Engineered Science
DOI: https://dx.doi.org/10.30919/es897

Soft Computing Techniques for Predicting Penetration and Uplift Resistances of Dual Pipelines in Cohesive Soils

Divesh Ranjan Kumar,[1] Pijush Samui,[1] Warit Wipulanusat,[2,*] Suraparb Keawsawasvong,[2] Kongtawan Sangjinda[2] and Wittaya Jitchaijaroen[2]

Abstract

It is a difficult task for practical engineers to calculate the uplift and penetration resistance of two overlapping pipelines that are buried in clay that increases in strength linearly. Hence, in this paper, four regression models, namely long short-term memory (LSTM), bidirectional long short-term memory (BI-LSTM), emotional neural network (ENN), and multivariate adaptive regression spline (MARS) models, are employed to create a data-driven prediction for the uplift and penetration resistance of two overlapping pipelines. For this purpose, a total of 256 samples of uplift conditions and 384 samples of penetration conditions, including three input parameters and one output parameter, are collected from the lower and upper bound finite element limit analysis (FELA) solutions. The predictive strength and robustness of the employed model were evaluated based on various performance metrics, rank analysis, error matrix, Taylor diagram and uncertainty analysis. Sensitivity analysis is also performed to determine the most and least effective parameters. Additionally, the results of the sensitivity analysis indicated that the pipe embedded depth ratio (w/D) was the most significant parameter in both uplift and penetration conditions. The MARS model produces more efficient performance (R^2=0.999 and $RMSE$=0.008 for uplift and R^2=0.999 and $RMSE$=0.009 for penetration condition) for uplift and penetration resistance prediction compared to the BI-LSTM, LSTM, and ENN models. The acquired findings demonstrated that the MARS model predicted the normalized uplift and penetration resistance with reasonable accuracy and yielded superior performance compared to the Bi-LSTM, LSTM, and ENN models. Therefore, it becomes one of the predictive tools practical engineers use in making preliminary decisions about things such as the uplift and penetration resistance of two overlapping pipelines buried in clay, which increases in strength linearly. It also provides a mathematical formulation for easy hand calculations.

Keywords: Penetration Resistance; Uplift Resistance; Pipeline; MARS; LSTM; Bi-LSTM; ENN.
Received: 05 April 2023; Revised: 18 May 2023; Accepted: 18 May 2023.
Article type: Research article.

1. Introduction

One offshore engineering issue is the pipeline problem. Marine pipelines are commonly constructed in a trench with a shallow embedment and are utilized to transport natural gas and oil from sources to manufacturing facilities. To ensure the safety of these pipelines, the penetration and uplift resistances are the essential issues to be predicted in the design process. Classic studies of the uplift capacity of pipelines were presented by Trautmann and O'Rourke, Trautmann *et al.*,[1] and Dickin[2] using experiments and analytical schemes. Later, some researchers also investigated the same pipeline problem by conducting laboratory and centrifuge model tests, such as Ling *et al.*,[3] Calvetti *et al.*,[4] Cheuk *et al.*,[5] and White *et al.*.[6] The displacement-based finite element method (FEM), which is one of the favorite numerical techniques for geotechnical stability problems, was also carried out by some researchers to investigate the penetration and uplift capacity of pipelines,[7] such as Guo *et al.*,[8,9] Chatterjee *et al.*,[10–12] Jung *et al.*,[13,14] and Maitra *et al.*.[15] Another favorite numerical technique for geotechnical stability problems is finite element limit analysis (FELA). This FELA consists of lower bound (LB) and upper bound (UB) methods, which can be used to provide LB and UB solutions for bracketing true solutions.[16] Using FELA, Martin and White[17] provided solutions for the penetration and

