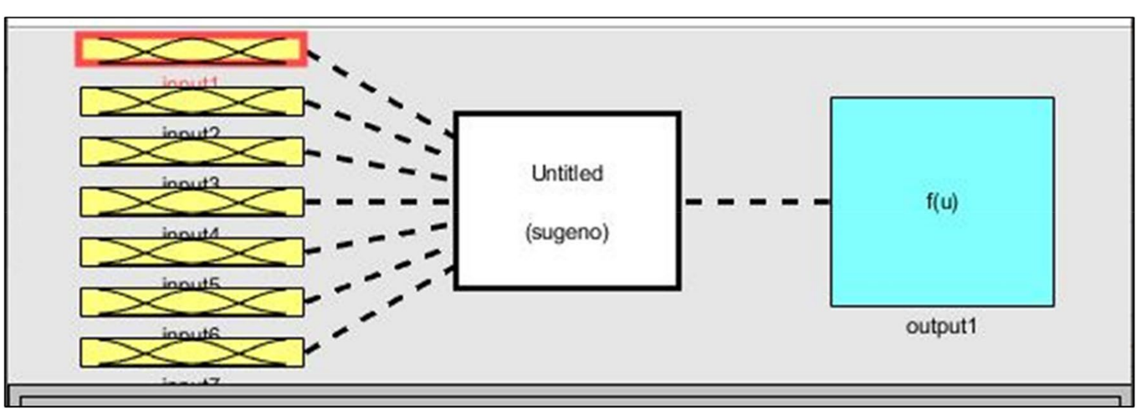

Fig. 6 Sugeno-FIS model

Table 1 Setting parameters for the ANFIS models

Parameter	Setting
Sampling	
Training record	85
Validation/testing	36
General	
Type	Sugeno
And method	Prod
Imp method	Prod
Or method	Probor
Agg method	Sum
Defuzzification method	Whatever
FIS properties	
FIS type	Sub-clustering
Training FIS method	Hybrid
Error tolerance	0
Epochs	50

liquid limit, and bulk density) and frictional angle. The results obtained from the developed ANFIS models developed showed satisfactory agreement with the experimental results. Rangel et al. (2005) presented an alternative approach to evaluating tunnel stability during construction using a Neuro-Fuzzy system. Rangel et al. (2005) verified the validity and efficiency of the model developed by considering two examples of actual tunnels. Conversely, RF has also being applied in the field of geotechnical engineering with overwhelming success. According to Puri et al. (2018), the relationships between in-situ density using SPT N-value, compression index using liquid limit and void ratio, and cohesion and frictional angle using SPT N-value also were developed. Because of the linearity of the model parameters relationship, mean absolute error was used to evaluate the accuracy of the developed model. The results from this work showed that the predicted and measured parameters are in close correlation with a coefficient of determination of about 0.988 with very minimal error. Also, the undrained shear strength (USS) of soil was forecasted by the random forest technique by Pham et al. (2020) with success. In this study, multiple predictors were studied experimentally on 127 soil samples from the USS model was developed. The performance and accuracy of the model showed close correlation between predicted and measured parameters with nonlinearity error, RMSE of 0.48, and coefficient of determination of 0.87. The results of both ANFIS and RF learning techniques from the literature showed a better approximation than the other widely used techniques. None of the above-mentioned pieces of literature was able to forecast models of CBR, UCS, and R of an expansive soil treated with activated ash for a more sustainable and green earthwork construction and

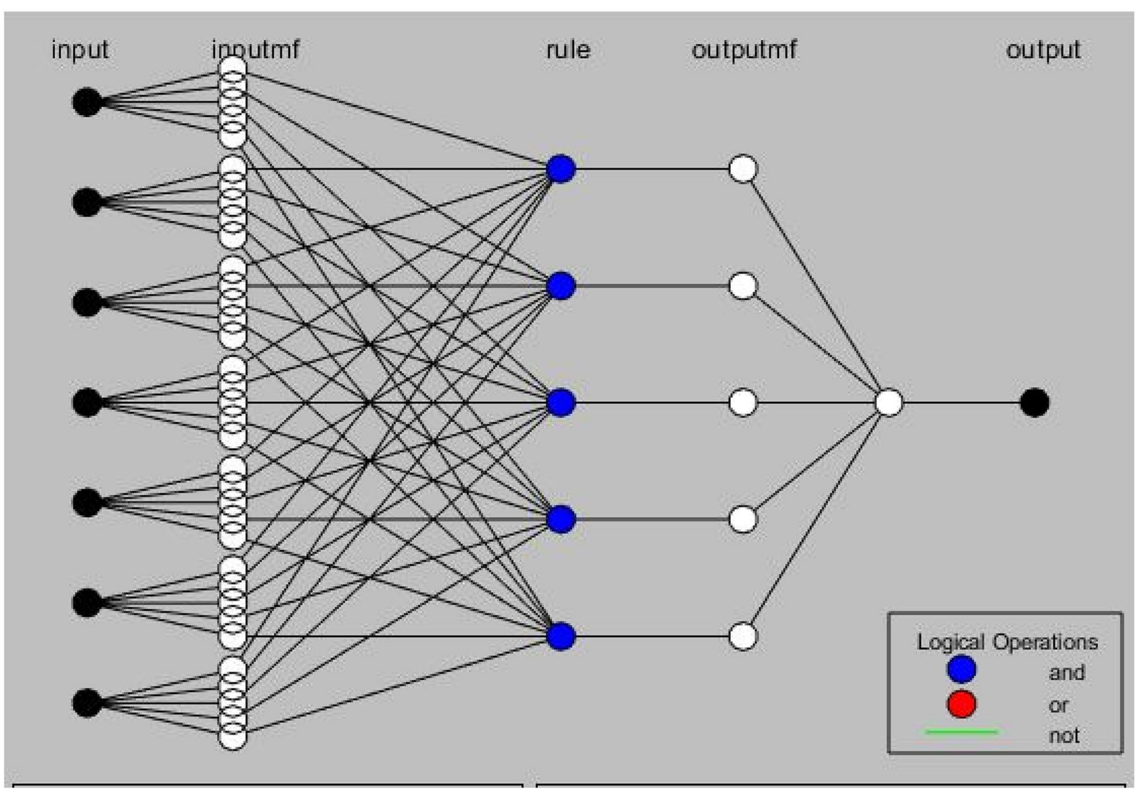

Fig. 7 Architecture of the proposed model

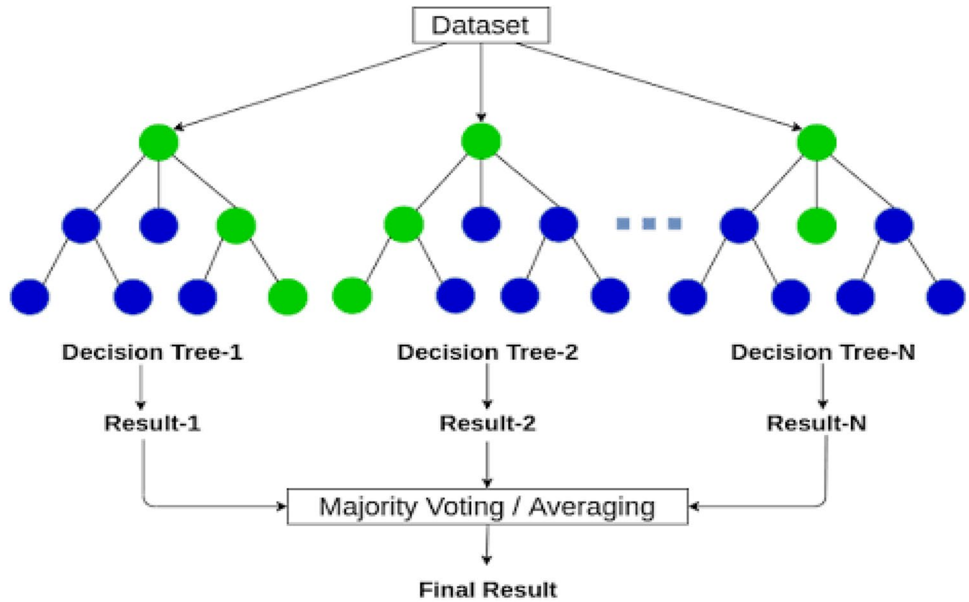

Fig. 8 Random forest (RF) method flowchart (Puri et al. 2018)

infrastructural performance evaluation, hence the focus of this research work is on the smart model's development for a robust geotechnics design, construction and performance monitoring employing the learning abilities of adaptive neuro-fuzzy inference system (ANFIS) and RF. To achieve this, HARHA—hydrated-lime-activated rice husk ash, w_L—liquid limit, w_P—plastic limit, I_P—plasticity index, w_{OMC}—optimum moisture content, A_C—clay activity, and δ_{max}—maximum drydensity were estimated and employed as predictors (independent variable) while, CBR-California bearing ratio, UCS_{28}-unconfined compressive strength at 28 days curing and R-resistance value were estimated and employed as the targets (dependent variables).

