Asian Journal of Civil Engineering (2023) 24:3099–3119
https://doi.org/10.1007/s42107-023-00697-z

RESEARCH

# Probabilistic analysis of gravity retaining wall against bearing failure

Rashid Mustafa[1] · Pijush Samui[1] · Sunita Kumari[1] · Edy Tonnizam Mohamad[2] · Ramesh Murlidhar Bhatawdekar[2]

Received: 21 April 2023 / Accepted: 2 May 2023 / Published online: 13 May 2023


**Abstract**

Machine learning (ML) models have been extensively used in the stability check of gravity retaining wall. They are renowned as the most capable methods for predicting factor of safety (*FOS*) of gravity retaining wall against bearing failure. In this work, *FOS* against bearing is predicted based on extreme gradient boosting (XGBoost), random forest (RF) and deep neural network (DNN). To establish homogeneity and distribution of datasets, Anderson−Darling (AD) and Mann−Whitney U (M−W) tests are carried out, respectively. These three machine learning models are applied to 100 datasets by considering six influential input parameters for predicting *FOS* against bearing failure. The execution of the established machine learning models is assessed by several performance parameters. The obtained results from computational approach shows that DNN attained the best predictive performance with coefficient of determination ($R^2$) = 0.998 and root mean square error (*RMSE*) = 0.006 in the training phase and $R^2$ = 0.929 and *RMSE* = 0.053 in the testing phase. The models result are also analyzed by using rank analysis, regression error characteristics curve, and accuracy matrix. Sensitivity analysis is carried to know the relative importance of input variables.

**Keywords** Reliability analysis · DNN · RF · XGBoost · Rank analysis · Uncertainty analysis · Statistical testing

## Introduction

Predicting the failure of civil engineering structures and providing corrective measures is the main concern for the researchers nowadays. In geotechnical field, retaining wall is used to hold counteract forces of gravity to protect the structure. Stability of gravity retaining wall is checked against sliding, overturning and bearing failure. To measure the retaining wall failure against bearing, a parameter called Factor of Safety (*FOS*), defined as the ratio of net allowable bearing capacity of the foundation soil ($q_{na}$) to the maximum soil pressure ($q_{max}$), is calculated. Reliability study in geotechnical field recognized over the years early from the probabilistic method. Several scholars have done reliability analysis on retaining walls. Basha and Babu (2008) conducted inverse reliability analysis on cantilever sheet pile wall. Goh et al. (2009) analyzed reliability investigation of partial safety factor concept for cantilever wall. Chouksey and Fale (2017) conducted reliability analysis of retaining structure. They have used first-order reliability method (FORM) and second-order reliability method (SORM) to calculate reliability index linked with several kinds of failure. Dao-Bing et al. (2013) analyzed probabilistic investigation of retaining wall against sliding and overturning modes of failure are analyzed using theory of upper-boundary. Kumar and Roy (2017) used reliability approach to design reliability analysis using imprecise probability. In this research, they have used copula centered technique to examine the effect of copulas for modelling tri-variate distributions on system reliability. Low et al. (2011) analyzed effective system reliability study for cantilever wall and a slope. Menon and

✉ Rashid Mustafa
rashidm.phd19.ce@nitp.ac.in

Pijush Samui
pijush@nitp.ac.in

Sunita Kumari
sunitafce@nitp.ac.in

Edy Tonnizam Mohamad
edy@utm.my

Ramesh Murlidhar Bhatawdekar
rmbhatawdekar2@graduate.utm.my

1 Department of Civil Engineering, National Institute of Technology Patna, Patna, Bihar 800 005, India

2 Centre of Tropical Geoengineering (GEOTROPIK), Institute for Smart Infrastructure and Innovative Construction (ISIIC), Department of Civil Engineering, Universiti Teknologi Malaysia, 81310 Johor, Malaysia

Mangalathu (2011) designed the cantilever retaining wall against sliding failure and also performed reliability analysis. They have used FORM, SORM and MCS techniques for reliability analysis. Wang et al., (2020a, 2020b) done reliability study of retaining wall under mountain torrent. Alghaffar and Wellington (2005) performed reliability study of walls using British and European standards. Xiao et al. (2014) done random reliability study of gravity retaining wall. They have used fuzzy random reliability and FOSM method to perform reliability analysis. Zhang et al. (2022) have done reliability study of gravity retaining wall to check external stability of wall under seismic condition.

However, use of computational method in the geotechnical research is very much in the trend. Chen et al. (2019) investigated the retaining wall and predicted the FOS through AI approach. Goh and Kulhawy (2005) used the neural network (NN) concept for the reliability study of wall. Mishra et al. (2021) used machine learning approach for the probabilistic analysis of retaining wall. Wu et al. (2022) used convolutional neural network (CNN) is used to predict wall deflection made by excavation. Kaveh et al. (2013) used multi-objective genetic algorithm for constructability optimal design of reinforced concrete cantilever retaining walls. In this research seismic analysis of cantilever retaining wall using Mononobe-Okabe method. Zhang et al., (2017a, 2017b) computed deflection of wall outlines produced by braced excavations using multivariate adaptive regression splines (MARS) technique. Xiang et al. (2018) computed extreme deflections of wall by braced excavation in clayey soil using MARS model. Yong et al. (2022) used FEM and ANN to predict the deflection of wall produced by braced excavations. Kaveh and Soleimani (2015) analyzed cantilever retaining wall under both static and seismic condition. They have used colliding bodies of optimization (CBO) and democratic particle swarm optimization (DPSO). The static earth pressures computed by Coulomb and Rankine theory and dynamic earth pressure computed by Mononobe-Okabe theory. Kaveh and Laien (2017) used colliding bodies of optimization (CBO), enhanced colliding bodies of optimization (ECBO) and vibrating particles system (VPS) for the optimal design of cantilever retaining wall under both static and seismic condition. Zhang et al., (2017a, 2017b) done inverse investigation of wall and backfill properties in braced excavation using MARS model. Apart from the retaining wall, computational approach are extensively used in the other geotechnical arena. Pradeep et al. (2021) predicted strain in rock with the help of DNN and hybrid model ANFIS and metaheuristic algorithm like particle swarm optimization (PSO), firefly algorithm (FFA), genetic algorithm (GA) and grey wolf optimization (GWO). Kumar et al. (2021) analyzed pile foundation using soft computing techniques like minimax probability machine regression (MPMR), emotional neural network (ENN), group method of data handling (GMDH) and ANFIS as a substitute to the conventional approach. Wang et al., (2020a, 2020b) proposed XGBoost technique for the effective reliability study of slope stability. Ray et al. (2021) done reliability study of shallow foundation using soft computing methods. In this research they have used MPMR, ANN-PSO and ANFIS-PSO model for reliability study. Babu and Srivastava (2007) computed bearing capacity and settlement using response surface method (RSM). Shahin et al. (2003) used multi-layer perception (MLPs) and B-spline Neuro-fuzzy networks to predict settlement of shallow foundation. Jena et al. (2019) used ANN for the mapping of earthquake risk assessment (ERA). From this research it has been observed that ANN is quite beneficial in probabilistic valuation of earthquake with high $R^2$ and very low RMSE. Mustafa et al. (2023) analyzed gravity retaining wall under seismic condition. They have used three machine learning models namely minimax probability machine regression (MPMR), group method of data handling (GMDH) and Gaussian process regression (GPR) to predict factor of safety against sliding, overturning and bearing failure. Zhang and Goh (2013) used multivariate adaptive regression splines (MARS) for the geotechnical engineering system. Other ML methods and also reliability study executed can be stated to latest literatures. Kaveh and Khalegi (1998) used artificial neural network (ANN) for the prediction of 7-day and 28-day strength of concrete specimen. Kaveh and Khavaninzadeh (2023) used four meta-heuristic optimization to predict FRP strength. Kaveh et al. (2008) used genetic algorithm (GA) and neural networks for the optimal design of transmission towers. Kaveh and Iranmanesh (1998) used two artificial neural networks namely backpropagation neural net (BPN) and counterpropagation neural net (CPN) for the analysis and design of large scale space structures. Ali and Burhan (2023) used hybrid machine learning (ML) technique for construction cost assessment. Hashmi et al. (2023) predicted compressive strength of concrete using genetic algorithm (GA) based hybrid artificial neural network (ANN) model. The key objective of current study is to execute probabilistic analysis of gravity retaining wall based on bearing failure criteria using three machine learning algorithms XGBoost, RF and DNN. The reason behind choosing DNN over ANN as it has numerous hidden layers of processing, whereas a simple neural network just has an input, output, and hidden layer. The input data for DNN is propagated through an input layer, several hidden layers, and finally the output layer in a layered architecture. The input data are subjected to a set of mathematical operations known as weights and biases in each layer, and the output of one layer is used as the input in the following layer. To reduce the error between the predicted output and the actual output, a deep learning model's weights and biases are modified during the training phase. Recent advances in deep learning have been made in a number of areas, including the prediction of forest cover, flood and typhoon activity, image and speech recognition, traffic and other aspects, low-flow hydrological time series forecasting, weather forecasting, and natural language processing. Numerous industry leaders in technology are steadily preparing to implement deep neural network. Considering the benefits of utilizing a deep learning technique as; it provides results of a high caliber, the ability to fully utilize unstructured data, the removal of unnecessary costs, and the need for data labelling. These models are also assessed by using the numerous statistical performance parameters.

## Details of present study

The *FOS* against bearing failure is define as the ratio of allowable bearing pressure ($q_{na}$) to the maximum applied pressure ($q_{max}$). Maximum applied pressure depends on unit weight of wall ($\gamma_{wall}$), dimension of the wall such as top width (a), bottom width (b), height of wall (H), base width of the wall (B) and cohesion ($c_b$), angle of internal friction ($\varphi_b$) and unit weight ($\gamma_b$) of backfill. Allowable bearing pressure depends on the B, depth of foundation ($D_f$), shape of footing and mainly on the cohesion ($c_f$), angle of internal friction ($\varphi_f$) and unit weight ($\gamma_f$) of the foundation soil. Otherwise $\gamma_{wall}$, a, b, H, B and $D_f$ are the constant for this study. Allowable bearing pressure can be computed by (IS 1981) as:

$$q_{nu} = c_f N_c s_c d_c i_c + q\left(N_q - 1\right) s_q d_q i_q + 0.5\, B\, \gamma_f N_\gamma s_\gamma d_\gamma i_\gamma R_w \tag{1}$$

where, $q_{nu}$ is the net ultimate bearing capacity, $q$ is the effective pressure at the base and $R_w$ is the water table correction factor. In this study $R_w$ is taken as unity as it has been assumed that water table is at or below a depth of ($D_f$+$B$). $N_c$, $N_q$ and $N_\gamma$ are the bearing capacity factor computed by Vesic's bearing capacity theory as in Murthy (2003) and Das (1998).

$$Nq = tan^2(45 + \varphi_f/2)e^{(\pi tan\varphi_f)} \tag{2}$$

$$N_c = (Nq - 1)\, cot\varphi_f \tag{3}$$

$$N\gamma = 2\left(N_q + 1\right) tan\varphi_f \tag{4}$$

$s_c$, $s_q$ and $s_\gamma$ are the shape factors and taken as unity for this study (as the strip footing is considered for this study). $d_c$, $d_q$ and $d_\gamma$ are the depth factors and computed as:

$$d_c = 1 + 0.2\left(D_f/B\right) tan\left(45 + \varphi_f/2\right) \tag{5}$$

$$d_q = d_\gamma = 1 \quad for \quad \varphi_f < 10^0$$

$$d_q = d_\gamma = 1 + 0.1\left(D_f/B\right) tan\left(45 + \varphi_f/2\right) \quad for \quad \varphi_f > 10^0 \tag{6}$$

Inclination factors $i_c$, $i_q$ and $i_\gamma$ can be computed as:

$$i_c = i_q = (1 - \alpha/90)^2 \tag{7}$$

$$i_\gamma = \left(1 - \alpha/\varphi_f\right)^2 \tag{8}$$

where, $\alpha$ is the load inclination with vertical and taken as zero degree for this study it has been assumed that load on the footing is vertical and uniformly distributed.