---

[1] *Department of Civil Engineering, National Institute of Technology Patna.*
[2] *Department of Civil Engineering, Faculty of Engineering, Thammasat School of Engineering, Thammasat University, Pathumthani, Thailand.*
**Email:* wwarit@engr.tu.ac.th (W. Wipulanusat)

uplift resistances of pipelines in cohesive soils. Consequently, Chakraborty and Kumar[18–20] and Chakraborty[21] carried out FELA solutions of vertical and horizontal resistances of pipelines in sand. By considering the effect of a sloping ground near a pipeline, Bhattacharya and Sahoo[22] also employed the FELA technique to investigate the stability of pipelines in slopes. In addition, Shiau *et al.*[23–26] presented FELA solutions for problems of pipeline burst-related ground stability by considering both collapse and blowout conditions. Recently, Seehavong and Keawsawasvong[27] carried out FELA solutions of two interfering pipelines in cohesive soils subjected to uplift and penetration resistances. In this study, the previous results by Seehavong and Keawsawasvong[27] were employed to conduct the machine learning models. To analyze the uplift and penetration resistances of two interfering pipelines in cohesive soils, the LSTM, BI-LSTM, ENN, and MARS models have not yet been applied, as per the findings of the authors' literature review; meanwhile, these machine learning models have been widely used to solve complex and non-linear civil engineering problems. Najafzadeh *et al.*[28] proposed a neuro-fuzzy based group method of data handling (NF-GMDH) based self-organized model to predict bridge pier score depth under debris flow effects. Furthermore, Najafzadeh & Movahed[29] proposed a gene-expression programming (GEP) based GMDH model to predict the free span expansion rates below pipelines under waves. Later, Movahed *et al.*[30] proposed a hybrid GMDH-ELM model to predict the longitudinal dispersion coefficients in water pipelines. Najafzadeh used various machine learning models including multivariate adaptive regression spline (MARS) to evaluate the river water quality index.

The numerical results were collected from a previous study by Seehavong and Keawsawasvong[27] as training and testing data. Then, four advanced regression machine learning models, namely long short-term memory (LSTM), bidirectional long short-term memory (BI-LSTM), emotional neural network (ENN), and multivariate adaptive regression spline (MARS), are utilized to predict the normalized uplift and penetration resistance of two overlapping pipelines buried in clay After performing four machine learning models, the performance and effectiveness of the used models have been analyzed using traditional and modern performance evaluation metrics, rank analysis, and error matrix. The generalization capability of the proposed models has been analyzed using uncertainty analysis. Most significantly, the models have been proposed into an empirical design equation so that they may be used easily by practical practitioners for predicting the normalized uplift and penetration resistance. The highlights of these proposed advanced ML algorithms are as follows: The proposed models are a useful tool for geotechnical engineers since they are straightforward to use and require a basic understanding of computers to run the software and obtain accurate results in a short amount of time. They consider the smallest RMSE error in both the training and testing phases, and their computational cost and generalization ability are rewarding.

2. Data collection

Figures 1(a) and (b) depict the problem definition for two interfering pipelines buried in clay under uplift (F_t) and penetration (F_c) forces, respectively. Two pipelines have the same diameter (D), where the spacing (center-to-center) between them is denoted by S, and the embedded depth is defined by w. The clay has a linear strength gradient (r) and unit weight (g). Notably, the shear strength at the ground surface is set to zero. The adhesion factor at the contact of pipelines and soils is set to 2/3. More details on the problem can be found in Seehavong and Keawsawasvong.[27]

Based on Seehavong and Keawsawasvong,[27] the normalized uplift and penetration resistances of two interfering pipelines can each be expressed in terms of three dimensionless parameters as follows Eq. (1):

$$\frac{q_t}{\rho D}\ and\ \frac{q_c}{\rho D} \propto f\left(\frac{S}{D},\frac{w}{D},\frac{\gamma}{\rho}\right) \tag{1}$$

where the vertical uplift pressure q_t and the vertical penetration pressure q_c can be calculated from F_t/D and F_c/D, respectively. In addition, S/D is the spacing ratio; w/D is the embedded ratio;