2 Materials and methods

2.1 Materials

Soft expansive clay soil with the following properties; percentage passing 0.075 mm size sieve, 45%, natural moisture content, 14%, liquid limit, 66%, plastic limit, 21%, plasticity index, 45%, swelling potential, 23%, AASHTO classification, A-7–6, maximum dry density, 1.25% at an optimum moisture content of 16% with a high degree of expansion was used in this research work. The above characteristics show that the soil is expansive, weak, highly plastic, and requires modification to be suitable as a foundation material. Hence, rice husk was collected from rice mills, and burnt to derive RHA. Meanwhile, to further enhance the

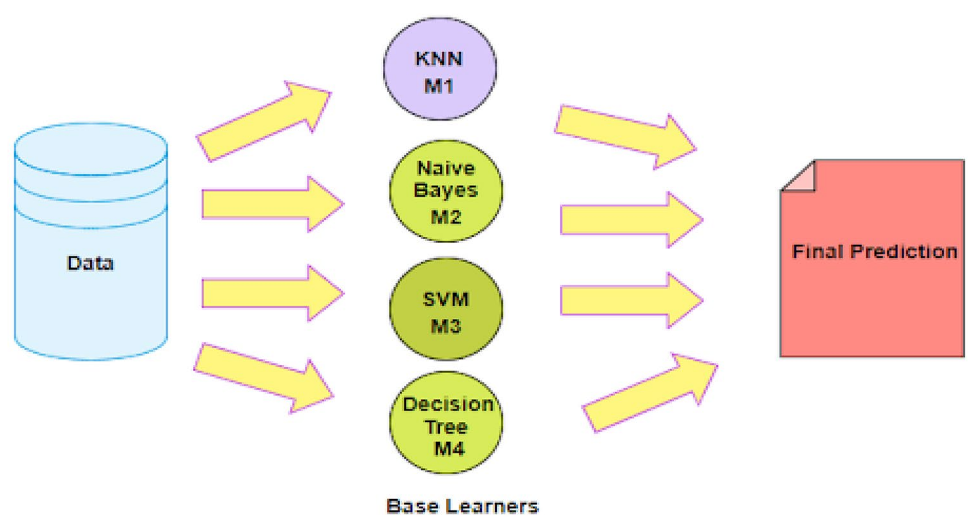

Fig. 9 Ensemble learning method (Pham et al. 2020)

pozzolanic properties of RHA, hydrated-lime was used to achieve hydrated-lime-activated rice husk ash (HARHA). The HARHA was utilized in varying percentages to modify the expansive soil and tests were conducted to produce values for the model. The size distribution and morphology of the soil and the rice husk ash are presented in Figs. 1, 2, and 3, respectively.

It can be seen in Fig. 2 that RHA, exhibits gel-like porous-valve structure at a magnification level of 100 µm. Additionally, the presence of bigger voids could be observed in the SEM micrograph, which illustrates the light-weight structure and porous structure of the agricultural waste. Similarly, the SEM image in Fig. 3 depicts a laminar structure with dispersive, larger, and thinner clay platelets. The smectites are seen to conform plate-like structures with the presence of thin hairline cracks. Additionally, the aggregates are mostly arranged in a face-to-face contact style in Fig. 3.

Table 2 Setting parameters for ERF regression

Parameter	Setting
Sampling	
Training record	121
Validation/testing	121
General	
Type	Regression
Sampling type	Automatic
Number of folds	5
Number of trees	100
Criterion	Least square
Maximum depth	10

2.2 Methods

ANFIS learning algorithm and ensemble RF regression was employed in the modeling exercise with seven predictors; HARHA—hydrated-lime-activated rice husk ash, w_L—liquid limit, w_P—plastic limit, I_P—plasticity index, w_{OMC}—optimum moisture content, A_C-clay activity, $\delta_{\max}$—maximum dry density and three targets,i.e., CBR-California bearing ratio, UCS_{28}—unconfined compressive strength at 28 days curing and R—resistance value. Various experiments on the treated expansive soil like the Atterberg limits, compaction, CBR, UCS, and R-value tests were conducted to produce 61 datasets.

2.2.1 Overview of ANFIS

Figure 4 represents the typical architecture of ANFIS working on two "if–then" rules. It is a type of ANN that is based on Takagi–Sugeno fuzzy inference system (FIS). Hence, a combination of FIS with ANN reduces the limitations of ANN approach. Each node in the first layer (layer 1) has a membership function (MF), which carries the degree of satisfaction of inputs according to the quantifier (low, average, high). The nodes in the second layer (layer 2) produce firing strengths, which are normalized in the consequent layer. Finally, the strength of all incoming signals is calculated as the output (layer 5). The current study employed a hybrid model for training FIS. FIS was generated using a subtractive-clustering method due to the large variance among the datspoints. In addition, Gaussian type of membership function was used. The step-wise procedure of the ANFIS modeling is presented in Fig. 5. This figure presents the step-by-step flow of prediction activities in the intelligent protocol for an ANFIS operation.

Table 3 Statistical functions for input and output parameters

Parameters	Minimum	Maximum	Mean	Median	Standard deviation	Skewness	Kurtosis
Input parameters							
HARHA	0	12	6	6	3.51	0	−1.2
w_L	27	66	47.99	49	11.5	−0.12	−1.25
w_p	12.8	21	17.2	17.7	2.41	−0.06	−1.24
I_p	14	45	30.8	31	9.14	−0.144	−1.24
W_{OMC}	16	19	18	18.2	0.76	−0.94	0.24
A_c	0.6	2	1.34	1.4	0.39	−0.2	−1.17
$\delta_{\max}$	1.25	1.99	1.68	1.69	0.24	−0.16	−1.4
Output parameter							
CBR	8	44.6	24	22.8	11.74	0.29	−1.17
UCS	125	232	172.8	172	31.65	0.26	−1.03
R	11.7	27	20.5	20.9	4.48	−0.43	−0.79

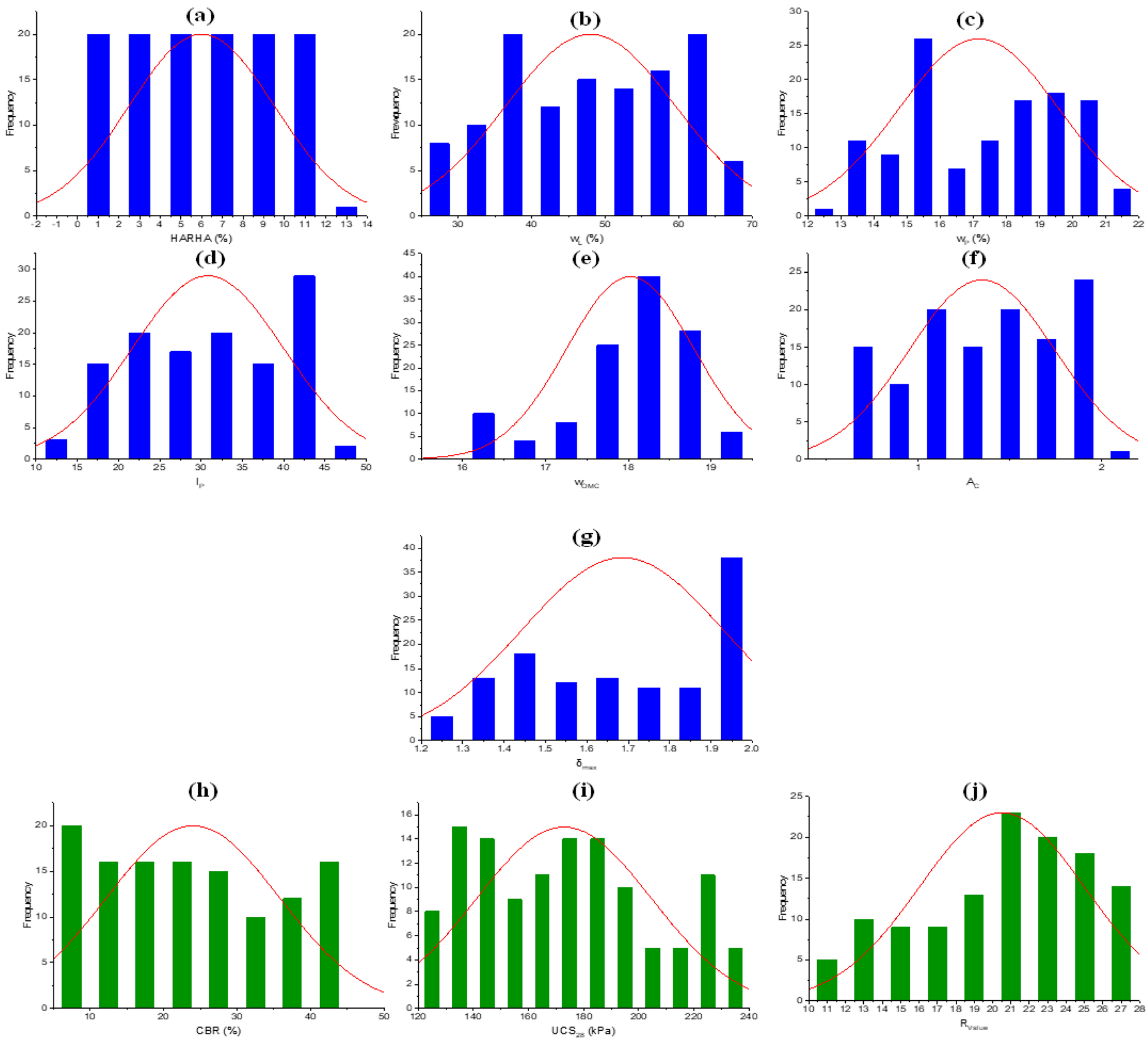

Fig. 10 Distribution histogram for input (in blue) and output (in green) parameters