The net allowable bearing pressure ($q_{na}$) can be computed as:

$$q_{na} = \frac{q_{nu}}{FOS} \tag{9}$$

Here, *FOS* is taken as 2 to 3 for bearing capacity as per Bowles (1997) and Terzaghi et al. (1996) and for this study it is taken as 3.0. The maximum soil pressure ($q_{max}$) can be computed as:

$$q_{max} = \frac{\sum W}{B}(1 + 6e/B) \tag{10}$$

where, $\sum W$ is the total vertical forces acting on the wall and $e$ is the eccentricity which can be computed as:

$$e = \frac{B}{2} - \frac{\sum M_R - \sum M_O}{\sum W} \tag{11}$$

where, $\sum M_R$ is the sum of resisting moment and $\sum M_O$ is the summation of overturning moment. *FOS* of gravity retaining wall against bearing failure ($FOS_{bearing}$) can be computed as:

$$FOS_{bearing} = \frac{q_{na}}{q_{\max}} \tag{12}$$

Reliability index (β) is computed using FOSM approach. FOSM method is exceptionally influential probabilistic approach to compute β. First order Taylor series approximation is used in FOSM method to express the performance function. In this method, $\mu_P$ and $\sigma_P$ are the average value and the standard deviations of the output function P, respectively. Allowable bearing pressure ($q_{na}$) and maximum soil pressure ($q_{max}$) are signified as resistance ($R$) and load ($S$), respectively; and $\mu_R$ and $\mu_S$ are the average value and $\sigma_R$ and $\sigma_S$ are the standard deviations of $R$ and $S$, respectively; the performance function (P) is expressed as per Christian (2004).

$$P = g(R, S) = R - S \begin{Bmatrix} > 0, \ Safe \\ = 0, \ Verge\ of\ failure \\ < 0, \ Failure \end{Bmatrix} \tag{13}$$

Stages to find β of the performance function as per Hasofer and Lind (1974) and Cornell (1969). (Cornell, 1969; Hasofer & Lind, 1974):

(1) Express the variables R and S in non-dimensional system:

$$\mu_1 = \frac{R-\mu_R}{\sigma_R}, \quad \text{and} \quad \mu_2 = \frac{S-\mu_S}{\sigma_S} \tag{14}$$

(2) Change the boundary state task associated by reduced variables which signifies straight lines.

$$g(R, S) = R - S \tag{15}$$

(3) The smallest perpendicular distance to the $g$ $(R, S)$ from the origin gives the β of the performance function $P$ as shown in Fig. 1.

$$\beta = \frac{\mu_P}{\sigma_P} = \frac{\mu_R-\mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \tag{16}$$

Bearing failure probability mainly depend on the average and variance of achieved *FOS*; hence the reliability index in terms of *FOS* is achieved as:

$$\beta = \frac{\mu_{FOS}-1}{\sigma_{FOS}} \tag{17}$$

where, $\mu_F$ is the average of factor of safety and σF is the standard deviation of FOS. The probability of failure ($P_f$) can be computed as:

$$P_f = 1 - \Phi(\beta) \tag{18}$$

where, $\Phi(\beta)$ is the standard normal cumulative probability.

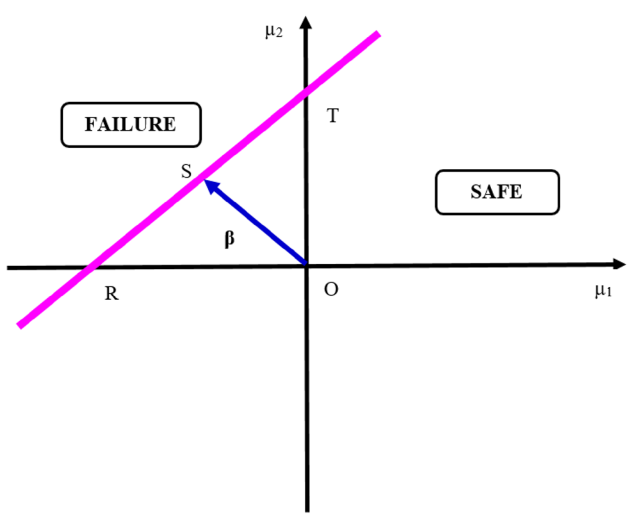



**Fig. 1** Reliability index (β) designated as the minimum distance from origin

## Methodology and theoretical background of models

### Extreme gradient boosting (XGBoost)

XGBoost is a machine learning (ML) system for tree boosting. It is boosted under the gradient boosting framework and established by Chen and Guestrin (2016). The vital aim of boosting is to merge a chain of feeble classifiers with less precision to form a robust classifier with enhanced classification conduct.

Supposing that a set of data is $S=\{(x_i, y_i): i=1....n, x_i \in R^m, y_i \in R\}$. Here $n$ is the whole number of samples with complete details m. Let $y_{p,i}$ be the predicted value of the model and can be defined as:

$$y_{p,i} = \sum_{j=1}^{j} f_j(x_i), \quad f_j \in F \tag{19}$$

where, $f_j$ signifies a self-regulating regression tree and $f_j(x_i)$ represents the prediction rank given by the *jth* tree to *ith* sample. In the tree model, $f_j$ can be learned by diminishing the objective function ($OBJ_f$):

$$OBJ_f = \sum_{i=1}^{n} l(y_i, y_{p,i}) + \sum_{j=1}^{j} \Omega(f_j) \tag{20}$$

where, $l$ is the training loss function, compute the difference between the predicted value ($y_{p,i}$) and observed value ($y_i$). To prevent over-fitting, the term Ω punishes the complication of the model:

$$\Omega(f_j) = \alpha L + \frac{1}{2}\lambda||w||^2 \tag{21}$$

where, λ and $\alpha$ are the regularization degree, $w$ and $L$ are the ranks on each leaf and numbers of leaves, respectively.

The tree ensemble model can be trained in supplement mode. Let $y_{p,i}^{(k)}$ be the prediction of the *ith* occurrence at the *kth* repetition, it requires to add $f_k$ to reduce the objective as follows:

$$OBJ_f(k) = \sum_{i=1}^{i} (y_i, y_{(p,i)}^{(k-1)} + f_{(k)}(x_i)) + \Omega(f_k) \tag{22}$$

Equation (23) can be obtained by simplifying the Eq. (22) using Taylor series expansion and removing the all constant terms as:

$$OBJ_f(k) = \sum_{i=1}^{i} [g_i f_k(x_i) + \frac{1}{2} h_i f_k(x_i)^2] + \Omega(f_k) \tag{23}$$

where $g_i = \partial_{y_{p,i}^{(k-1)}} l\,(y_i, y_{p,i}^{(k-1)})$ and $h_i = \partial^2_{y_{p,i}^{(k-1)}} l\,(y_i, y_{p,i}^{(k-1)})$ are the first and second order gradient on $l$. The $OBJ_f^{(k)}$ can be further expressed as:

$$OBJ_f(k) = \sum_{i=1}^{n}[g_i f_k(x_i) + \frac{1}{2}h_i f_k(x_i)^2] + \alpha L + \frac{1}{2}\lambda\sum_{j=1}^{L} w_j^2$$
$$= \sum_{j=1}^{L}\left(\left(\sum_{i\in I_j} g_i\right)s_j + \frac{1}{2}\left(\sum_{i\in I_j} h_i + \lambda\right)w_j^2\right] + \alpha L \tag{24}$$

where, $I_j = \{ i \mid Q(x_i) = j\}$ indicates the instance set of leaf $j$. For a fixed tree structure Q, the optimal weight $w_j^*$ of leaf $j$. The corresponding optimal value can be computed as:

$$w_j^* = -\frac{G_j}{H_j + \lambda} \tag{25}$$

$$OBJ_f^* = -\frac{1}{2}\sum_{j=1}^{L}\frac{G_j^2}{H_j+\lambda} + \lambda L \tag{26}$$

where, $G_j = \sum_{i\in Ij} g_i$, $H_j = \sum_{i\in Ij} h_i$, $OBJ$ represents the class of a tree structure Q. Lesser value indicates superior assembly of the tree. Since it is very difficult to specify entire tree assemblies, a self-indulgent algorithm is used to add branches of the tree repeatedly. For the purpose of evaluating split candidates, the gain formula is used and can be expressed as:

$$G = \frac{1}{2}\left[\frac{(\sum_{i\in I_L} g_i)^2}{\sum_{i\in I_L} h_i + \lambda} + \frac{(\sum_{i\in I_R} g_i)^2}{\sum_{i\in I_R} h_i + \lambda} - \frac{(\sum_{i\in I} g_i)^2}{\sum_{i\in I} h_i + \lambda}\right] - \alpha \tag{27}$$

where, $I_R$ and $I_L$ are the instance sets of the right and left nodes after splitting. Working steps of XGBoost are shown in Fig. 2

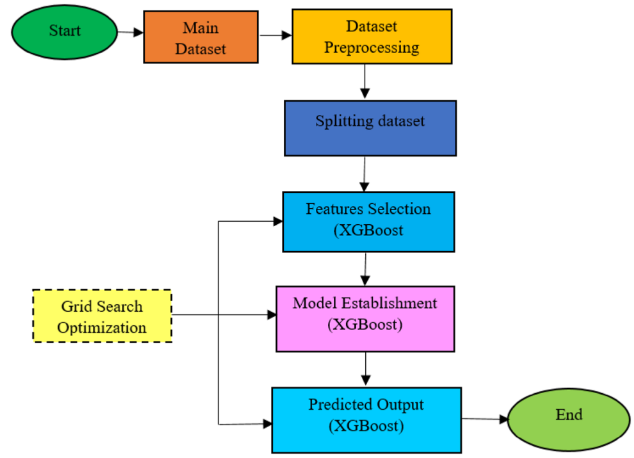



**Fig. 2** Flowchart of XGBoost

## Random forest (RF)

RF is another ML algorithm established on the idea of ensemble learning as per Liaw and Wiener (2002). RF is a classifier that comprises various decision tress on numerous subgroups of the specified data and taking average to advance the predictive validity of that data. Rather depend on single decision tree, RF takes the prediction from each individual tree and predicts the ultimate output established by majority votes of predictions (average voting). The major advantages of using RF are that, it takes lesser training time and predict output with great precision. Even for very large datasets it works proficiently. Process and characters of RF was defined by Leo Breiman (2001) as:

RF is a classifier comprising of a group of tree-structured classifiers {c (x, $\Theta_n$, n = 1, 2, ……} where {$\Theta_n$} are self-regulating unformly spread random vectors and each tree casts an element vote for the utmost widely held class at input vector x.

Each tree is planted based on random variable and training sample set. The random variable analogous to nth tree is $\Theta_n$ and obtain classifiers after n times running as {$c_1(x)$, $c_2(x)$, ……$c_n(x)$}. With the help of these classifiers create along with classification model system and the concluding outcome drawn by ordinary majority vote. Let us assume $C(x)$ be the blend of model classification, $c_i$ is the single decision tree, $Y$ is the output variable and $I(.)$ is the indicator function, then decision function can be obtained as:

$$C(x) = \underset{Y}{\arg\max}\sum_{i=1}^{n} I(c_i(x) = Y) \tag{28}$$

Margin function in RF is used to judge when the average votes at $X, Y$ for the right class surpass that for the wrong classs. The margin finction ($MF$) can be defined as:

$$MF(X,Y) = av_n I\left(c_n(X) = Y\right) - max_{J\neq Y} av_n I\left(c_n(X) = j\right) \tag{29}$$

Higher the margin value indicates greater precision of the classification prediction results in further reliance in classification.

Generalization error of the classifier can be defined as:

$$PE^* = P_{X,Y}(MF(X,Y < 0) \tag{30}$$

when there is very high number of decision tree ($n$), $c_n(X) = c(X, \Theta_n)$ follow the resilient law of huge quantity. When $n$ is very high $PE^*$ converges to:

$$P_{X,Y}(P_\theta\left(c_n(X,\theta) = Y\right) - max_{J\neq Y}P_\theta(c(X,\theta) = j) < 0) \tag{31}$$

Upper bound of the simplification error can be expressed as:

$$PE^* \leq \mu \left(1 - \tau^2\right) / \tau^2 \tag{32}$$

where $\tau$ is classifiers strength { $c\ (x,\ \theta)$}, $\mu$ is the average value of the correlation. From Eqs. 32 it is concluded that generalization error influenced by the strength of the individual trees and the correlation among these trees. Lesser value of thses indicates better prediction result of RF.