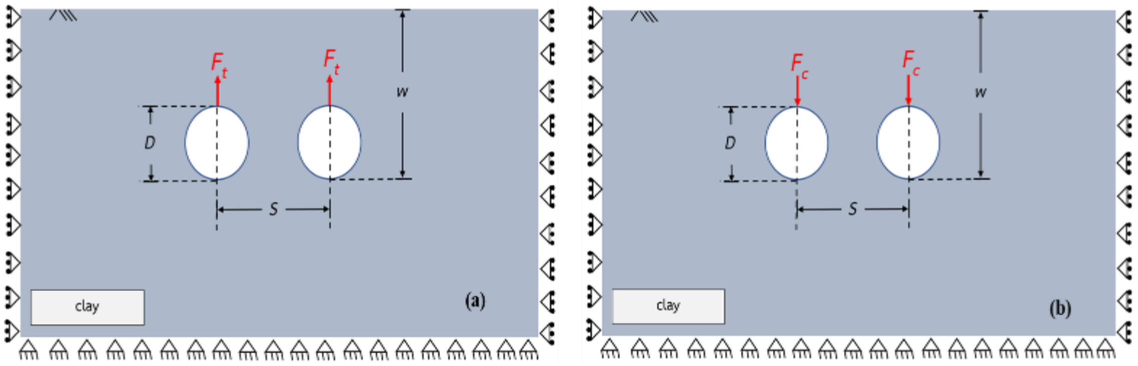


Fig. 1 Problem description of vertical forces of two interfering pipelines in clay: (a) uplift condition and (b) penetration condition.

and γ/ρ is the normalized unit weight and increasing strength. The selected ranges of these three dimensionless inputs are shown in Table 1, according to Seehavong and Keawsawasvong.[27]

Table 1. List of parametric values used in this study.

Parameters	Selected values
S/D	1.25, 1.5, 1.75, 2, 2.5, 3, 3.5, 4, 5, 6, 8, 10
w/D	1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3
g/r	0, 1, 3, 5

The FELA techniques were employed by Seehavong and Keawsawasvong[27] to determine the LB and UB solutions of q_t/rD and q_c/rD in the instances of uplift and penetration resistances, respectively. Examples of both uplift and penetration cases simulated in the FELA called OptumG2 (Krabbenhoft *et al.*)[31] are presented in Figs. 2(a) and 2(b), respectively. Note that these two models in Fig. 2 have developed only half of the model due to the problem symmetry. To increase the correctness of LB and UB solutions, an automatically adaptable mesh refinement technique Ciria *et al.*,[32] was also employed by Seehavong and Keawsawasvong[27] by setting the initial meshes to 5,000 elements and the final meshes after five adaptive meshing iterations to 10,000 elements (*e.g.*, Keawsawasvong and Ukritchon, Keawsawasvong *et al.*, Yodsomjai *et al.*,).[33–38] Examples of the final adaptive meshes are shown in Figs. 3(a) and 3(b) in the instances of uplift and penetration resistances, respectively. The potential slip surfaces represented by the absolute velocity contours are also shown in Figs. 4(a) and 4(b) in the instances of uplift and penetration resistances, respectively, which proves that the interference effect on the two pipelines exists and results in the capacity of both pipelines.

2.1 Statistical analysis of the dataset

In this analysis, the statistical description describes the characteristics of the dataset used for uplift and penetration conditions presented in Tables 2 and 3, respectively. There are two main groups of measurements that make up descriptive statistics: (1) variability (or spread) measurements and (2) central tendency measures. Measures of variability (or spread) include range (*i.e.*, minimum and maximum of the variables), standard deviation, and variance. In contrast, measures of central tendency comprise mean, median, mode, skewness, kurtosis, and quartile. As presented in Tables 2 and 3, the values of skewness and kurtosis for S/D and g/r are higher than those of other parameters. Consequently, these parameters deviate less from the normal distribution than the other variables.

Table 2. Statistical description of the dataset for uplift conditions.

	Mean	Max	Min	Stdev	Standard Error	Skewness	kurtosis	Q25%	Q75%
S/D	2.438	4	1.250	0.927	0.058	0.363	-1.232	1.688	3.125
w/D	2.125	3	1.250	0.574	0.036	0	-1.239	1.688	2.563
γ/ρ	2.250	5	0.000	1.924	0.120	0.280	-1.431	0.750	3.500
$\frac{q_t}{\rho D}$	6.121	17.52	0.769	4.124	0.258	0.908	0.106	2.803	8.356