Table 4 Pearson correlation matrix for inputs and out parameter (CBR)

	HARHA	w_L	w_p	I_p	W_{OMC}	A_c	δ_{max}	CBR
HARHA	1							
w_L	−0.99724	1						
w_p	−0.98926	0.991515	1					
I_p	−0.99652	0.999411	0.986472	1				
W_{OMC}	0.201388	−0.1435	−0.17491	−0.1348	1			
A_c	−0.99388	0.997543	0.984584	0.998142	−0.12039	1		
δ_{max}	0.985771	−0.98176	−0.97696	−0.98026	0.23936	−0.97417	1	
CBR	0.991609	−0.99425	−0.98026	−0.99514	0.097679	−0.9951	0.969326	1

Table 5 Pearson correlation matrix for inputs and out parameter (UCS)

	HARHA	w_L	w_p	l_p	W_{OMC}	A_c	δ_{max}	UCS
HARHA	1							
w_L	−0.99724	1						
w_p	−0.98926	0.991515	1					
l_p	−0.99652	0.999411	0.986472	1				
W_{OMC}	0.201388	−0.1435	−0.17491	−0.1348	1			
A_c	−0.99388	0.997543	0.984584	0.998142	−0.12039	1		
δ_{max}	0.985771	−0.98176	−0.97696	−0.98026	0.23936	−0.97417	1	
UCS	0.990886	−0.99098	−0.97628	−0.99206	0.134931	−0.99283	0.967127	1

Table 6 Pearson correlation matrix for inputs and out parameter (R)

	HARHA	w_L	w_p	l_p	W_{OMC}	A_c	δ_{max}	R
HARHA	1							
w_L	−0.99724	1						
w_p	−0.98926	0.991515	1					
l_p	−0.99652	0.999411	0.986472	1				
W_{OMC}	0.201388	−0.1435	−0.17491	−0.1348	1			
A_c	−0.99388	0.997543	0.984584	0.998142	−0.12039	1		
δ_{max}	0.985771	−0.98176	−0.97696	−0.98026	0.23936	−0.97417	1	
R	0.984407	−0.9721	−0.96953	−0.97003	0.363941	−0.96588	0.972762	1