During formation of RF, by using random features selection tree is planted on the new training set and the new training set is drawn from the actual set of training by bagging methods. The main purpose of using bagging is that, it improve accuracy and gives the idea of strength and correlation. Let S be the total actual training set with N samples, the nth set of training is drawn from S with replacement by bagging, every $S_n$ comprises N samples. The probability of each sample can not be contain is $(1 - 1/N)^N$. When N is very high, $(1 - 1/N)^N$ is converges to $e^{-1}$ or we can say that 36.8% sample of the S is not contained in $S_n$ and sample is called out-of-bag (OOB) data. Strength and correlation can be assessed using OOB techniques. The classic structure of RF model is shown in Fig. 3

## Deep neural network (DNN)

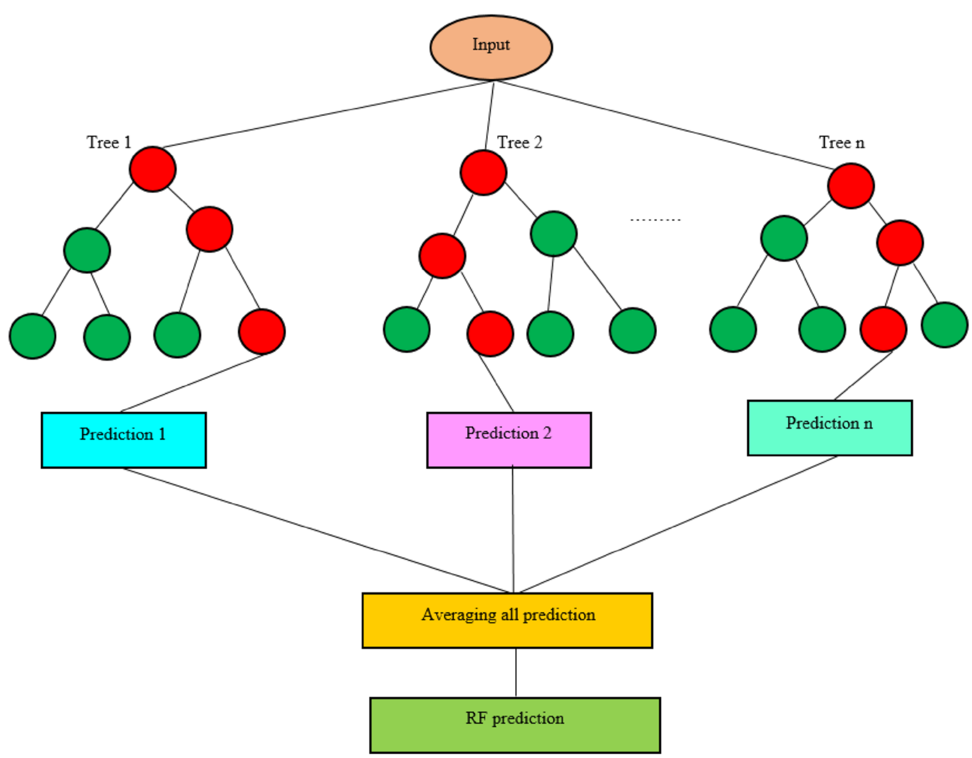
Deep neural networks (DNNs) are the improvement over conventional ANN with multi-layered architecture as in Jiang et al. (2019). The fully linked, three-layer feedforward network is the topology of ANN used in supervised learning the most frequently. All of the network's input values are connected to every neuron in the hidden layer, and every neuron in the output layer is connected to every neuron in the hidden layer's outputs, which together make up the entire network's output when the output neurons are activated. DNN is a type of ANN with various hidden layer in among input and output layers. In DNN back-propagation (BP) methods are used to learn challenging configurations in dataset. To work out the illustration of individual layer from the illustrations of the prior layer, BP approaches regulate the learning factors of DNNs. In DNNs, an input layer, number of hidden layers and an output layer exists. Various hyper-parameters like broader vs. deeper networks, neuron count in hidden layers, optimizer, batch size, loss function and epochs affect the architecture of DNNs. Under-fitting and over-fitting are the two main issues of DNN. Under-fitting issue can be removed by increasing network capacity and by regularization strategies like weight constraint, weight decay and early termination with dropout can handle this type of issue.

DNN can be trained by BP algorithm. The weight upgrades can be done by stochastic gradient descent as per Benuwa et al. (2016) by using the Eq. (33):

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \frac{\partial C}{\partial w_{ij}} \tag{33}$$

where, $C$ is the cost function, and α is the rate of learning. $C$ is influenced by several features such as type of learning (supervised, unsupervised, reinforcement) and activation

**Fig. 3** Basic structure of RF



function. When executing supervised learning, softmax and cross entropy function are the collective varieties for the activation function and cost function respectively. The softmax function can be defined as:

$$p_j = \frac{\exp(x_j)}{\sum_m \exp(x_m)} \tag{34}$$

where, $p_j$ is the class probability to output of the unit $j$ and $x_j$ and $x_m$ denotes the entire input to units $j$ and $m$, respectively of the similar level. Cross entropy ($C$ in supervised learning on multiclass grouping difficulties) can be expressed as:

$$C_r = \sum_j t_j \log(p_j) \tag{35}$$

where, $t_j$ indicates the target probability for output unit $j$. Figure 4 shows the structure of DNN model.

## Dataset preparation

Gravity retaining wall is considered to study probabilistic study against bearing failure (Fig. 5). The input data have been created randomly in excel using NORM.INV (RAND (), mean, standard_dev) command. Specifically, input variables, i.e., cohesion ($c_b$), angle of internal friction ($\varphi_b$), and unit weight ($\gamma_b$) of backfill and cohesion ($c_f$), angle of internal friction ($\varphi_f$) and unit weight ($\gamma_f$) of foundation soil have been created to find the output variables, *FOS* against bearing $(FOS)_{bearing}$ involving Eqs. (1–12). For this purpose mean and coefficient of variation of backfill properties were taken from previous research's GuhaRay et al. (2018) and Zhou et al. (2014).The mean value and coefficient of variation (CoV) of foundation soil were taken from the research paper Zevgolis and Bourdeau (2006). Figure 5 indicates the geometry of gravity wall and numerous dimension taken

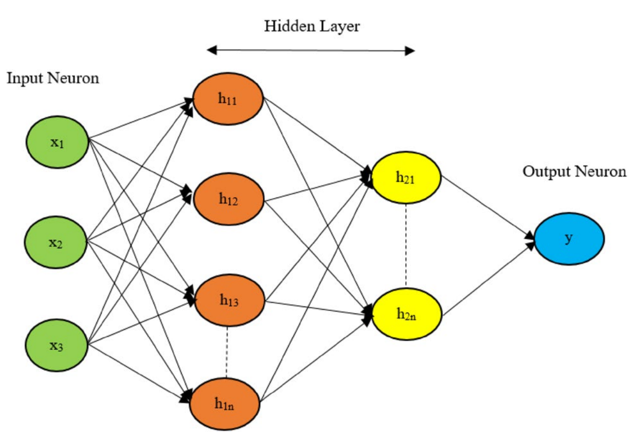



Fig. 4 Structure of deep neural network (DNN) model

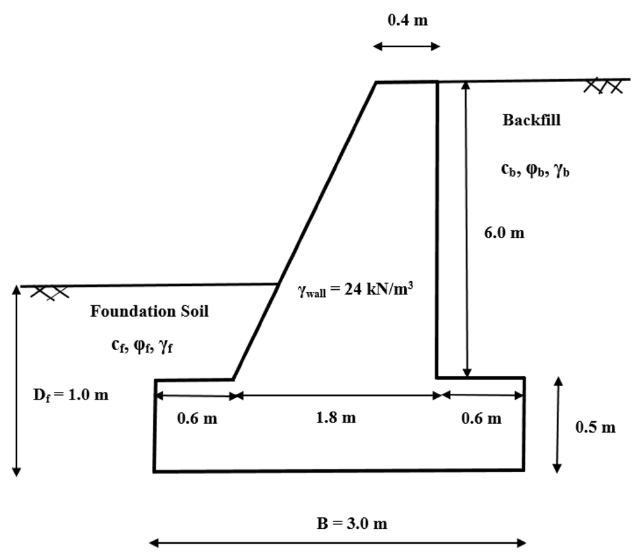



Fig. 5 Gravity retaining wall geometry

from the research paper Kumar and Roy (2017). The statistical depiction of input parameters are given in Table 1.

100 data sets were taken and the input ($c_b$, $\varphi_b$, $\gamma_b$, $c_f$, $\gamma_f$, and $\varphi_f$) and output (Factor of safety against bearing failure) variables have been normalized among 0 and 1, before spending in the model. The normalization of the dataset can be done as follows:

$$D_{Normalized} = \frac{D - D_{\min}}{D_{\max} - D_{\min}} \tag{36}$$

where, $D_{max}$ and $D_{min}$ are the maximum and minimum values of the parameter ($D$), respectively. After the normalization, dataset is separated into two subsets namely training (TR) and testing (TS). For this, 70% of the total dataset is taken randomly for training phase (70 data) and rest 30% is taken for the testing phase (30 data). The methodology flowchart is presented in Fig. 6.

## Models accuracy assessment

The prediction power of AI based computational models used in this work like XGBoost, RF and DNN were inspected using numerous performance parameters. Statistical parameters are further subdivided into trend measuring statistical parameters (TMSP) and error measuring statistical parameters (EMSP).

### Trend measuring statistical parameters (TMSP)

To know the predictive power of used model, seven variety of trend measuring statistical parameters are used. The coefficient

**Table 1** Statistical depiction of input and output parameters

| Parameters | INPUT | | | | | | OUTPUT |
|---|---|---|---|---|---|---|---|
| | Backfill Properties | | | Foundation soil properties | | | |
| | Cohesion ($c_b$) (in kN/m$^2$) | Unit weight ($\gamma_b$) (in kN/m$^3$) | Angle of shearing resistance ($\varphi_b$) (in degree) | Cohesion ($c_f$) (in kN/m$^2$) | Unit weight ($\gamma_f$) (in kN/m$^3$) | Angle of shearing resistance ($\varphi_f$) (in degree) | FOS against bearing failure |
| Mean | 11 | 16 | 29 | 30 | 16 | 28 | 6.33 |
| Standard deviation | 2.2 | 0.96 | 3.48 | 9 | 1.12 | 5.6 | 4.58 |
| Minimum | 5.64 | 13.49 | 21.14 | 11.6 | 12.20 | 17.11 | 1.16 |
| First Quartile | 9.83 | 15.16 | 26.47 | 23.42 | 15.42 | 24.52 | 2.98 |
| Second Quartile | 11.53 | 15.98 | 28.72 | 31.46 | 16.08 | 28.77 | 5.18 |
| Third Quartile | 13.13 | 16.39 | 30.97 | 36.61 | 16.76 | 31.70 | 7.98 |
| Maximum | 16.93 | 17.84 | 37.52 | 51.14 | 17.81 | 39.42 | 29.65 |
| Sample Variance | 5.85 | 0.96 | 11.49 | 81.00 | 1.10 | 26.01 | 21.01 |
| Range | 11.29 | 4.34 | 16.39 | 39.54 | 5.60 | 22.31 | 28.49 |
| Standard Error | 0.22 | 0.096 | 0.35 | 0.90 | 0.11 | 0.56 | 0.46 |
| 5% Trimmed mean | 11.44 | 15.85 | 28.79 | 30.03 | 16.06 | 28.59 | 6.04 |
| Skewness | – 0.097 | – 0.019 | 0.286 | – 0.09 | – 0.65 | 0.03 | 2.19 |
| Kurtosis | – 0.138 | – 0.455 | – 0.105 | – 0.49 | 0.80 | – 0.53 | 7.07 |
| Geometric mean | 11.18 | 15.81 | 28.63 | 28.56 | 15.99 | 28.12 | 5.14 |
| Harmonic mean | 10.89 | 15.79 | 28.44 | 26.87 | 15.96 | 27.65 | 4.21 |

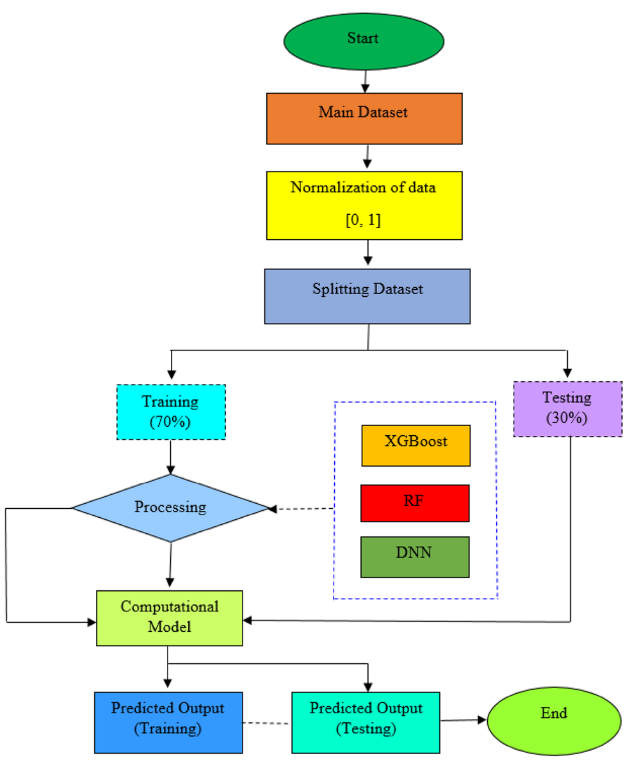



**Fig. 6** Methodology flowchart

of determination ($R^2$), performance index (PI), variance account factor (*VAF*), Willmott's index of agreement (*WI*), Legate and McCabe's index (*LMI*), Kling Gupta efficiency (*KGE*), and *a-20 Index* are seven employed TMSP used to evaluate the efficacy of the predictive models in this study. The mathematical terms for these parameters are expressed as follows:

$$R^2 = \frac{\sum_{i=1}^{n}(F_{o,i} - \overline{F_o})^2 - \sum_{i=1}^{n}(F_{o,i} - F_{p,i})^2}{\sum_{i=1}^{n}(F_{o,i} - \overline{F_o})^2} \tag{37}$$

$$VAF = \left(1 - \frac{\mathrm{var}(F_{o,i} - F_{p,i})}{\mathrm{var}(F_{o,i})}\right) \times 100 \tag{38}$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{n}(F_{o,i} - F_{p,i})^2}{\sum_{i=1}^{n}(|F_{p,i} - \overline{F_o}| + |F_{o,i} - \overline{F_o}|)^2}\right] \tag{39}$$

$$LMI = 1 - \left[\frac{\sum_{i=1}^{n}|F_{o,i} - F_{p,i}|}{\sum_{i=1}^{n}|F_{o,i} - \overline{F_o}|}\right] \tag{40}$$

$$PI = AdjR^2 + 0.01VAF - RMSE \tag{41}$$

$$KGE = 1 - \sqrt{(R^2-1)^2 + \left(\frac{\overline{F_p}}{\overline{F_o}}-1\right)^2 + \left(\frac{COV_p}{COV_o}-1\right)^2} \tag{42}$$

$$a-20\,Index = \frac{K20}{n} \tag{43}$$

where, $F_{o,i}$ and $F_{p,i}$ are the actual and predicted i$^{th}$ value, respectively, $\overline{F}_o$ and $\overline{F}_p$ are the average of actual and predicted value, respectively, $n$ is the number of training or testing samples and $K20$ is the amount of data with observed/predicted ratio between 0.80 and 1.20.