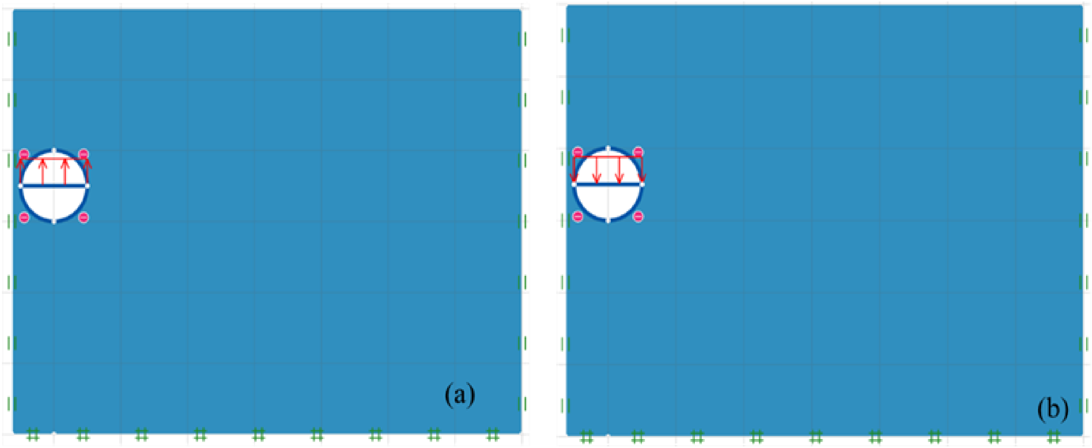


Fig. 2 Typical model geometries in OptumG2: (a) vertical uplift resistance and (b) vertical penetration resistance (S/D = 1.25, w/D = 3 and g/r = 1).

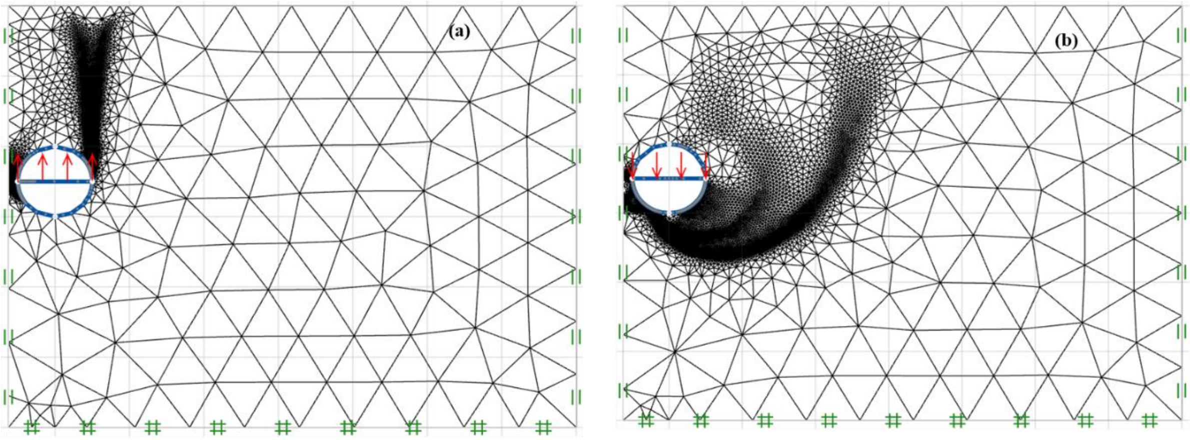


Fig. 3 Typical adaptive meshes: (a) Vertical uplift resistance and (b) Vertical penetration resistance (*S/D* = 1.25, *w/D* = 3 and *g/r* = 1).

Table 3. Statistical description of the dataset for penetration conditions.