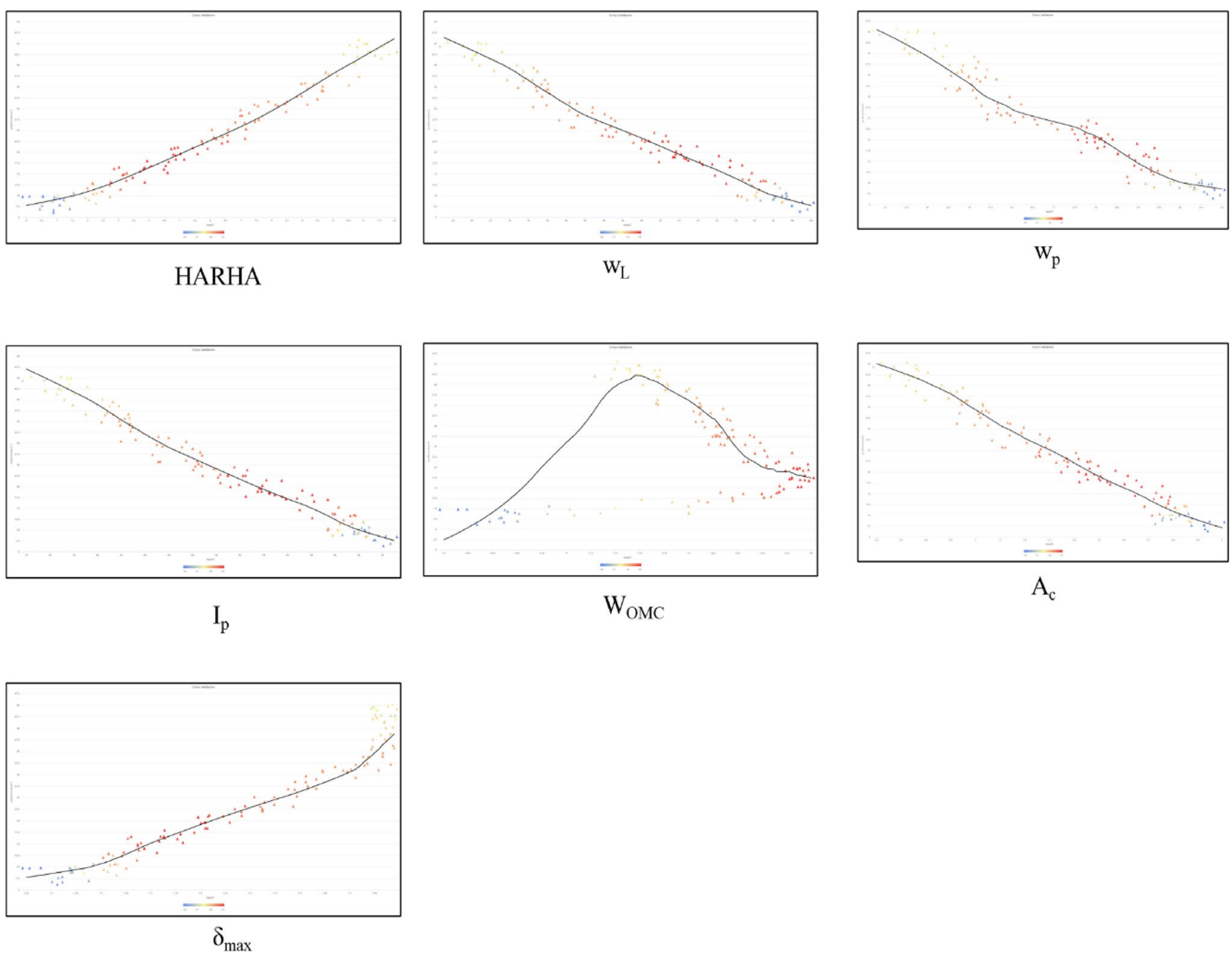

Fig. 11 Parametric study of CBR using ERF regression

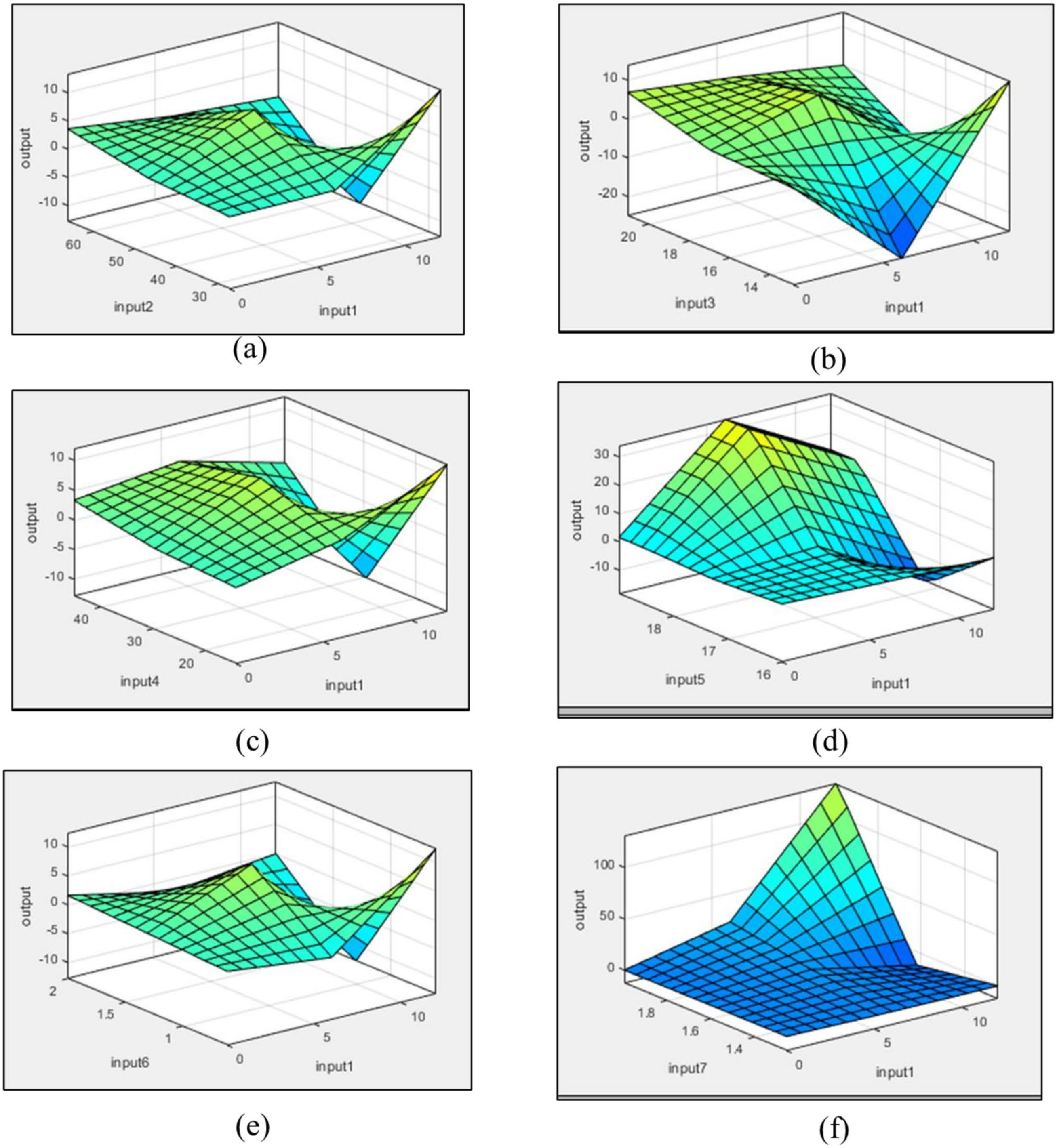

Fig. 12 Comparative analysis of HARHA contribution as compared to **a** w_L, **b** w_p, **c** l_p, **d** W_{OMC}, **e** A_c and **f** δ_{max} for CBR

2.2.2 ANFIS model setting parameters

The model was based on Sugeno-FIS fed with seven independent variables as depicted in Fig. 6. This depicts the computer–human interface that allows the inputs and outputs operations that give rise to the forecasting of the target parameters. The setting parameters of the ANFIS are listed in Table 1. It can be seen that 70% of the total data set was used for training the model, while 30% of the data was equally divided among testing and validation data sets. The analysis was carried via fuzzy logic toolbox of MATLAB R2020b. 50 iterations were used for all three models while training FIS. Figure 7 presents the architecture of the

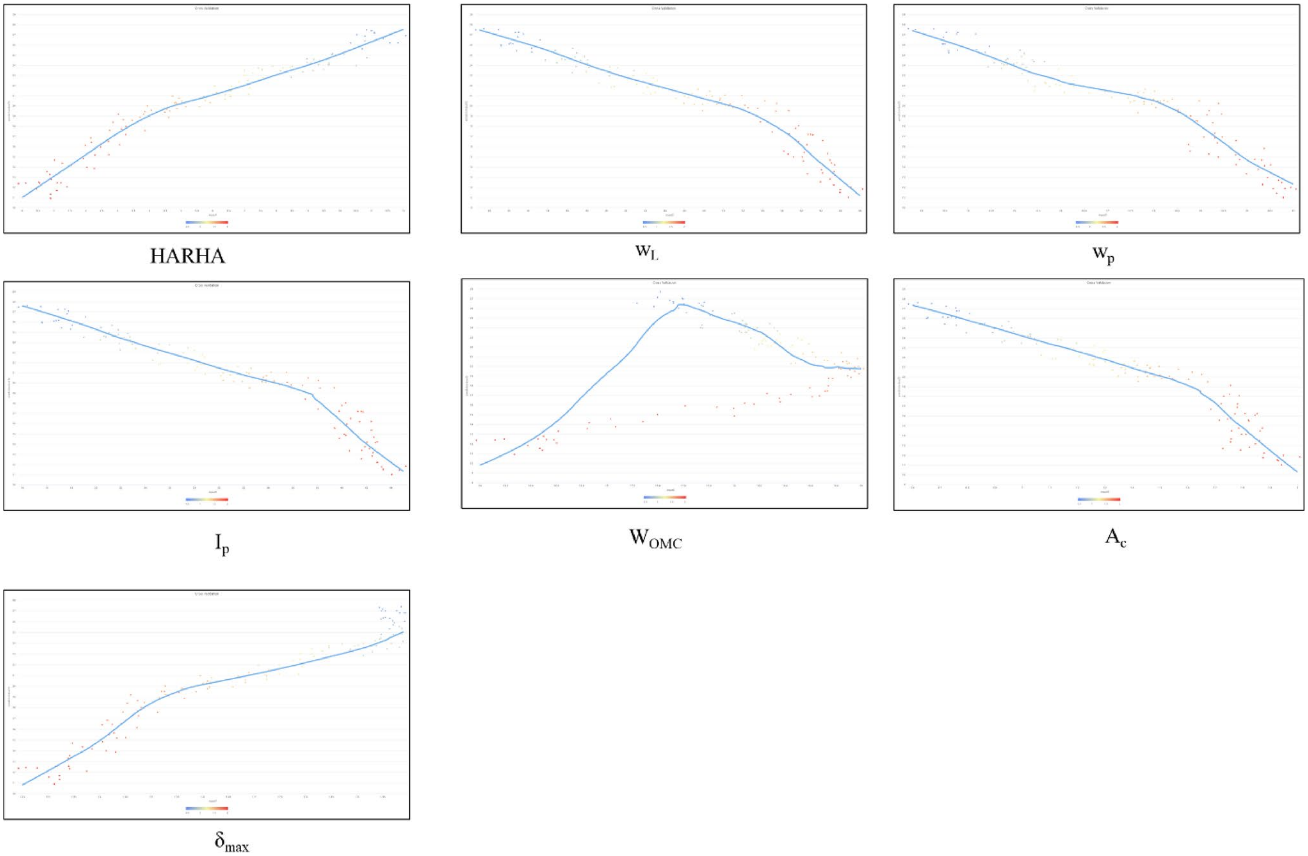

Fig. 13 Parametric study of UCS using ERF regression

proposed model showing the number of membership functions (MFs) in modeling.

2.2.3 Ensemble RF regression overview

Ensemble learning combines the predictions from multiple machine learning algorithms into one to make more perfect forecasting or models than any individual model. The ensemble approach employed in the random forest learning algorithm utilizes the growing tree of the RF into the assembled algorithm to improve the performance and accuracy of the base algorithm (Pham et al. 2020). The RF trees and the ensemble learning method are shown in Figs. 8 and 9. These machine figures show the iterative operations and steps involved in making a decision that best fits the target prediction depending on the trees' performance.

2.2.4 ERF regression setting parameters

Unlike the ANFIS model, a complete data set was used for training and validation of the model using five (5) numbers of folds. A sampling type was kept automatic. Several iterations were carried out varying the number of trees from 100 to 300 with an increment of 50, but no significant improvement in the model was observed; hence, final results were reported on basis of 100 numbers of trees as depicted in Table 2. The process developed for random forest regression is provided in supplementary data.

2.3 Experimental database

The extensive laboratory testing of 121 specimens for each output parameter (CBR, UCS, and R) were conducted with variation in input parameters as presented in Table 3. The distribution histograms were plotted for the input and output parameters, as shown in Fig. 10. A slight or no skewness was observed in both types of parameters used. The essential statistical functions have been listed in Table 3, depicting the satisfying values of skewness and kurtosis.