### Error measuring statistical parameter (EMSP)

To inspect model accuracy, seven EMSP are used. These parameters are root mean squared error (RMSE), mean absolute error (*MAE*), mean bias error (*MBE*), expanded uncertainty ($U_{95}$), scatter index (*SI*), median absolute deviation (*MAD*), and mean square error (*MSE*). Mathematical expression of EMSP are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(F_{o,i}-F_{p,i})^2}{n}} \tag{44}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(F_{p,i}-F_{o,i})| \tag{45}$$

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(F_{p,i}-F_{o,i}) \tag{46}$$

$$SI = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_{o,i}-F_{p,i})^2}}{\overline{F_o}} = \frac{RMSE}{\overline{F_o}} \tag{47}$$

$$MAD = Median\,(|F_{p,1}-F_{o,1}|,\ |F_{p,2}-F_{o,2}|,\ ----\ |F_{p,n}-F_{o,n}|) \tag{48}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(F_{p,i}-F_{o,i})^2 \tag{49}$$

$$U_{95} = 1.96\,(SD^2 + RMSE^2)^{\frac{1}{2}} \tag{50}$$

where, *SD* is the standard deviation of the dataset generated.

## Results and discussion

### Prediction capability

The prediction capability of the proposed models for the prediction of FOS against bearing failure is investigated in this section. For this study several statistical parameters for TR and TS datasets are computed for different model. Tables 2 and 3 shows statistical performance details for TR and TS dataset. The prediction power of proposed three models (XGBoost, RF and DNN) was evaluated by using numerous error measuring statistical parameters namely *MAD*, *MSE*, *RMSE*, *MBE*, *MAE*, $U_{95}$ and *SI* and trend measuring statistical parameters namely $R^2$, *PI*, *VAF*, *WI*, *LMI*, *KGE* and *a-20 index*. Result shows DNN has attained superior prediction capability in both TR and TS stage as compare to the other two models. In both

**Table 2** Details of performance parameters for the established computational model (TR dataset)

| Parameters | Name of the parameters | Ideal value | XGBoost (Training) | RF (Training) | DNN (Training) |
|---|---|---|---|---|---|
| MAD | Median absolute deviation | 0 | 0.006 | 0.015 | 0.001 |
| MSE | Mean square error | 0 | 0.001 | 0.002 | 3.86E-05 |
| $R^2$ | Coefficient of determination | 1 | 0.994 | 0.924 | 0.998 |
| RMSE | Root mean square error | 0 | 0.011 | 0.037 | 0.006 |
| PI | Performance index | 2 | 1.979 | 1.803 | 1.990 |
| VAF | Variance account factor | 100 | 99.644 | 92.302 | 99.795 |
| MBE | Mean bias error | 0 | 0.007 | -0.001 | 0.001 |
| WI | Willmott's index of agreement | 1 | 0.998 | 0.976 | 0.999 |
| MAE | Mean absolute error | 0 | 0.008 | 0.023 | 0.003 |
| LMI | Legate and McCabe's index | 1 | 0.918 | 0.772 | 0.968 |
| KGE | Kling Gupta efficiency | 1 | 0.911 | 0.766 | 0.989 |
| $U_{95}$ | Expanded uncertainty | 0 | 0.268 | 0.277 | 0.267 |
| SI | Scatter index | 0.1 | 0.066 | 0.231 | 0.039 |
| a-20 Index | a-20 index | 1 | 0.871 | 0.714 | 0.943 |

**Table 3** Details of performance parameters for the established computational model (TS dataset)

| Parameters | Name of the parameters | Ideal value | XGBoost (Testing) | RF (Testing) | DNN (Testing) |
|---|---|---|---|---|---|
| MAD | Median absolute deviation | 0 | 0.032 | 0.037 | 0.009 |
| MSE | Mean square error | 0 | 0.008 | 0.016 | 0.003 |
| $R^2$ | Coefficient of determination | 1 | 0.799 | 0.585 | 0.929 |
| RMSE | Root mean square error | 0 | 0.089 | 0.128 | 0.053 |
| PI | Performance index | 2 | 1.459 | 0.954 | 1.791 |
| VAF | Variance account factor | 100 | 79.998 | 60.512 | 98.399 |
| MBE | Mean bias error | 0 | 0.003 | -0.028 | -0.014 |
| WI | Willmott's index of agreement | 1 | 0.935 | 0.802 | 0.979 |
| MAE | Mean absolute error | 0 | 0.053 | 0.067 | 0.022 |
| LMI | Legate and McCabe's index | 1 | 0.622 | 0.515 | 0.843 |
| KGE | Kling Gupta efficiency | 1 | 0.706 | 0.065 | 0.860 |
| $U_{95}$ | Expanded uncertainty | 0 | 0.433 | 0.469 | 0.410 |
| SI | Scatter index | 0.1 | 0.389 | 0.560 | 0.232 |
| a-20 Index | a-20 index | 1 | 0.533 | 0.433 | 0.867 |

the two stages DNN has higher value of $R^2$ (TR = 0.998, TS = 0.929), $PI$ (TR = 1.990, TS = 1.791), $VAF$ (TR = 99.795, TS = 98.399), $WI$ (TR = 0.999, TS = 0.979), $LMI$ (TR = 0.968, TS = 0.843), $KGE$ (TR = 0.989, TS = 0.860) and $a\text{-}20\ index$ (TR = 0.943, TS = 0.867) and lower value of $MAD$ (TR = 0.001, TS = 0.009), $MSE$ (TR = 0.3.86E-05, TS = 0.003), $RMSE$ (TR = 0.006, TS = 0.053), $MBE$ (TR = 0.001, TS = −0.014), $MAE$ (TR = 0.003, TS = 0.022), $U_{95}$ (TR = 0.267, TS = 0.410) and $SI$ (TR = 0.039, TS = 0.232).

## Rank analysis

Rank analysis is used to compare the predictive model' performance as shown in Table 4. Each three model's TR and TS data are used to compute the rank. The number of models governs the rank from 1 to 3 (as three models are used in this study). Best performance model is ranked 1 and rank 3 is assigned for poorest performance model. The model's final rank is evaluated by totaling all of the rank for TR and testing TS data as:

**Table 4** Rank analysis of different model

| Model | XGBoost | | RF | | DNN | |
|---|---|---|---|---|---|---|
| | Training | Testing | Training | Testing | Training | Testing |
| MAD | 2 | 2 | 3 | 3 | 1 | 1 |
| MSE | 2 | 2 | 3 | 3 | 1 | 1 |
| $R^2$ | 2 | 2 | 3 | 3 | 1 | 1 |
| RMSE | 2 | 2 | 3 | 3 | 1 | 1 |
| PI | 2 | 2 | 3 | 3 | 1 | 1 |
| VAF | 2 | 2 | 3 | 3 | 1 | 1 |
| MBE | 2 | 1 | 3 | 3 | 1 | 2 |
| WI | 2 | 2 | 3 | 3 | 1 | 1 |
| MAE | 2 | 2 | 3 | 3 | 1 | 1 |
| LMI | 2 | 2 | 3 | 3 | 1 | 1 |
| KGE | 2 | 2 | 3 | 3 | 1 | 1 |
| $U_{95}$ | 2 | 2 | 3 | 3 | 1 | 1 |
| SI | 2 | 2 | 3 | 3 | 1 | 1 |
| a-20 Index | 2 | 2 | 3 | 3 | 1 | 1 |
| Total rank | 28 | 27 | 42 | 42 | 14 | 15 |
| Final rank | 55 | | 84 | | 29 | |

$$Final\,Rank = [\sum_{i=1}^{m} R_i + \sum_{j=1}^{n} R_j] \tag{51}$$

where, $R$ is the rank, $i$ and $j$ are the TR and TS performance gauges, and m and n are the number of performance gauges in the TR and TS phase respectively. DNN ($R_{TR} = 14$, $R_{TS} = 15$ and Final Rank = 29) is the best performing model followed by XGBoost ($R_{TR} = 28$, $R_{TS} = 27$ and Final Rank = 55) and RF ($R_{TR} = 42$, $R_{TS} = 42$ and Final Rank = 84). Illustration of rank analysis is also represented in the form of a radar diagram in Fig. 7

## Reliability index of the model

Reliability index (β) of proposed models are calculated with the help of FOSM approach and compared with the actual value of β. Probability of failure ($P_f$) is also calculated for these models by using reliability index. As we know that higher value of β shows better accuracy of the model. Models are ranked with the help of β and $P_f$ and can be shown in Table 5. Higher β and lesser $P_f$ (Top performing model) is ranked 1 and lower β and higher $P_f$ (Low performing model) is ranked 3. Table 5 shows DNN is the top performing model among three and RF is the lowermost performing model.

**Table 5** Comparison among different model based on reliability index and probability

| Models | Actual β | Actual $P_f$ | Model's β | Model's $P_f$ | Rank |
|---|---|---|---|---|---|
| XGBoost | 1.763 | 0.039 | 1.351 | 0.088 | 2 |
| RF | | | 1.240 | 0.107 | 3 |
| DNN | | | 1.711 | 0.043 | 1 |

## Performance curve

Performance or regression curve is the graphical representation among the observed and predicted factor of safety against bearing failure. It offers R-value computed and are shown in Tables 2 and 3. Figures 8 and 9 show the graphical representation among the actual FOS (Normalized) and predicted FOS (Normalized) against bearing failure using TR and TS dataset. In Figs. 8 and 9, dotted line indicates ± 15% deviance of the predicted data from the actual line. In the both training and testing phase, very less deviation was observed in the DNN model followed by XGBoost and RF.