	Mean	Max	Min	Stdev	Standard Error	Skewness	kurtosis	Q25%	Q75%
S/D	4.042	10	1.250	2.639	0.135	0.997	-0.091	1.938	5.250
w/D	2.125	3	1.250	0.574	0.029	0.000	-1.239	1.688	2.563
γ/ρ	2.250	5	0.000	1.923	0.098	0.279	-1.430	0.750	3.500
$\frac{q_c}{\rho D}$	17.579	31.846	7.151	6.143	0.313	0.336	-0.747	12.665	21.946

2.2 Data preprocessing

In data preprocessing, the whole dataset is normalized in the interval of [0, 1] using the min-max approach to confirm that no single feature's range had an outsized impact on the neuron input and improved the accuracy of the proposed model. As a result of the normalization, only the central tendencies of the neurons were affected, rather than the ranges of individual properties. The mathematical Eq. (2) was used for the normalization of the dataset.

$$X_N = \frac{X - \min(X)}{\max(X) - \min(X)} \tag{2}$$

where X_N and X represent the normalized and actual feature values, respectively.

In this study, a total of 256 samples were collected for calculation of uplift resistance, and 384 samples were gathered for penetration resistance calculations. The entire dataset is randomly divided into training data and testing data, with training data containing 70% of input data and testing data comprising 30% of input data, to obtain robust results by forcing the algorithms to operate completely independently of the specific training data for both the uplift and penetration cases. Training data are utilized to build the model, whereas testing data are employed to validate the model.

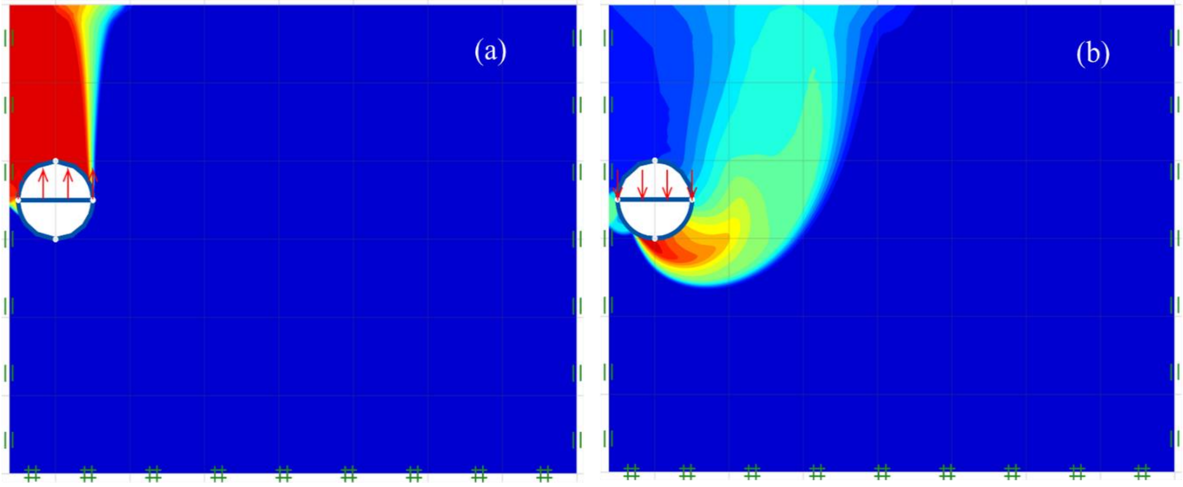


Fig. 4 Absolute velocity contours: (a) Vertical uplift resistance and (b) Vertical penetration resistance (*S/D* = 1.25, *w/D* = 3 and *g/r* = 1).