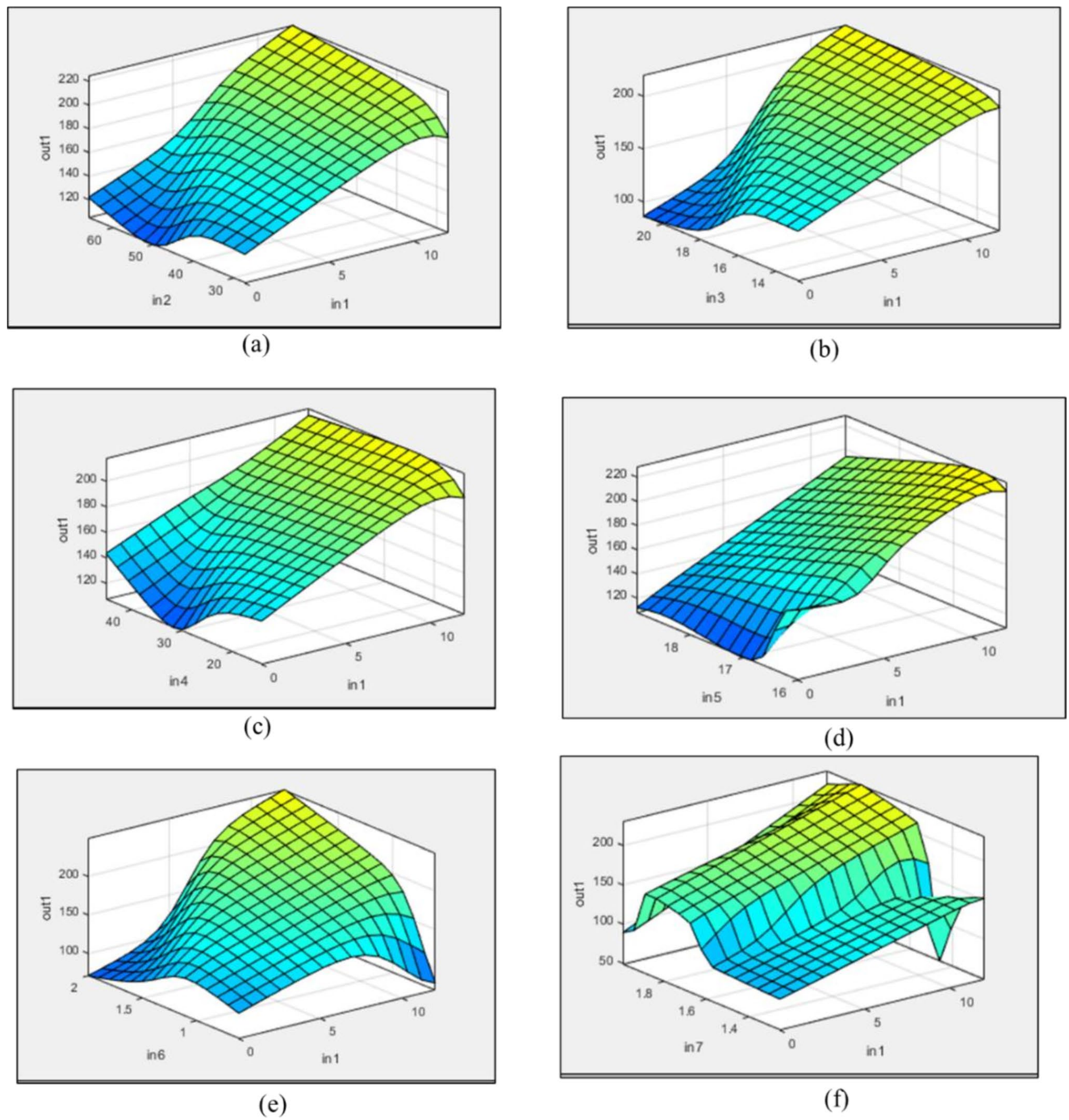

Fig. 14 Comparative analysis of HARHA contribution as compared to **a** w_L, **b** w_p, **c** l_p, **d** W_{OMC}, **e** A_c and **f** δ_{max} for UCS

3 Results and discussion

3.1 Pearson's correlations

According to previous studies, the current research employed Pearson correlation coefficients as presented in Tables 4, 5, and 6 to measure the linear relationship between the input and output parameters (Adler and Parmryd 2010; Benesty et al. 2008; Benesty et al. 2009). The use of HARHA influenced the values of CBR, UCS, and R almost in a similar manner. The CBR value depicted a strong positive linear relationship with the addition of HARHA. CBR seems to unaffected in contrast to OMC. Moreover, the maximum dry density significantly influenced the value of CBR, thus depicting a strong positive relationship. A similar type of trend was observed for the value of UCS and R values as well.

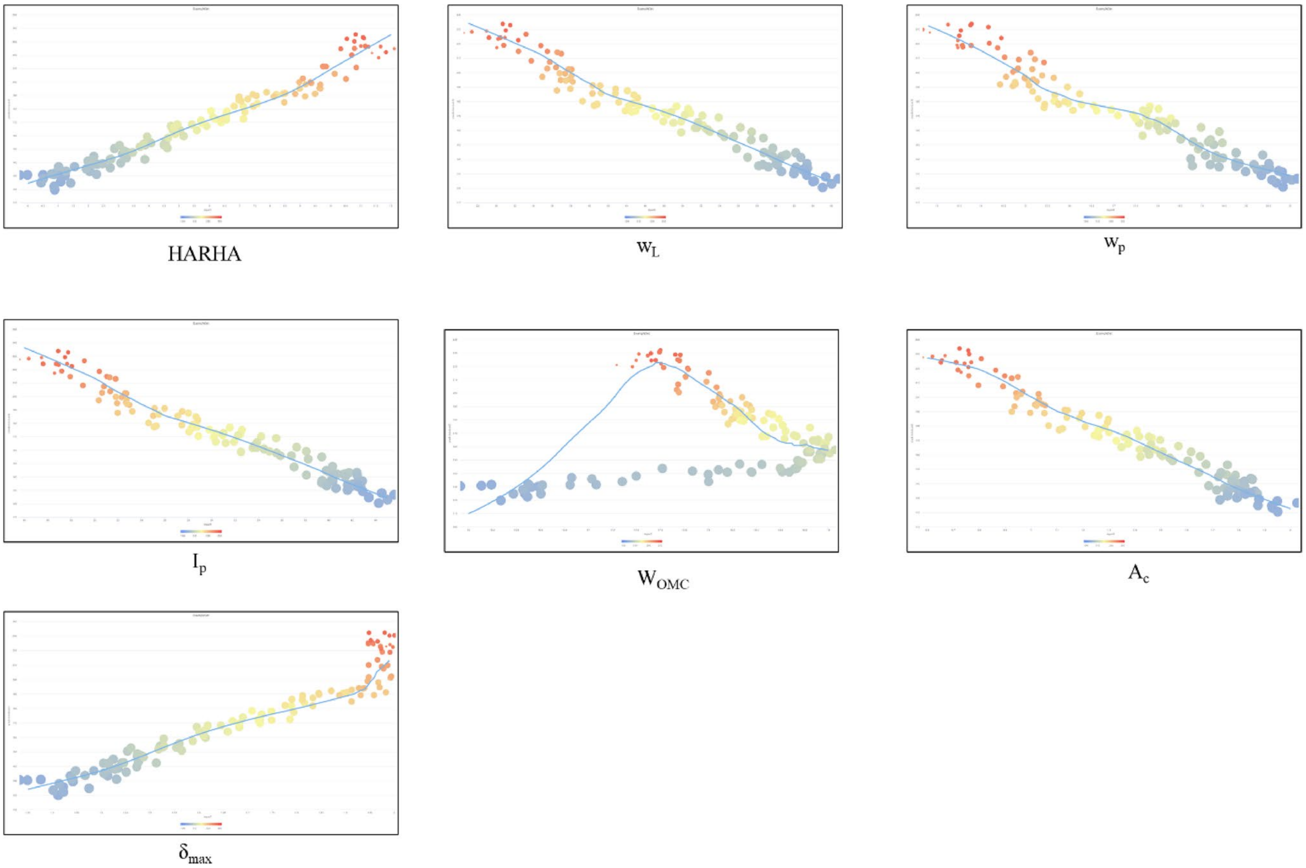

Fig. 15 Parametric study of *R* using ERF regression

3.2 Parametric study

3.2.1 California bearing ratio (CBR)

The individual effect of input parameters in CBR values was investigated from both models. ERF regression was primarily used to see the effect of each input in predicting the CBR results, whereas, the surface results achieved from ANFIS model were used to investigate the contribution of HARHA as compared to other inputs. It is clear from Fig. 11 that the percentage of HARHA linearly increased the CBR value as evidence for the Pearson correlation matrix. OMC (up to 17.5%) initially increased the CBR and then start decreasing (Yadu et al. 2011; Brooks 2009). The maximum dry density also linearly increased the value of CBR. The random tree generated for CBR via ERF regression is provided in supplementary data. A similar pattern of results was obtained from surface generated via ANFIS model for CBR as shown in Fig. 12.

3.2.2 Unconfined compression strength (UCS)

The random trees generated for the UCS model are provided in supplementary data. The results obtained from ERF regression revealed that UCS linearly increased with the increase of HARHA and maximum dry density, whereas OMC initially increased and then declined the UCS value. The plastic index has a negative influence on compression results (Fig. 13). The 3-D plots of HARHA relative to remaining attributes revealed enhanced contribution of HARHA relatively (Fig. 14).

3.2.3 Resistance values (*R*)

Random trees generated for *R* model are provided in supplementary data. A similar pattern of impacting inputs was recorded for *R* values as manifested in Fig. 15. The relative contribution of HARHA reflected more contribution as compared to other attributes (Fig. 16).