## Regression error characteristics (REC) curve

To visualize decent outcomes, graphical analyses are offered. The REC curve helps in finding the inaccuracy in absolute deviation form (Fig. 10). The REC curve is receiver operating characteristics (ROC) plot

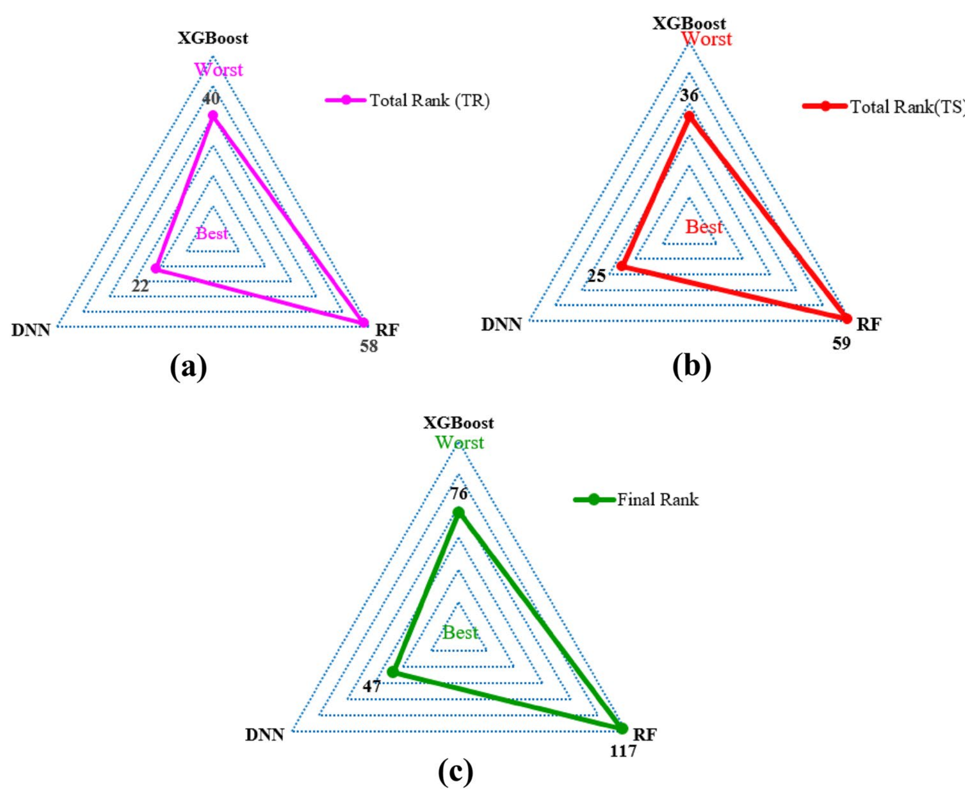



**Fig. 7** Rank analysis presenting in form of radar diagram: **a** TR stage; **b** TS stage; and **c** for final rank

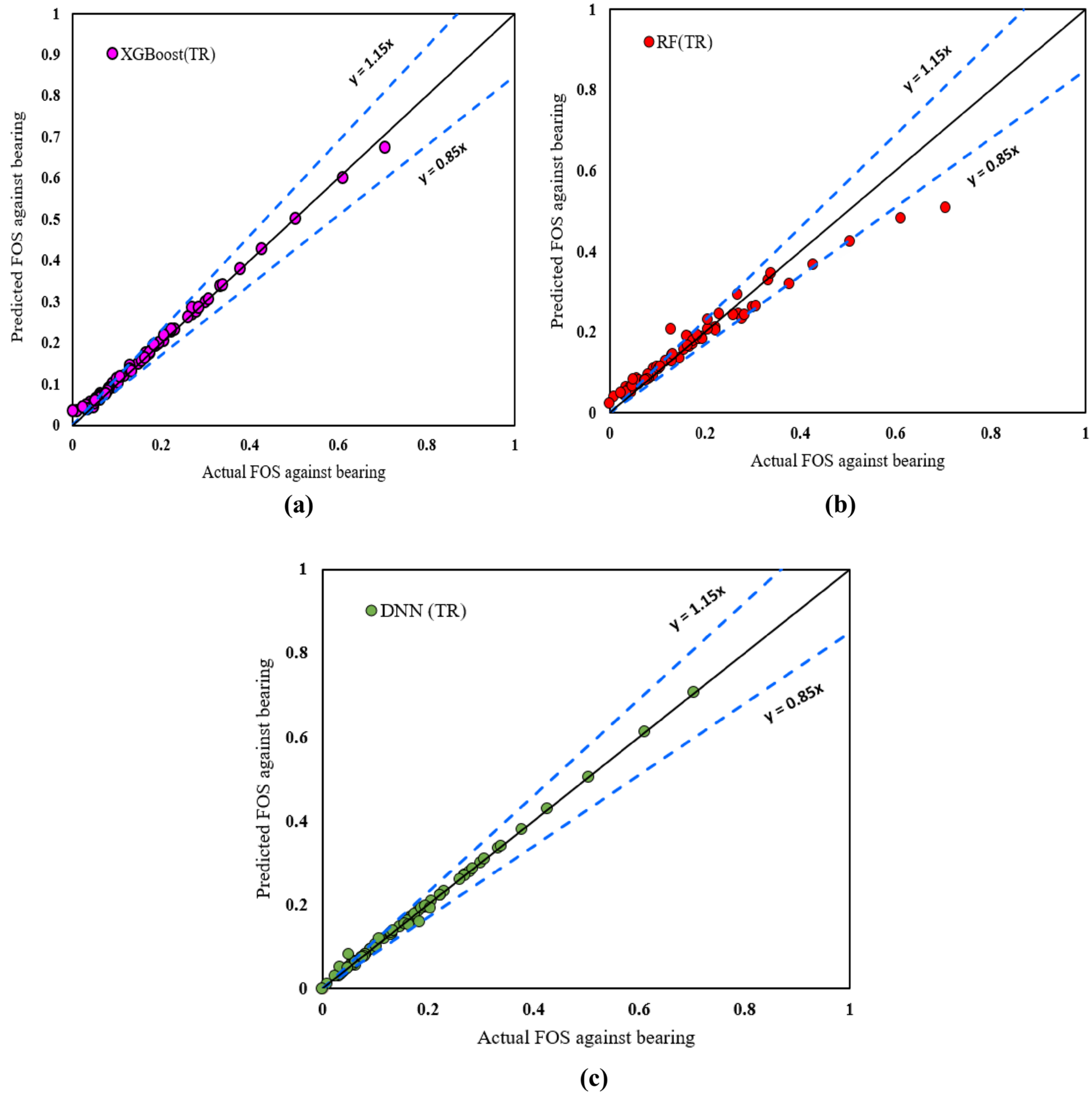



**Fig. 8** Performance curve of training (TR) data for **a** XGBoost; **b** RF; and **c** DNN models

in two-dimensional form in which absolute deviation is represented on abscissa and accuracy represented on the ordinates. The subsequent curve evaluates the cumulative distribution function (cdf) of the error among the observed and predicted values. The area over the curve (AOC) offers effective level of the enactment of a regression model. The AOC values of proposed models are given in Table 6. Gini coefficient ($G$) of each model is computed with the help of area under the curve (AUC). High value of $G$ specifies better performance of the model. Table 7 illustrates DNN model has higher value of G for both TR and TS phase followed by XGBoost and RF. On account of $G$, models are ordered accordingly. Models which has higher value of Gini coefficient is ranked 1 and lower value is ranked 3.

$$Gini\ Coefficient\,(G) = 2AUC - 1 \tag{52}$$

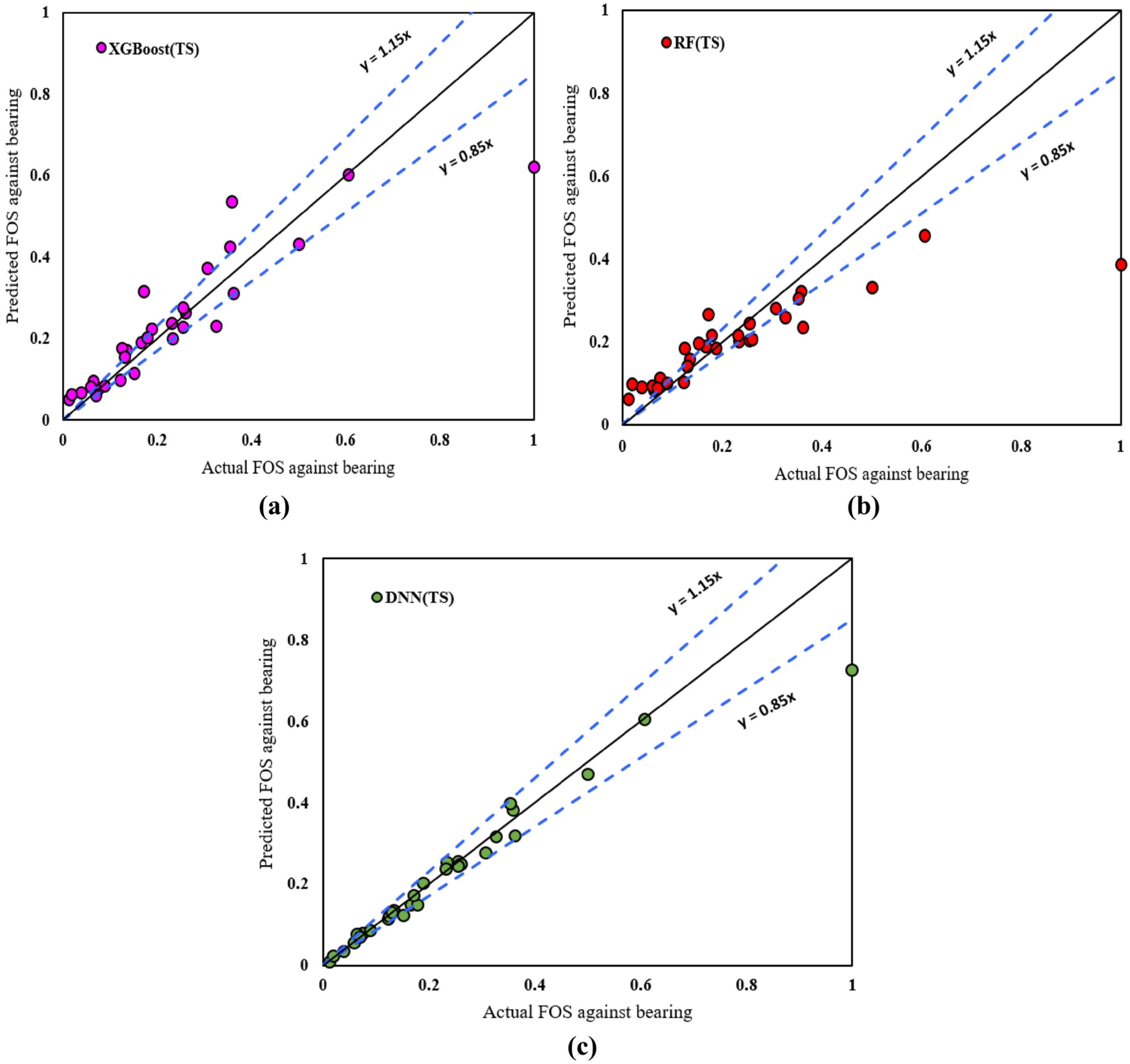



**Fig. 9** Performance curve of testing (TS) data for **a** XGBoost; **b** RF; and **c** DNN models

## Accuracy and error matrix

Accuracy matrix is a plot in the pattern of heat map, is inspected to assess the effectiveness of the models. The matrix contained numerous performance factors to regulate the predictive presentation of the model for TR and TS datasets. Figures 11 and 12 show the accuracy matrix of three models in forms of predicting factor of safety. So, the degree of error ($E\ \%$) can be calculated for error measuring statistical parameters (EMSP) and trend measuring statistical parameters (TMSP), respectively.

$$E_e = \left|(I_e - |S_e|)\right| \times 100 \tag{53}$$

$$E_t = \frac{(I_t - |S_t|)}{S_t} \times 100 \tag{54}$$

where, $E_e$ and $E_t$ represents the error for EMSP and error for TMSP respectively; $I_t$ and $I_e$ are the ideal values for TMSP and EMSP, respectively; $S_e$ and $S_t$ are the predicted values for EMSP and TMSP respectively. Using Eq. 53–54 error is calculated which are shown in Tables 8 and 9.

**Fig. 10** REC plot for the **a** TR dataset; and **b** TS dataset

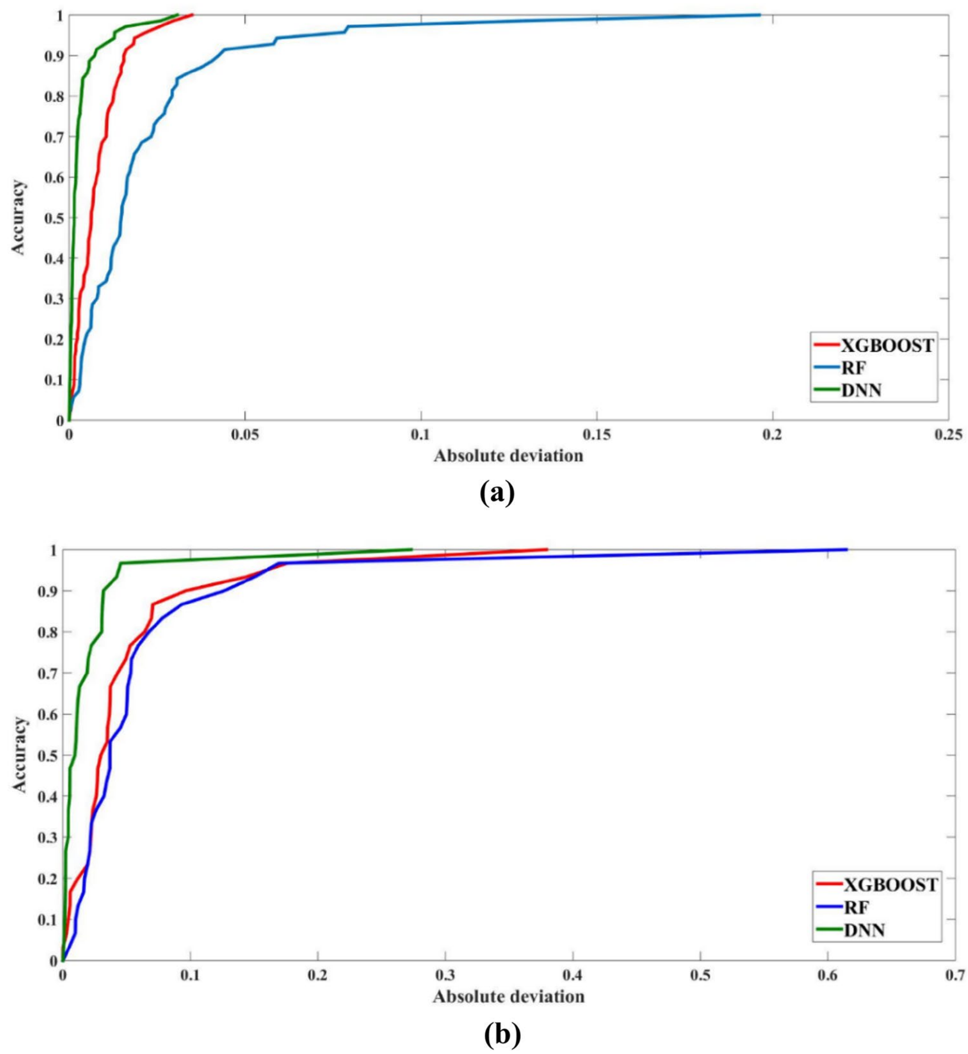



**Table 6** AOC values for the models

| Phase/Model | XGBoost | RF | DNN |
|---|---|---|---|
| Training (TR) | 0.0078 | 0.0211 | 0.003 |
| Testing (TS) | 0.0462 | 0.0571 | 0.0173 |

**Table 7** Gini coefficient values for proposed model

| Model | Phase | AUC | Gini Coefficient (G) | | Rank |
|---|---|---|---|---|---|
| | | | Ideal value | Calculated value | |
| XGBoost | TR | 0.9922 | 1 | 0.9842 | 2 |
| | TS | 0.9538 | | 0.9076 | 2 |
| RF | TR | 0.9789 | 1 | 0.9578 | 3 |
| | TS | 0.9429 | | 0.8858 | 3 |
| DNN | TR | 0.9970 | 1 | 0.9940 | 1 |
| | TS | 0.9827 | | 0.9654 | 1 |

Figure 11a and b indicate the error matrix plot for the EMSP and TMSP of the training TR) and testing (TS) dataset. In both the figure it can be seen that lesser error observed while predicting FOS against bearing failure in DNN model as compare to the other two models. The lesser range of error is revealed by the shade of green and the highest range of error shown by the shade of red. Figure 12 displays the accuracy plot for the TR and TS dataset of the proposed model. Once again highest range of accuracy witnessed for the DNN model among all the three proposed model. Uppermost scale of accuracy is presented by shade of green and lowest scale of accuracy presented by shade of red.

## Uncertainty analysis and statistical testing

### Uncertainty analysis

Quantitative analysis of the models in predicting FOS against bearing failure is presented in this sub-section. The analysis was done for the testing datasets. In uncertainty analysis the absolute error ($E_{a,\,i}$) can be computed as:

| | XGBoost (TR) | RF(TR) | DNN(TR) | XGBoost (TS) | RF(TS) | DNN(TS) |
|---|---|---|---|---|---|---|
| $R^2$ | 0.60% | 7.60% | 0.20% | 20.10% | 41.50% | 7.10% |
| PI | 1.10% | 9.90% | 0.50% | 27.10% | 52.30% | 10.50% |
| VAF | 0.40% | 7.70% | 0.20% | 20% | 39.50% | 1.60% |
| WI | 0.20% | 2.40% | 0.10% | 6.50% | 19.80% | 2.10% |
| LMI | 8.20% | 22.80% | 3.20% | 37.80% | 48.50% | 15.70% |
| KGE | 8.90% | 23.40% | 1.10% | 29.40% | 93.50% | 14% |
| a-20 Index | 12.90% | 28.60% | 5.70% | 46.70% | 56.70% | 13.30% |

**(a)**

| | XGBoost(TR) | RF(TR) | DNN(TR) | XGBoost(TS) | RF(TS) | DNN(TS) |
|---|---|---|---|---|---|---|
| MAD | 0.60% | 1.50% | 0.10% | 3.20% | 3.70% | 0.90% |
| MSE | 0.10% | 0.20% | 0.01% | 0.80% | 1.60% | 0.30% |
| RMSE | 1.10% | 3.70% | 0.60% | 8.90% | 12.80% | 5.30% |
| MBE | 0.70% | 0.10% | 0.10% | 0.30% | 2.80% | 1.40% |
| MAE | 0.80% | 2.30% | 0.30% | 5.30% | 6.70% | 2.20% |
| $U_{95}$ | 26.80% | 27.70% | 26.70% | 43.30% | 46.90% | 41% |
| SI | 6.60% | 23.10% | 3.90% | 38.90% | 56% | 23.20% |

**(b)**

**Fig. 11** Error matrix for the TR and TS dataset of proposed model **a** TMSP; and **b** EMSP

| | XGBoost (TR) | RF(TR) | DNN(TR) | XGBoost (TS) | RF(TS) | DNN(TS) |
|---|---|---|---|---|---|---|
| $R^2$ | 99.40% | 92.40% | 99.80% | 79.90% | 58.50% | 92.90% |
| PI | 98.95% | 90.15% | 99.50% | 72.95% | 47.70% | 89.55% |
| VAF | 99.64 | 92.30% | 99.80% | 79.99% | 60.51% | 98.40% |
| WI | 99.80% | 97.60% | 99.90% | 93.50% | 80.20% | 97.90% |
| LMI | 91.80% | 77.20% | 96.80% | 61.20% | 51.50% | 84.30% |
| KGE | 91.10% | 76.60% | 98.90% | 70.60% | 6.50% | 86% |
| a-20 Index | 87.10% | 71.40% | 94.30% | 53.30% | 43.30% | 86.70% |

**Fig. 12** Accuracy matrix for the TR and TS dataset of the proposed model

$$E_{a,i} = \left| F_{O,i} - F_{P,i} \right| \tag{55}$$

where, $F_{O,\,i}$ and $F_{P,\,i}$ be the observed and predicted factor of safety, respectively. The mean of absolute error (*MOAE*) and standard deviation ($\sigma$) of prediction can be computed as:

$$MOAE = \frac{\sum_{i=1}^{n} E_{a,i}}{n} \tag{56}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (E_{a,i} - MOAE)^2}{n-1}} \tag{57}$$

where, $n$ is number of testing dataset. Width of confidence bound (*WCB*) was computed by using margin of error (*MOE*) at 95% confidence interval. The standard error ($E_S$), lower bound ($L_B$) and upper bound ($U_B$) were computed as:

$$Standard\ Error = \frac{Standard\ Deviation}{\sqrt{n-1}} \tag{58}$$

$$Upper\ Bound = Mean\ of\ Absolute\ Error + Margin\ of\ Error \tag{59}$$

$$Lower\ Bound = Mean\ of\ Error - Margin\ of\ Error \tag{60}$$

$$Width\ of\ confidence\ Bound = Upper\ Bound - Lower\ Bound \tag{61}$$

Table 10 shows uncertainty analysis details of different parameters. The parameters shown in Table 10 are very much useful to measure the capability of the models. Lesser *WCB* indicated higher certainty of the model and the model will give less error and higher accuracy while predicting the output. All the three models used in this study were ranked according the *WCB* value. Table 10 shows DNN has lesser value of *WCB* among all the three model and secured rank 1 followed by XGBoost and RF. Figure 13 a–c shows the uncertainty analysis graph in the form of bar chart showing *MOAE*, *WCB*, and *MOE* value for a superior appraisal.

### Statistical testing

***t*-Test** *t*-test is the statistical test to compare the models. One-tailed test was performed to assess the substantial dissimilarity of the preferred DNN's execution to the others two models namely XGBoost and RF in predicting the FOS of gravity wall against bearing failure. This test was performed on the MBE values with the hypothesized mean difference MD = 0. At MD = 0 and 95% confidence interval, the hypotheses of one-tailed t-test are $HP_0$: $MBE_{DNN} - MBE_{RF/XGBoost\ models} = 0$ and $HP_A$: $MBE_{DNN} - MBE_{RF/XGBoost\ models} < 0$, where $HP_A$ and $HP_0$ denotes alternate hypothesis and null hypothesis, respectively. The outcomes are shown in Table 11. It can be observed that the refusal (failed to accept) of $HP_0$ (i.e. t-stat < t-critical) specifies the suggested DNN model outperformed all two models in decreasing the MAE value in both training and testing phase.

**Table 8** Calculated error for TMSP for TR and TS dataset

| TMSP | Ideal values | XGBoost | | Error ($E_t$) in % | | RF | | Error ($E_t$) in % | | DNN | | Error ($E_t$) in % | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | TR | TS | TR | TS | TR | TS | TR | TS | TR | TS | TR | TS |
| $R^2$ | 1 | 0.994 | 0.799 | 0.6 | 20.1 | 0.924 | 0.585 | 7.6 | 41.5 | 0.998 | 0.929 | 0.2 | 7.1 |
| PI | 2 | 1.979 | 1.459 | 1.1 | 27.1 | 1.803 | 0.954 | 9.9 | 52.3 | 1.990 | 1.791 | 0.5 | 10.5 |
| VAF | 100 | 99.644 | 79.998 | 0.4 | 20.0 | 92.302 | 60.512 | 7.7 | 39.5 | 99.795 | 98.399 | 0.2 | 1.6 |
| WI | 1 | 0.998 | 0.935 | 0.2 | 6.5 | 0.976 | 0.802 | 2.4 | 19.8 | 0.999 | 0.979 | 0.1 | 2.1 |
| LMI | 1 | 0.918 | 0.622 | 8.2 | 37.8 | 0.772 | 0.515 | 22.8 | 48.5 | 0.968 | 0.843 | 3.2 | 15.7 |
| KGE | 1 | 0.911 | 0.706 | 8.9 | 29.4 | 0.766 | 0.065 | 23.4 | 93.5 | 0.989 | 0.860 | 1.1 | 14 |
| a-20 Index | 1 | 0.871 | 0.533 | 12.9 | 46.7 | 0.714 | 0.433 | 28.6 | 56.7 | 0.943 | 0.867 | 5.7 | 13.3 |

**Table 9** Calculated error for EMSP for TR and TS dataset

| EMSP | Ideal values | XGBoost | | Error ($E_e$) in % | | RF | | Error ($E_e$) in % | | DNN | | Error ($E_e$) in % | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | TR | TS | TR | TS | TR | TS | TR | TS | TR | TS | TR | TS |
| MAD | 0 | 0.006 | 0.032 | 0.6 | 3.2 | 0.015 | 0.037 | 1.5 | 3.7 | 0.001 | 0.009 | 0.1 | 0.9 |
| MSE | 0 | 0.001 | 0.008 | 0.1 | 0.8 | 0.002 | 0.016 | 0.2 | 1.6 | 3.86E-05 | 0.003 | 0.01 | 0.3 |
| RMSE | 0 | 0.011 | 0.089 | 1.1 | 8.9 | 0.037 | 0.128 | 3.7 | 12.8 | 0.006 | 0.053 | 0.6 | 5.3 |
| MBE | 0 | 0.007 | 0.003 | 0.7 | 0.3 | −0.001 | −0.028 | 0.1 | 2.8 | 0.001 | −0.014 | 0.1 | 1.4 |
| MAE | 0 | 0.008 | 0.053 | 0.8 | 5.3 | 0.023 | 0.067 | 2.3 | 6.7 | 0.003 | 0.022 | 0.3 | 2.2 |
| $U_{95}$ | 0 | 0.268 | 0.433 | 26.8 | 43.3 | 0.277 | 0.469 | 27.7 | 46.9 | 0.267 | 0.410 | 26.7 | 41.0 |
| SI | 0.1 | 0.066 | 0.389 | 6.6 | 38.9 | 0.231 | 0.560 | 23.1 | 56.0 | 0.039 | 0.232 | 3.9 | 23.2 |

**Table 10** Outcomes of uncertainty investigation

| Models | MOAE | σ | $E_S$ | MOE | $L_B$ | $U_B$ | WCB | Rank |
|---|---|---|---|---|---|---|---|---|
| XGBoost | 0.053 | 0.090 | 0.016 | 0.032 | 0.02 | 0.085 | 0.065 | 2 |
| RF | 0.067 | 0.131 | 0.024 | 0.047 | 0.03 | 0.114 | 0.093 | 3 |
| DNN | 0.022 | 0.054 | 0.009 | 0.019 | 0.003 | 0.041 | 0.039 | 1 |

**Non-parametric testing** In this sub-section 'Anderson–Darling' (A-D) test and 'Mann–Whitney U' (M-W) tests are done to inspect the normality and probability distribution of the observed and predicted factor of safety using three established machine learning (ML) models. In A-D test, if null hypothesis ($HP_0$) is accepted at a certain level of significance then the given data follows normal distribution as in Razali and Wah (2011). Table 12 shows the AD value as well as Adj. AD value for the entire actual and predicted datasets. It is observed that for actual and all the three models P-values are under the 5% significance level and failed to accept $HP_0$. It is also observed that model DNN is best performing model to predict factor of safety against bearing failure.