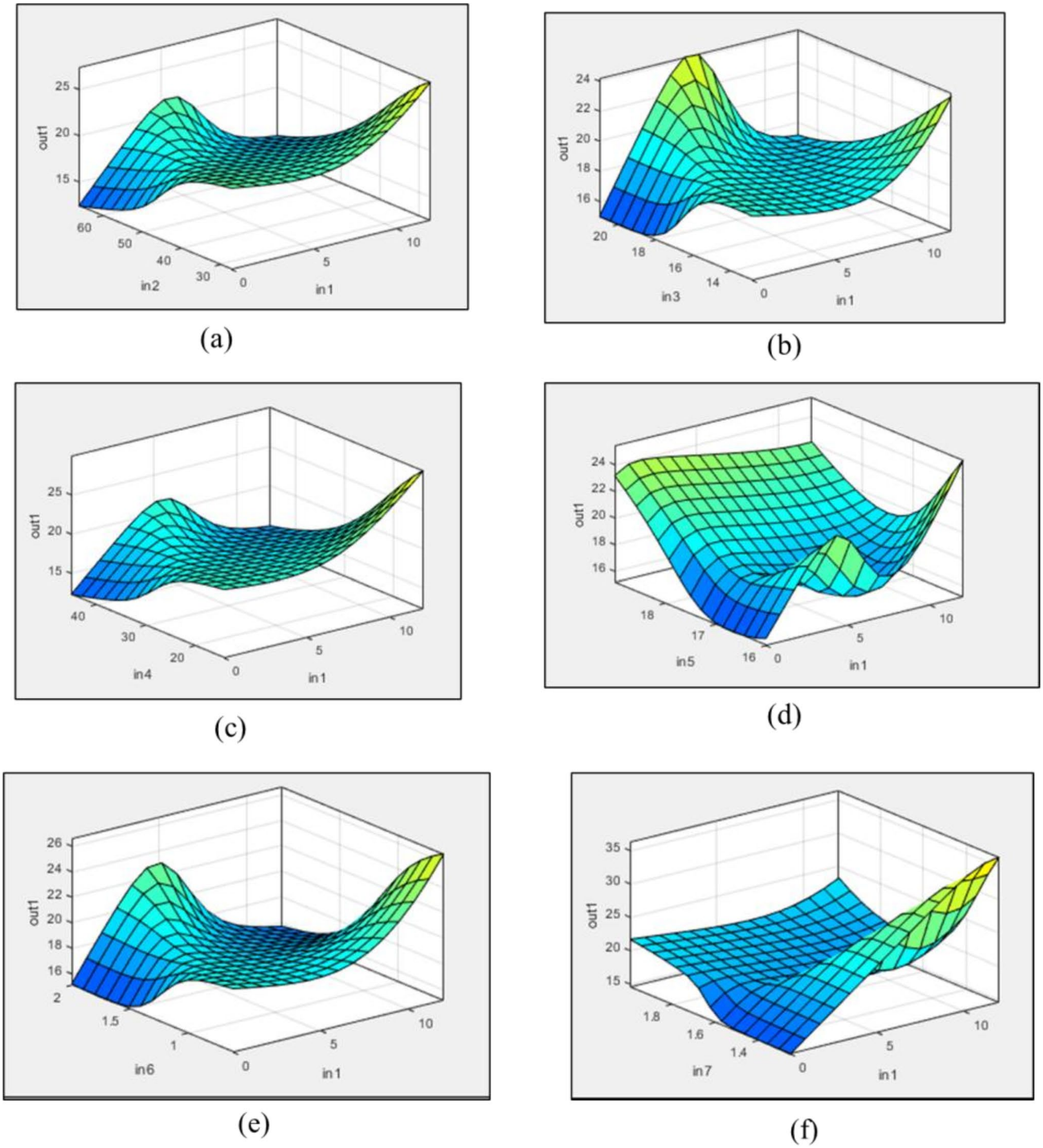

Fig. 16 Comparative analysis of HARHA contribution as compared to **a** w_L, **b** w_p, **c** l_p, **d** W_{OMC}, **e** A_c and **f** δ_{max} for R

3.3 Performance evaluation of models

The statistical results used for the evaluation of the ANFIS models are listed in Table 7. The coefficient of determination (R^2) for all the three models is more significant than 0.94 representing the close agreement of the experimental results to the predicted values. The other functions such as mean absolute error (MAE) (Willmott and Matsuura 2005; Willmott et al. 2009), relative squared error (RSE), root mean squared error (RMSE) (Iqbal et al. 2020), relative root mean square error (RRMSE), performance indicator (ρ)

Table 7 Calculation of statistical parameters for performance evaluation of the proposed ANFIS models

Data set	Statistical evaluation	CBR	UCS	R
Training set	RSQ	0.999922	0.9997	0.999909
	RMSE	5.118456	13.34563	4.558767
	MAE	0.088974	0.362986	0.040364
	RSE	9.82E-07	1.17E-05	3E-06
	RRMSE	0.195371	0.074931	0.219358
Testing set	RSQ	0.989268	0.999600	0.846968
	RMSE	4.020516	12.373348	4.448637
	MAE	0.509516	0.289150	0.695923
	RSE	0.038275	0.001042	0.617788
	RRMSE	0.24285	0.080667	0.232981
Validation set	RSQ	0.942248	0.999500	0.596443
	RMSE	1.197683	1.197683	1.197683
	MAE	0.83023	0.289150	1.352781
	RSE	0.294606	0.004327	17.47836
	RRMSE	0.056882	0.007146	0.058156

Table 8 Objective functions of the proposed ANFIS models

Data set	CBR	UCS	R
Training set	0.999961	0.037468	0.109681
Testing set	0.99462	0.043979	0.121325
Validation set	0.970694	0.003602	0.032814

(Babanajad et al. 2017), and objective function OBF were also used for the model evaluation. The values of RSE are almost zero for the training data sets of all the three models. The values of MAE manifest an error of less than 0.5% of the target values. The mathematical equations of the statistical evaluation functions are presented as Eqs. 1–7.

$$R = \frac{\sum_{i=1}^{n}(e_i - \bar{e}_i)(m_i - \bar{m}_i)}{\sqrt{\sum_{i=1}^{n}(e_i - \bar{e}_i)^2(m_i - \bar{m}_i)^2}}, \quad (1)$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|e_i - m_i|}{n}, \quad (2)$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(e_i - m_i)^2}{n}}, \quad (3)$$

$$\mathrm{RSE} = \frac{\sum_{i=1}^{n}(e_i - m_i)^2}{\sum_{i=1}^{n}(\bar{e} - m_i)^2}, \quad (4)$$

Table 9 Calculation of statistical parameters for performance evaluation of the proposed ERF regression models

Model	Statistical parameter	Training set	CoV (±)
CBR	Root mean squared error	0.265	0.048
	Absolute error	0.197	0.045
	Relative error	0.87%	0.08%
	Normalized absolute error	0.021	0.007
	Root relative squared error	0.024	0.007
	Squared error	0.072	0.024
	Correlation	1	0.000
	Squared correlation	1	0.000
	Spearman rho	1	0.000
	Kendal tau	0.999	4.995
UCS	Root mean squared error	1.001	0.238
	Absolute error	0.681	0.128
	Relative error	0.38%	0.05%
	Normalized absolute error	0.027	0.007
	Root relative squared error	0.033	0.006
	Squared error	1.047	0.487
	Correlation	0.999	0.000
	Squared correlation	0.999	0.000
	Spearman rho	0.998	0.003
	Kendal tau	0.987	0.010
R	Root mean squared error	0.093	0.013
	Absolute error	0.072	0.011
	Relative error	0.37%	0.06%
	Normalized absolute error	0.020	0.003
	Root relative squared error	0.022	0.002
	Squared error	0.009	0.002
	Correlation	1	0.000
	Squared correlation	1	0.000
	Spearman rho	1	0.000
	Kendal tau	0.997	0.002

$$\mathrm{RRMSE} = \frac{1}{|\bar{e}|}\sqrt{\frac{\sum_{i=1}^{n}(e_i - m_i)^2}{n}}, \quad (5)$$

$$\rho = \frac{RRMSE}{(1 + R)}, \quad (6)$$

where e_i and m_i are nth experimental and model TSR(%), respectively; $\bar{e}_i$ and $\bar{m}_i$ denotes the average values of experimental and model TSR(%), respectively; n is the number of samples in the data set.

$$\mathrm{OBF} = \left(\frac{n_T - n_v}{n}\right)\rho_T + 2\left(\frac{n_v}{n}\right)\rho_V, \quad (7)$$

where the subscript *T* and *V* represents the training and validation data, and n is the total number of sample points.