M-W test is used to differentiate the values among two sets (observed and predicted) which means whether they originate from the alike distribution as in Mann and Whitney (1947). In M-W test, rank is allotted to each sample in the initial stage and then assigned ranks are summed up. Afterward the 'Mann–Whitney $U$' value is computed as the minimum value among $U_1$ and $U_2$ and computed as:

$$U_1 = m_1 m_2 + \frac{m_1(m_1 + 1)}{2} - R_{t1}$$

$$U_2 = m_1 m_2 + \frac{m_2(m_2 + 1)}{2} - R_{t2} \tag{62}$$

where, $m_1$ and $m_2$ are total of samples of the corresponding group; $R_{t1}$ and $R_{t2}$ are the total ranks for group 1 and 2 respectively. z-stat can be computed as:

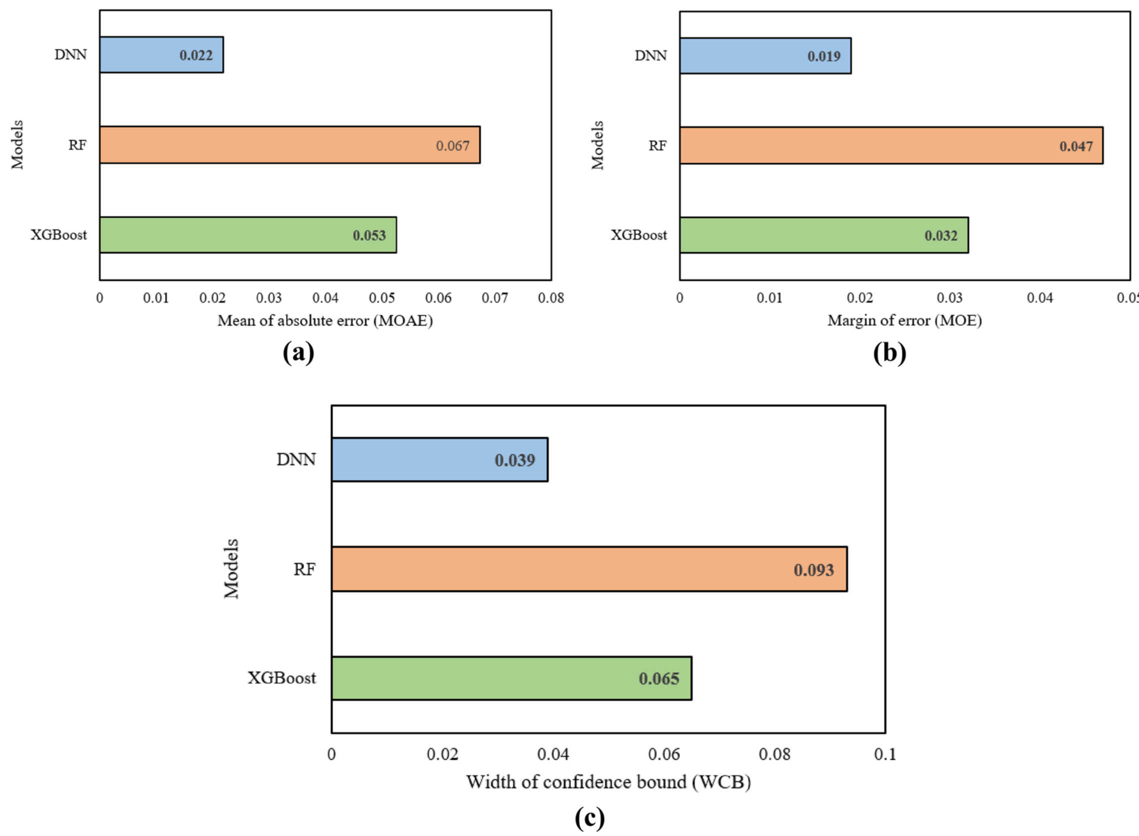



**Fig. 13** Bar plots for uncertainty analysis presenting: (a) MOAE; (b) MOE; and (c) WCB

**Table 11** Outcomes of t-test

| Phase | Models | Total sample | Degree of freedom (DOF) | MD | t-stat | t-critical | $HP_o$ |
|---|---|---|---|---|---|---|---|
| Training | XGBoost | 70 | 69 | 0 | 1.048 | 1.667 | Reject |
| | RF | 70 | 69 | 0 | 0.098 | 1.667 | Reject |
| Testing | XGBoost | 30 | 29 | 0 | 0.166 | 1.699 | Reject |
| | RF | 30 | 29 | 0 | 1.207 | 1.699 | Reject |

**Table 12** Features of A-D test for entire dataset

| Output variables | Observe/models | Total dataset | Mean | Standard deviation | AD value | Adj. AD value | *P*-value |
|---|---|---|---|---|---|---|---|
| Factor of Safety against Bearing Failure | Observe | 100 | 0.1815 | 0.1609 | 4.263 | 4.296 | 1.146E-10 (<<0.05) |
| | XGBoost | 100 | 0.1872 | 0.1427 | 4.010 | 4.041 | 4.687E-10 (<<0.05) |
| | RF | 100 | 0.1729 | 0.1043 | 5.075 | 5.114 | 1.242E-12 (<<0.05) |
| | DNN | 100 | 0.1780 | 0.1473 | 3.859 | 3.889 | 1.097E-09 (<<0.05) |

$$z{-}stat = \frac{U{-}Mean}{S\tan dard\ deviation} \quad (63)$$

$$Mean = \frac{m_1 m_2}{2}, \quad \text{Standard}\ deviation = \sqrt{\frac{m_1 m_2 (m_1 + m_2 + 1)}{12}} \quad (64)$$

**Table 13** Features of M-W test for entire dataset

| Output variables | Model group | Mann- Whitney U | Wilcoxon W | z-stat | Asymptotic Significance (2-Tailed:P value) |
|---|---|---|---|---|---|
| Factor of Safety against Bearing Failure | Observe-XGBoost | 4674 | 205 | −0.7965 | 0.4458 |
| | Observe-RF | 4640 | 865 | −0.8796 | 0.3945 |
| | Observe-DNN | 4993 | 545 | −0.0171 | 0.9859 |

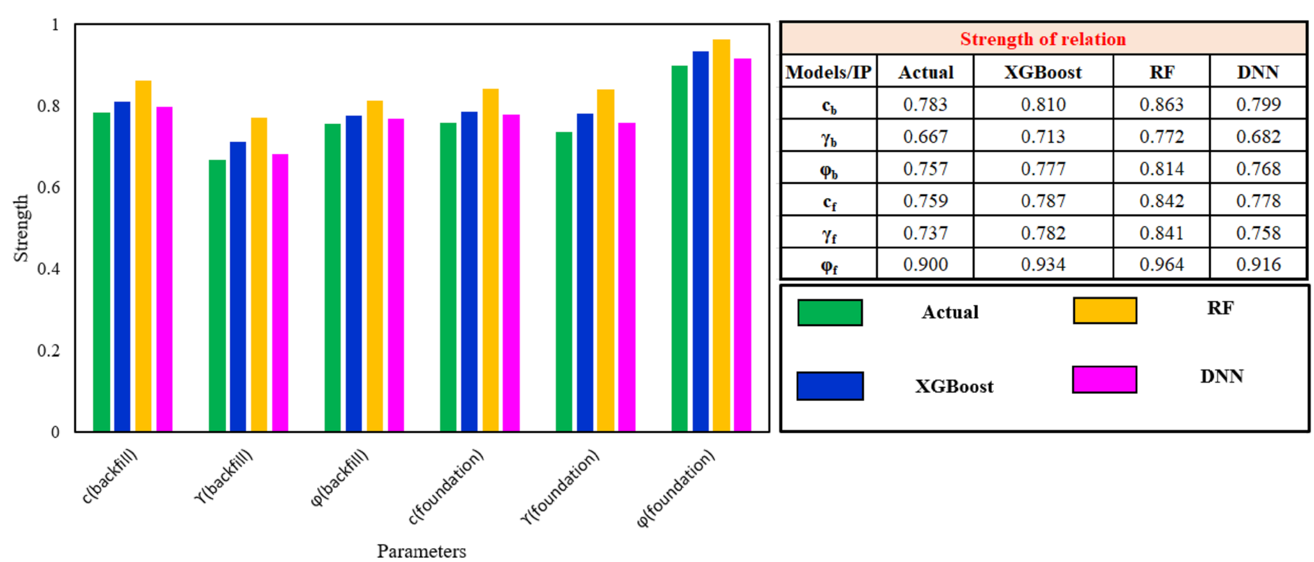
The results of M-W test for all three models are shown in Table 13. It is concluded that all the models have homogeneities due to lesser deviation is perceived when z-stat and asymptotic significance (i.e. 2-t: P values) are taken as a source. At 5% significance level ($\alpha = 0.05$), z –stat is less than 1.96 specify that there is no substantial dissimilarity between the observed and predicted values. However, the obtained value nearer to ideal value (i.e. $z_{0.025} = 1.96$) shows a more reliable model. Hence, DNN (z-stat = −0.0171) is a better performing model in predicting the FOS against bearing failure.

## Sensitivity analysis

To know the influence of each input parameters on the output, sensitivity study is performed. For this extensively used method i.e. cosine amplitude method (CAM) was implemented for the study as per Asteris and Mokos (2020) and Kardani et al. (2021). The strength of relation ($S_{O_{n,m}}$) of input parameters ($c_b$, $\varphi_b$, $\gamma_b$, $c_f$, $\varphi_f$ and $\gamma_f$) in predicting the FOS against bearing failure for the actual data and used model can be computed as:

$$S_{o_{n,m}} = \frac{\sum_{i=1}^{j} \mathrm{I}_{m,i} O_{n,i}}{\sqrt{\sum_{i=1}^{j} (\mathrm{I}_{m,i})^2 \sum_{i=1}^{j} (O_{n,i})^2}} \tag{65}$$

where, $I_{m,\,i}$ signifies the ith value of *mth* independent variable; *j* and *m* are the total observations and total input parameters, respectively; $O_{n,\,i}$ signifies the ith value of nth dependent variable; $S_{O_{n,m}}$ is the strength of relation of mth independent variable to nth dependent variable; and n is the total dependent variables. In this study, m = 6, n = 1 and j = 100. The strength of relation of different input parameters are shown in Fig. 14. The figure shows that the $\varphi_f$ is the utmost influential parameter for computing factor of safety against bearing failure followed by $c_b$, $c_f$, $\gamma_f$ and $\gamma_b$ for all the cases. It can also be observed that DNN nearly modelled the actual output in predicting the factor of safety.




| Strength of relation | | | | |
|---|---|---|---|---|
| Models/IP | Actual | XGBoost | RF | DNN |
| $c_b$ | 0.783 | 0.810 | 0.863 | 0.799 |
| $\gamma_b$ | 0.667 | 0.713 | 0.772 | 0.682 |
| $\varphi_b$ | 0.757 | 0.777 | 0.814 | 0.768 |
| $c_f$ | 0.759 | 0.787 | 0.842 | 0.778 |
| $\gamma_f$ | 0.737 | 0.782 | 0.841 | 0.758 |
| $\varphi_f$ | 0.900 | 0.934 | 0.964 | 0.916 |

**Fig. 14** Bar plots of sensitivity study for the developed model

## Conclusions

It is key to specify that a reliable and accurate valuation of gravity retaining wall failure against bearing can reduce resources used up in the creation and nurture of civil engineering structures. In this study, three machine learning (ML) models were used for predicting *FOS* against bearing failure. These ML models namely XGBoost, RF, and DNN. The main idea of this study is to recommend a quick and perfect ML model for the prediction of *FOS* against bearing failure. Different ML models were established and examined for their speed and based on their results, the following inferences are pointed out:

1. DNN model is the most capable and robust in predicting FOS against bearing failure among three models. This is because of higher value of $R^2$, *VAF*, *WI*, *LMI*, *KGE*, and *a-20 index* and lower value of *MAD*, *MSE*, *RMSE*, *MBE*, *MAE*, $U_{95}$ and *SI* among all the three used ML models in the training and testing stage.
2. The presentation of the models have also been judged by rank analysis, performance curve, and by computing reliability index of the models. Obtained results shown that DNN outperforms the others.
3. The uncertainty analysis and statistical testing (t-test, A-D test and M-W test) were also done to evaluate the reliability of the proposed DNN model in predicting the factor of safety against bearing failure.
4. ML model is effectively applied for this study to predict FOS against bearing failure. It seems to be an accurate and computationally proficient due to its simple execution technique, greater prediction validity, very low computational cost, and being representational.
5. Despite the numerous advantages of ML model, it can be improved in the future as follows: (i) using huge datasets to more precisely predict the desired output(s); (ii) through evaluation of the testing dataset's outcomes using cross-validation of many traditional machine learning models; (iii) using data endorsed by experts to address considerably wider diversities; (iv) extending the use of proposed models in many more fields; and (v) more input parameters amplified the computational time unusually due to additional number of rules.

**Acknowledgements** Not applicable.

**Author contributions** RM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing-original draft, Writing-review & editing; PS: Data curation, Software, Supervision, Validation; SK: Supervision, Validation; ETM: Validation; RMB: Validation.

**Funding** No funding was received for conducting this study.

## Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

**Availability of data and material** Data presented in the paper are available with authors.

## References

Alghaffar, M., & Wellington, C. (2005). Reliability analysis of retaining walls designed to British and European standards. *Structure and Infrastructure Engineering, 1*, 271–284.

Ali, Z. H., & Burhan, A. M. (2023). Hybrid machine learning approach for construction cost estimation: An evaluation of extreme gradient boosting model. *Asian Journal of Civil Engineering*. https://doi.org/10.1007/s42107-023-00651-z

Asteris, P.G., & Mokos, V.G. (2020). Concrete compressive strength using artificial neural networks. *Neural Comput. Appl.*, 32, 11807–11826.Babu, G.L.S., Srivastava, A. (2007). Reliability analysis of allowable pressure on shallow foundation using response surface method. *Comput. Geotech*, 34(3), 187–194.

Basha, B. M., & Babu, G. L. S. (2008). Optimum design of cantilever sheet pile walls in sandy soils using inverse reliability approach. *Computers and Geotechnics, 35*, 134–143.

Benuwa, B. B., Zhan, Y., Ghansah, B., Wornyo, D. K., & Kataka, F. B. (2016). A review of deep machine learning. *International Journal of Engineering Research in Africa, 24*, 124–136.

Bowles, J.E. (1997). Foundation Analysis and Design. *Fifth Edition, The McGraw Hill Companies, Inc.*

Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *22nd ACM SIGKDD Int. Conf. Knowl.,* 785–794. https://doi.org/10.1145/2939672.2939785

Chen, H., Asteris, P., & Armaghani, D. J. (2019). Assessing dynamic conditions of the retaining wall: Developing two hybrid intelligent models. *Applied Sciences*. https://doi.org/10.3390/app9061042

Chouksey, S. K., & Fale, A. (2017). Reliability analysis of counterfort retaining wall. *International Journal of Civil Engineering and Technology, 8*, 1058–1073.

Christian, J. T. (2004). Geotechnical engineering reliability: How well do we know what we are doing? *Journal of Geotechnical Geoenvironmental Engineering, 130*, 985–1003.

Cornell, C. (1969). *A Probability-Based Structural Code. Inj. Proc., 66*, 974–985.

Dao-Bing, Z., Zhi-bin, S., & Chuan-qu, Z. (2013). Reliability analysis of retaining walls with multiple failure modes. *Journal of Centeral South University., 20*, 2879–2886. https://doi.org/10.1007/s11771-013-1809-z

Das, B. M. (1998). *Principles of Foundation Engineering* (4th ed.). Brooks/Cole Publishing Company.

Goh, A. T. C., & Kulhawy, F. H. (2005). Reliability assessment of serviceability performance of braced retaining walls using a neural network approach. *International Journal for Numerical and Analytical Methods in Geomechanics, 29*, 627–642.

Goh, A. T., Phoon, K. K., & Kulhawy, F. H. (2009). Reliability analysis of partial safety factor design method for cantilever retaining walls in granular soils. *J. Geotech. Geoenviron. Eng., 135*, 616–622.

GuhaRay, A., Mondal, S., & Mohiuddin, H. H. (2018). Reliability analysis of retaining wall subjected to blast loading by finite element approach. *J. Inst. Eng. India Ser. A, 99*, 95–102. https://doi.org/10.1007/s40030-018-0273-z

Hashmi, A. F., Ayaz, M., Bilal, A., et al. (2023). GA-based hybrid ANN optimization approach for the prediction of compressive strength of high-volume fly ash concrete mixes. *Asian J Civ Eng, 24*, 1115–1128. https://doi.org/10.1007/s42107-022-00557-2

Hasofer, A. M., & Lind, N. C. (1974). Exact and invariant second-moment code format. *Journal of the Engineering Mechanics Division, 100*, 111–121.

IS: 6403 (1981). Code of practice for determination of bearing capacity of shallow foundations. *Bureau of Indian Standards, New Delhi, India*.

Jena, R., Pradhan, B., Beydoun, G., Nizamuddin, A., Sofyan, H., & Affan, M. (2019). Integrated model for earthquake risk assessment using neural network and analytic hierarchy process: Aceh province. *Indonesia. Geosci. Front., 11*(2), 613–634.

Jiang, S., Xiao, R., Wang, L., Luo, X., Huang, C., Wang, J. H., Chin, K. S., & Nie, X. (2019). Combining Deep Neural Networks and Classical Time Series Regression Models for Forecasting Patient Flows in Hong Kong. *IEEE Access, 7*, 118965–118974.

Kardani, N., Bardhan, A., Samui, P., Nazem, M., Zhou, A., & Armaghani, D. J. (2021). A novel technique based on the improved firefly algorithm coupled with extreme learning machine (ELM-IFF) for predicting the thermal conductivity of soil. *Engineering Computations*. https://doi.org/10.1007/s00366-021-01329-3

Kaveh, A., Kalateh-Ahni, M., & Fahimi-Farzam, M. (2013). Constructability optimal design of reinforced concrete retaining walls using a multi-objective genetic algorithm. *Structural Engineering and Mechanics (Techno Press)*, 2(47), 227–245. https://doi.org/10.12989/SEM.2013.47.2.227

Kaveh, A., Gholipour, Y., & Rahami, H. (2008). Optimal design of transmission towers using genetic algorithm and neural networks. *International Journal of Space Structures, 23*(1), 1–19.

Kaveh, A., & Iranmanesh, A. (1998). Comparative study of backpropagation and improved counterpropagation neural nets in structural analysis and optimization. *International Journal of Space Structures, 13*, 177–185.

Kaveh, A., & Jafarpour Laien, D. (2017). Optimal design of reinforced concrete cantilever retaining walls using CBO, ECBO and VPS algorithms. *Asian Journal of Civil Engineering, 18*(4), 657–671.

Kaveh, A., & Khalegi, A. (1998). Prediction of strength for concrete specimens using artificial neural network. *Asian Journal of Civil Engineering, 2*(2), 1–13.

Kaveh, A., & Khavaninzadeh, N. (2023). Efficient training of two ANNs using four meta-heuristic algorithms for predicting the FRP strength. *In Structures, 52*, 256–272.

Kaveh, A., & Soleimani, N. (2015). CBO and DPSO for optimum design of reinforced concrete cantilever retaining walls. *Asian Journal of Civil Engineering, 6*(16), 751–774.

Kumar, A., & Roy, P. (2017). Reliability analysis of retaining wall using imprecise probability. *International Conf. on structural safety and Reliability*, 288–297.

Kumar, M., Bardhan, A., Samui, P., Hu, J. W., & Kaloop, M. R. (2021). Reliability Analysis of Pile Foundation Using Soft Computing Techniques: A Comparative Study. *Processes, 9*, 486. https://doi.org/10.3390/pr9030486

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, 2(3), 18–22. http://CRAN.R-project.org/doc/Rnews/

Low, B. K., Zhang, J., & Tang, W. H. (2011). Efficient system reliability analysis illustrated for a retaining wall and a soil slope. *Computers and Geotechnics, 38*, 196–204.

Mann, H.B., & Whitney, D.R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 50–60.

Menon, D., & Mangalathu, S. (2011). Reliability analysis and design of cantilever RC retaining walls against sliding failure. *International Journal of Geotechnical Engineering, 5*, 131–141.

Mishra, P., Samui, P., & Mahmoudi, E. (2021). Probabilistic Design of Retaining Wall Using Machine Learning Methods. *Applied Science*, 2–14.

Murthy, V. N. S. (2003). *Geotechnical Engineering: Principles and Practices of Soil Mechanics and Foundation Engineering*. Marcel Dekker Inc.

Mustafa, R., Samui, P., Kumari, S. (2023). Seismic Performance of Gravity Retaining Walls Under Quasi-static Approach Using Probabilistic Analysis. *Transp. Infrastruct. Geotech*. https://doi.org/10.1007/s40515-023-00298-y

Pradeep, T., Bardhan, A., Burman, A., & Samui, P. (2021). Rock Strain Prediction Using Deep Neural Network and Hybrid Models of ANFIS and Meta-Heuristic Optimization Algorithms. *Infrastructures, 6*(9), 129. https://doi.org/10.3390/infrastructures6090129

Ray, R., Kumar, D., Samui, P., Goh, A. T. C., & Zhang, W. (2021). Application of soft computing techniques for shallow foundation reliability in geotechnical engineering. *Geoscience Frontiers, 12*, 375–383.

Razali, N. M., & Wah, Y. B. (2011). Power comparison of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics, 2*(1), 21–33.

Salmi, Z. J., Khodakarami, M. I., & Behnamfar, F. (2023). Development of seismic fragility curves for RC/MR frames using machine learning methods. *Asian J Civ Eng, 24*, 823–836. https://doi.org/10.1007/s42107-022-00533-w

Shahin, M.A., Maier, H.R., Jaksa, M.B. (2003). Neural and neuro-fuzzy techniques applied to modelling settlement of shallow foundations on granular soils. Int. *Congress on Modelling and Simulation*, 1886–1891.

Terzaghi, K., Peck, R.B., & Mesri, G. (1996). Soil Mechanics in Engineering Practice. *Third Edition, A Willey-Interscience Publication.*

Wang, H., Chen, H., Wang, Y., & Han, L. (2020a). Reliability analysis for stability of the gravity retaining wall under mountain torrent. *System Science & Control Engineering, 8*, 434–440.

Wang, L., Wu, C., Tang, L., Zhang, W., Lacasse, S., Liu, H., & Goh, L. (2020b). Efficient reliability analysis of earth dam slope stability using extreme gradient boosting method. *Acta Geotechnica, 15*, 3135–3150.

Wu, C., Hong, L., Wang, L., Zhang, R., Samui, P., & Zhang, W. (2022). Prediction of wall deflection induced by braced excavation in spatially variable soils via convolutional neural network. *Gondwana Research*. https://doi.org/10.1016/j.gr.2022.06.011

Xiang, Y., Goh, A. T. C., Zhang, W., & Runhong, Z. (2018). A multivariate adaptive regression splines model for estimation of maximum wall deflections induced by braced excavation in clays. *Geomechanics and Engineering, 14*, 315–324.

Xiao, Z., Huang, J., Wang, Y., & Xia, H. (2014). Random Reliability Analysis of Gravity Retaining Wall Structural System. *International Conference on Mechanics and Civil Engineering*, 199–204. https://dx.doi.org/https://doi.org/10.2991/icmce-14.2014.36

Yong, W., Zhang, W., Nguyen, H., Bui, X., Choi, Y., Nguyen, T., Zhou, J., & Tran, T. T. (2022). Analysis and prediction of diaphragm wall deflection induced by deep braced excavations using finite element method and artificial neural network optimized by metaheuristic algorithms. *Reliability Engineering and System Safety, 221*, 108335.

Zevgolis, I.E., & Bourdeau, P.L. (2006). System Reliability of Cantilever Retaining Wall with Correlated Failure Modes. *In GeoCongress: Geotechnical Engineering in the Information Technology Age*, 1–6. https://doi.org/10.1061/40803

Zhang, B., Zhang, Z., & Zheng, P. (2022). Reliability analysis of gravity retaining wall. *Proceedings of the 8th International Conference on Civil Engineering*. https://doi.org/10.1007/978-981-19-1260-3_44

Zhang, W. G., & Goh, A. T. C. (2013). Multivariate adaptive regression splines for analysis of geotechnical engineering systems. *Computers and Geotechnics, 48*, 82–95.

Zhang, W., Zhang, R., & Goh, A. T. C. (2017a). Multivariate adaptive regression splines approach to estimate lateral wall deflection profiles caused by braced excavations in clays. *Geotechnical and Geological Engineering, 36*, 1349–1363.

Zhang, W. G., Zhang, Y. M., & Goh, A. T. C. (2017b). Multivariate adaptive regression splines for inverse analysis of soil and wall properties in braced excavation. *Tunneling and Underground Space Technology, 64*, 24–33.

Zhou, G. M., Li, Y., & Zhang, F. (2014). Analysis of Reliability Calculation and System Analysis of Gravity Retaining Walls. *Applied Mechanics and Materials, 556–562*, 862–866. https://doi.org/10.4028/www.scientific.net/amm.556-562.862

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.