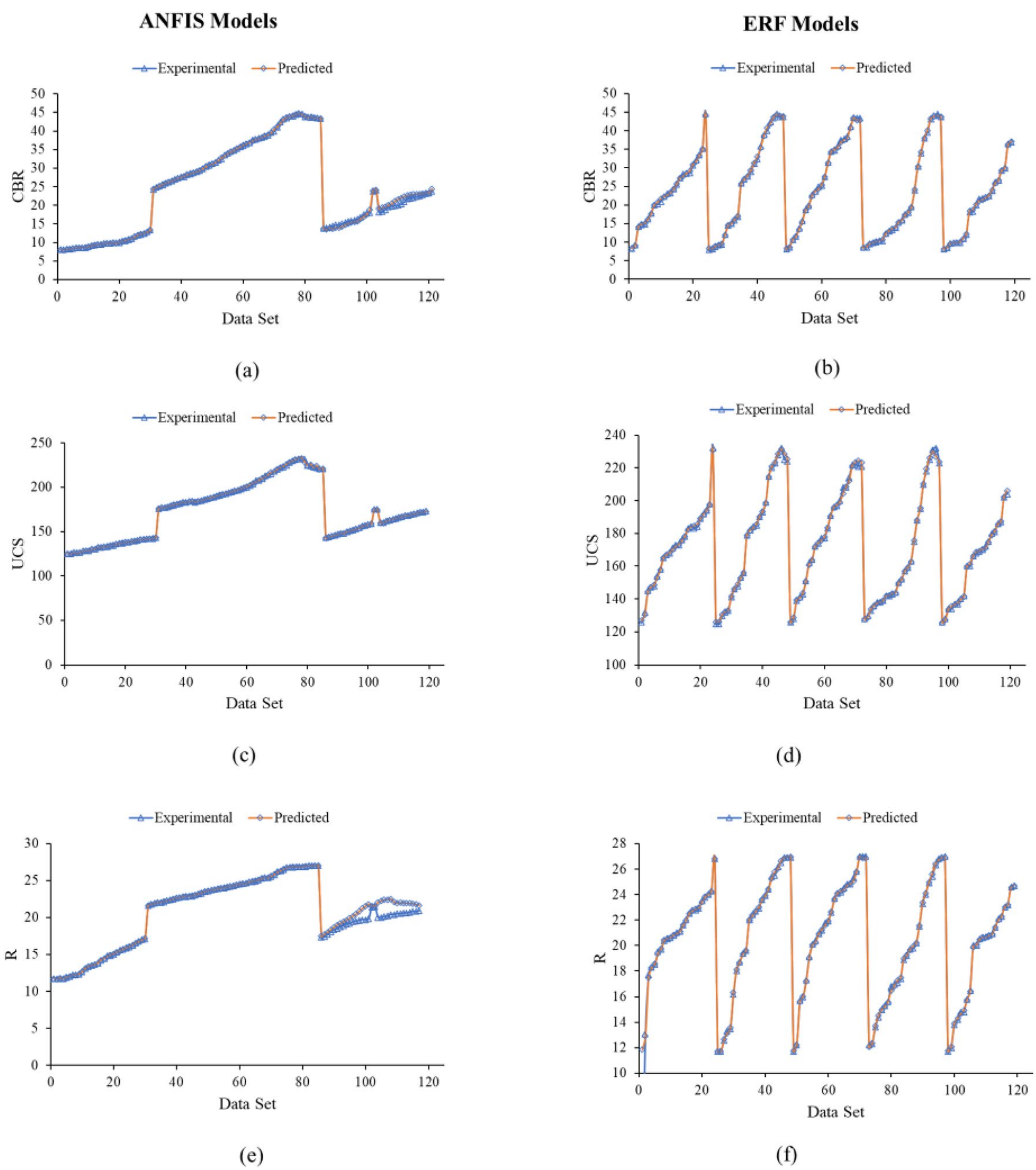

Fig. 17 Comparison of experimental and predicted trends using ANFIS model (**a**), (**c**), (**e**) and EFR regression model (**b**), (**d**), (**f**) for CBR, UCS, and R, respectively

All the statistical error evaluation functions satisfied the performance of the three models. The proximal values of OBF to zero reflect that the models are not overfitted (Table 8).

The ERF regression models for triple targets were evaluated with the above-mentioned statistical evaluations with additional checks of Spearman's rho and Kendall's rho (Table 9). The evaluation of ERF regression models depicted a more accurate prediction of experimental results as compared to ANFIS models. For the CBR and R models, the correlation and squared correlation are exactly 1 while for the UCS model it is 0.999. It can also be seen that the coefficient of variation of the errors is also very small.

3.4 Comparison of experimental and predicted results

Figure 17 represents the tracing of model predictions in comparison to experimental results for ANFIS and ERF regression models. Both types of intelligent models have

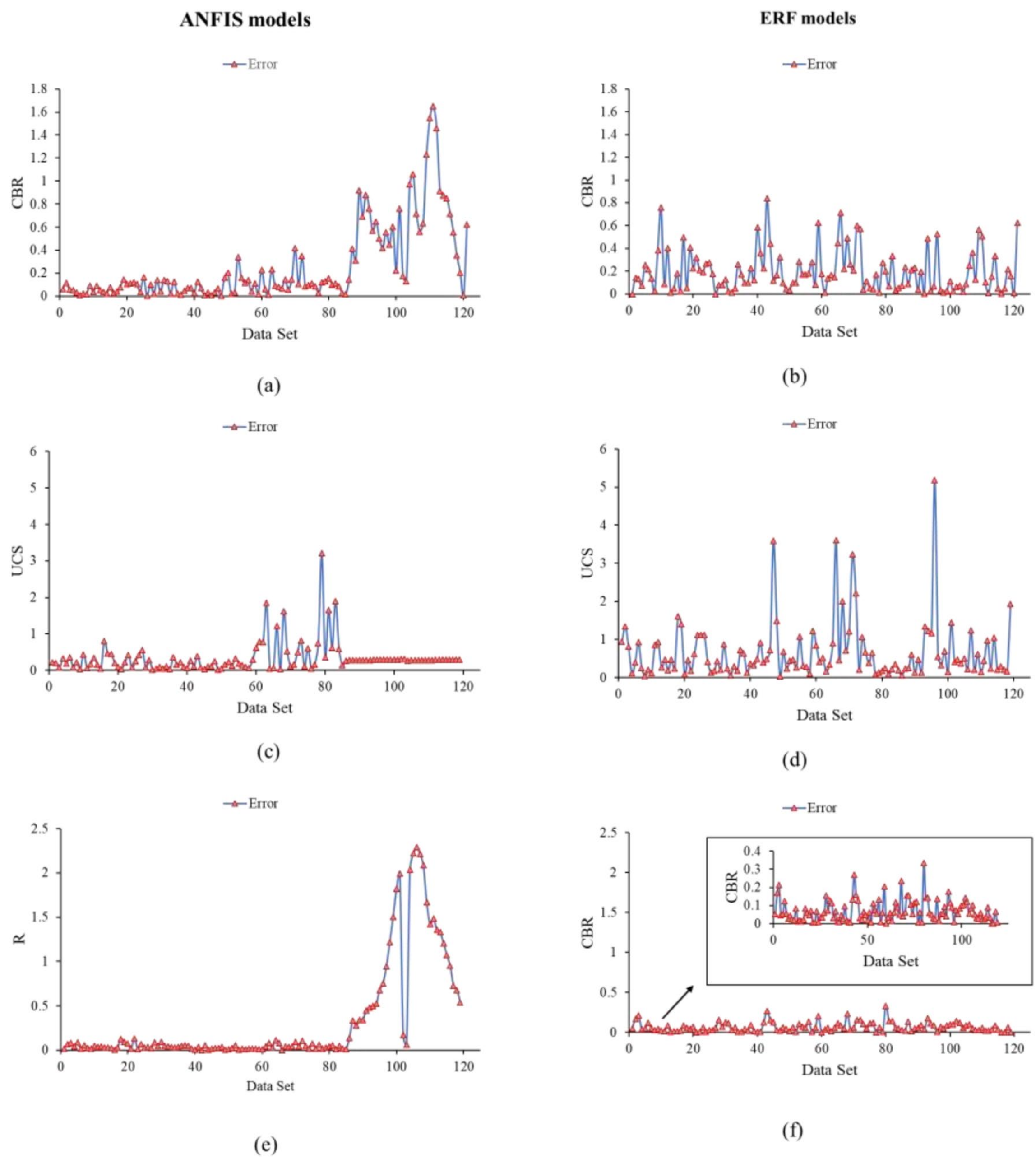

Fig. 18 Error analysis of ANFIS model (**a**), (**c**), (**e**) and EFR regression model (**b**), (**d**), (**f**) for CBR, UCS, and R, respectively

closely followed the experimental trend. The validation set of ANFIS models have relatively more deviation as compared to the training states. The error analysis of two types of models is illustrated in Figure. The maximum absolute error for CBR models is observed to be 1.6% and 0.8% for ANFIS model and ERF regression model, respectively. The maximum error ordinate in case of UCS model is more in ERF regression, whereas, R model has maximum error ordinate in case of ANFIS model. For the training set, error in ANFIS-CBR value reaches a maximum of 0.4%, while in testing and validation sets, it is recorded as 1.6%. Similarly, the error in case of ANFIS-UCS model, a peak of 3.2 was recorded In the case of resistance values, a maximum error ordinate of 2.4 was observed (Fig. 18).

4 Conclusions

The California bearing ratio (CBR), unconfined compressive strength (UCS), and resistance value (R) have been modeled employing the learning techniques of Sugeno or Takagi–Sugeno type fuzzy inference system ANFIS and ensemble random forest (ERF) regression. This was achieved with 61 datasets from experimental results and on seven independent variables and three targets. From the current study, the following have been concluded:

- The expansive soil used in this exercise was classified to belong to A-7–6 group by AASHTO classification. Besides, it exhibits a very high plasticity index with high clay content. The soil has percentage particle passing sieve no 200 as 45%.
- The rice husk ash was activated with hydrated lime to generate HARHA.
- The soil was treated with HARHA in increasing proportions by weight of dry soil and the effect on the other six predictors were monitored and recorded. The effect of the addition of HARHA on the targets was also recorded.
- The observed values from both targets and predictors of the soft clays were tabulated as the datasets and were deployed to the learning abilities of both ERF regression and ANFIS to proposed models.
- The results of the models' performance evaluation and validation show that there was close agreement between the parameters of the exercise, their agreement between the measurement parameters, and the modeled values with a correlation of more than 0.94.
- The evaluation of models using a variety of statical checks reflected that ERF regression model excels ANFIS models of target strength characteristics more precisely.
- It was also concluded that the learning algorithms of ANFIS and ERF regression are viable forecasting techniques for a smart and more sustainable design and performance determination of geotechnical infrastructures.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s41939-021-00092-8.

References

Adler J, Parmryd I (2010) Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytometry A 77(8):733–742

Alaneme GU, Onyelowe KC, Onyia ME, Van Bui D, Mbadike EM, Dimonyeka MU, Attah IC, Ogbonna C, Iro UI, Kumari S, Firoozi AA, Oyagbola I (2020a) Modelling of the swelling potential of soil treated with quicklime-activated rice husk ash using fuzzy logic. Umudike J EngTechnol (UJET) 6(1):1–12. https://doi.org/10.33922/j.ujet_v6i1_1

Alaneme GU, Onyelowe KC, Onyia ME, Van Bui D, Mbadike EM, Ezugwu CN, Dimonyeka MU, Attah IC, Ogbonna C, Abel C, Ikpa CC, Udousoro IM (2020b) Modeling volume change properties of hydrated-lime activated rice husk ash (HARHA) modified soft soil for construction purposes by artificial neural network (ANN). Umudike J EngTechnol (UJET) 6(1):1–12. https://doi.org/10.33922/j.ujet_v6i1_9

Alaneme GU, Onyelowe KC, Onyia ME, Bui Van D, Dimonyeka MU, Nnadi E, Ogbonna C, Odum LO, Aju DE, Abel C, Udousoro IM, Onukwugha E (2021) Comparative modelling of strength properties of hydrated-lime activated rice-husk-ash (HARHA) modified soft soil for pavement construction purposes by artificial neural network (ANN) and fuzzy logic (FL). J Kejuruteraan 33(2) (in press)

Al-Hmouz A, Shen J, Al-Hmouz R, Yan J (2012) Modeling and simulation of an adaptive neuro-fuzzy inference system (ANFIS) for mobile learning. IEEE Trans Learn Technol 5(3):226–237

Babanajad SK, Gandomin AH, Alavi AH (2017) New prediction models for concrete ultimate strength under true-triaxial stress states: an evolutionary approach. AdvEngSoftw 110:55–68

Benesty J, Chen J, Huang Y (2008) On the importance of the Pearson correlation coefficient in noise reduction. IEEE Trans Audio Speech Lang Process 16(4):757–765

Benesty J et al (2009) Pearson correlation coefficient. noise reduction in speech processing. Springer, pp 1–4

Brooks RM (2009) Soil stabilization with fly ash and rice husk ash. Int J Res Rev ApplSci 1(3):208–217

BS 1924 (1900) Methods of tests for stabilized soil. British Standard Institute

Cabalar AF, Cevik A, Gokceoglu C (2011) Some applications of Adaptive Neuro-Fuzzy Inference System (ANFIS) in geotechnical engineering. ComputGeotech 40:14–33. https://doi.org/10.1016/j.compgeo.2011.09.008

Chen FH (1988) Foundations on expansive soils, 2nd edn. Elsevier Services Publications, New York

Erdirencelebi D, Yalpir S (2011) Adaptive network fuzzy inference system modeling for the input selection and prediction of anaerobic digestion effluent quality. Appl Math Model 35:3821–3832. https://doi.org/10.1016/j.apm.2011.02.015

Ghorbani A, Hasanzadehshooiili H (2018) Prediction of UCS and CBR of microsilica-lime stabilized sulfatesilty sand using ANN and EPR models; application to the deep soil mixing. Soils Found 58:34–49. https://doi.org/10.1016/j.sandf.2017.11.002

Gokceoglu C, Yesilnacar E, Sonmez H, Kayabasi AA (2004) Neuro-fuzzy model for modulus of deformation of jointed rock masses. ComputGeotech 31:375–383

Iqbal MF et al (2020) Prediction of mechanical properties of green concrete incorporating waste foundry sand based on gene expression programming. J Hazard Mater 384:121322

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

Jokar MH, Mirasi S (2017) Using adaptive neuro-fuzzy inference system for modeling unsaturated soils shear strength. MethodolAppl Soft Comput. https://doi.org/10.1007/s00500-017-2778-1

Karaboga D, Kaya E (2019) Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey. ArtifIntell Rev 52:2263–2293. https://doi.org/10.1007/s10462-017-9610-2

Kayadelen C, Gunaydin O, Fener M, Demir A, Ozvan A (2009) Modeling of the angle of shearing resistance of soils using soft computing systems. Expert SystAppl 36:11814–11826

Liu Y, Su Y, Namdar A, Zhou G, She Y, Yand Q (2019) Utilization of cementitious material from residual rice husk ash and lime in

stabilization of expansive soil. AdvCivEng 2019:1–17. https://doi.org/10.1155/2019/5205276 (**ID: 5205276**)
Mazari M, Rodriguez DD (2016) Prediction of pavement roughness using a hybrid gene expression programming-neural network technique. J Traffic TranspEng (English Edition) 3:448–455
Mohammad K, Vali RS, Reza AC, Ebrahim T, Abbaspour-Gilandeh Y, Golpour I (2018) ANFIS and ANNs model for prediction of moisture diffusivity and specific energy consumption potato, garlic and cantaloupe drying under convective hot air dryer. Inf Process Agric 5:372–387. https://doi.org/10.1016/j.inpa.2018.05.003
Mohammed M, Sharafati A, Al-Ansari N, Yaseen ZM (2020) Shallow foundation settlement quantification: application of hybridized adaptive neuro-fuzzy inference system model. AdvCivEng 2020:1–14. https://doi.org/10.1155/2020/7381617 (**Article ID 7381617**)
Onyelowe KC, Bui VD (2018) Predicting strength behaviour of stabilized lateritic soil- ash matrix using regression model for hydraulically bound materials purposes. Int J Pavement Res Technol. https://doi.org/10.1016/j.ijprt.2018.08.004
Onyelowe KC, Onwa KC, Uwanuakwa I (2018) Predicting the behaviour of stabilized lateritic soils treated with green crude oil (GCO) by analysis of variance approaches. Int J Min Geo-Eng 52(1):37–42. https://doi.org/10.22059/ijmge.2017.240176.594690
Onyelowe KC, Alaneme G, Igboayaka C, Orji F, Ugwuanyi H, Van Bui D, Van Nguyen M (2019a) Scheffe optimization of swelling, California bearing ratio, compressive strength, and durability potentials of quarry dust stabilized soft clay soil. Mater Sci Energy Technol 2(1):67–77. https://doi.org/10.1016/j.mset.2018.10.005
Onyelowe KC, Alaneme G, Van Bui D, Van Nguyen M, Ezugwu C, Amhadi T, Sosa F, Orji F, Ugorji B (2019b) Generalized review on EVD and constraints simplex method of materials properties optimization for civil engineering. Civil Eng J 5(3):729–749. https://doi.org/10.28991/cej-2019-03091283
Panahi M, Gayen A, Pourghasemi HR, Rezaie F, Lee S (2020) Spatial prediction of landslide susceptibility using hybrid support vector regression (SVR) and the adaptive neuro-fuzzy inference system (ANFIS) with various metaheuristic algorithms. Sci Total Environ 741:139937. https://doi.org/10.1016/j.scitotenv.2020.139937
Pham BT, Qi C, Ho LS et al (2020) A novel hybrid soft computing model using random forest and particle swarm optimization for estimation of undrained shear strength of soil. Sustainability 12:2218. https://doi.org/10.3390/su12062218
Puri N, Prasad HD, Jain A (2018) Prediction of geotechnical parameters using machine learning techniques. In: 6th International Conference on Smart Computing and Communications, ICSCC 2017, 7–8, December, Kurukshetra, India, Procedia Computer Science, 125, pp 509–517
Rangel JL, Iturraran-Viveros U, Ayala AG, Cervantes F (2005) Tunnel stability analysis during construction using a neuro-fuzzy system. Int J Numer Anal Methods Geomech 29:1433–1456
Salahudeen AB, Sadeeq JA, Badamasi A, Onyelowe KC (2020) Prediction of unconfined compressive strength of treated expansive clay using back-propagation artificial neural networks. Nigerian J Eng 27(1):45–58
Sharma A (2020) Decision tree vs. random forest-which algorithm should you use? Analytics Vidhya
Sugeno M (1985) Industrial applications of fuzzy control. Elsevier Science Inc.
Venkatesh K, Bind YK (2020) ANN and neuro-fuzzy modeling for shear strength characterization of soils. In: Proceedings of the National Academy of Sciences, India Section A: Physical Sciences: 1–7. https://doi.org/10.1007/s40010-020-00709-6
Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Res 30(1):79–82
Willmott CJ, Matsuura K, Robeson SM (2009) Ambiguities inherent in sums-of-squares-based error statistics. Atmos Environ 43(3):749–752
Yadu L, Tripathi RJ, Singh D (2011) Comparison of fly ash and rice husk ash stabilized black cotton soil. Int J Earth SciEng 4(6):42–45

Authors and Affiliations

Mudassir Iqbal[1] · Kennedy C. Onyelowe[2,3] · Fazal E. Jalal[4]

[1] Shanghai Key Laboratory for Digital Maintenance of Buildings and Infrastructure, School of Naval Architecture, Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, China

[2] Department of Civil Engineering, Michael Okpara University of Agric, Umudike, Nigeria

[3] Department of Civil and Mechanical Engineering, Kampala International University, Kampala, Uganda

[4] Department of Civil Engineering, Shanghai Jiao Tong University, Shanghai, China