2.3 Performance and error metrics for evaluation

As performance metrics, the coefficient of determination (R^2), adjusted determination coefficient ($AdjR^2$), the Nash Sutcliffe efficiency (NS), the variance account factor (VAF), the performance index (PI), the Willmott`s index of agreement (*WI*), and error metrices the weighted mean absolute percentage error (*WMAPE*), the root mean square error (*RMSE*), the mean absolute error (*MAE*), and RMSE to observation's standard deviation ratio (*RSR*), were utilized to evaluate the predictive ability of the proposed models.[39–45] The coefficient of determination (R^2) measures how well the actual data agree with the predicted data. Thus, the value of R^2 corresponding to 1 indicates a perfect fit, while the prediction accuracy decreases with decreasing R^2 value. The adjusted determination coefficient ($AdjR^2$) measures the accuracy with which a model can predict a target outcome based on the provided input variable to construct the models. A $AdjR^2$ value of 1 indicates a perfect model. Nash–Sutcliffe efficiency (NS) measures the ratio of residual error variance to predicted variance in actual data. The ideal value of NS is 1, and its obtained value should not be less than 1. The variance account factor (*VAF*) is the percentage of sequence among the original and predicted values of the output parameters. The ideal value of *VAF* is 100% for the perfect model. The performance index (*PI*) is used to evaluate the performance of the proposed model. The ideal value of *PI* is 2, and for an accurate model, the value of *PI* should be greater than 1. Willmott's index of agreement (WI) measures the error associated with the predicted models. The value of WI ranges between 0 and 1. For a perfect model, the value of WI should be 1. Weighted mean absolute percentage error (WMAPE) statistically evaluates the proposed model's accuracy. Compared to the MAPE, which does not consider weighted error, this method is more accurate. The ideal value of WMAPE is 1. The RMSE to observation SD ratio (RSR) is calculated to check the statistical evaluation of the proposed model. The ideal value of RSR ranges from 0 to a large positive value, and a value higher than 0.7 is not desirable. The root mean square error (RMSE) is the most commonly utilized error metric used to evaluate errors associated with the proposed models. The ideal value of RMSE is 0. The mean absolute error (MAE) is used to check whether the model's predicted result is overestimated or underestimated. A value of MAE close to zero indicates perfect agreement. The mathematical expressions of performance and error metrices are presented in Eqs. (3) to Eq. (12).

$$R^2 = \frac{\sum_{i=1}^{n}\left(d_i - d_{avg}\right)^2 - \sum_{i=1}^{n}(d_i - y_i)^2}{\sum_{i=1}^{n}\left(d_i - d_{avg}\right)^2} \tag{3}$$

$$AdjR^2 = 1 - \frac{n-1}{n-p-1}(1 - R^2) \tag{4}$$

$$NS = 1 - \frac{\sum_{i=1}^{n}(d_i - y_i)^2}{\sum_{i=1}^{n}(d_i - d_{mean})^2} \tag{5}$$

$$VAF = \left(1 - \frac{var(d_i - y_i)}{var(d_i)}\right) \times 100 \tag{6}$$

$$PI = adj.R^2 + (0.01 \times VAF) - RMSE \tag{7}$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{n}(d_i - y_i)^2}{\sum_{i=1}^{n}\left\{\left|y_i - d_{avg}\right| + \left|d_i - d_{avg}\right|\right\}^2}\right] \tag{8}$$

$$WMAPE = \frac{\sum_{i=1}^{n}\left|\frac{d_i - y_i}{d_i}\right| \times d_i}{\sum_{i=1}^{n} d_i} \tag{9}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i - y_i)^2} \tag{10}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|(y_i - d_i)\right| \tag{11}$$

$$RSR = \frac{RMSE}{\sqrt{\frac{1}{n} * \sum_{i=1}^{n}\left(d_i - d_{avg}\right)^2}} \tag{12}$$

3. Theoretical background of the employed models

In this section, the theoretical background of the used models, namely, LSTM, Bi-LSTM, ENN, and MARS models, are described in detail to predict the normalized uplift and penetration resistance of two overlapping pipelines buried in clay, which increases in strength linearly. The methodology flow steps, including data collection, data preprocessing, proposed machine learning models, hyperparameter selection, and comparative analysis, are presented in Fig. 5.

3.1 Long Short-Term Memory (LSTM)

LSTM is a powerful machine learning technique proposed by Hochreiter and Schmidhuber. LSTM is an extended form of a recurrent neural network (RNN) in the form of a repeating block chain designed to address the issues of RNNs, such as learning long data sequences and vanishing and exploding gradients. The typical LSTM network architecture presented in Fig. 6 represents the cell state from C_{t-1} to C_t which contains four unique modules, three gates and one memory cell state. Information may be preserved in a steady state and continue to flow above the cell state because it acts like a conveyor belt that runs directly through the whole chain with just slight linear interactions. The three main gates of an LSTM network are the forget gate, the output gate, and the input gate. These gates regulate the flow of information into and out of each cell.

The input gate determines how much new information can enter the cell. The forget gate determines which information from the internal cell state should be deleted. Finally, the output gate determines how much of the cell's current memory should be used to generate the output. The mathematical

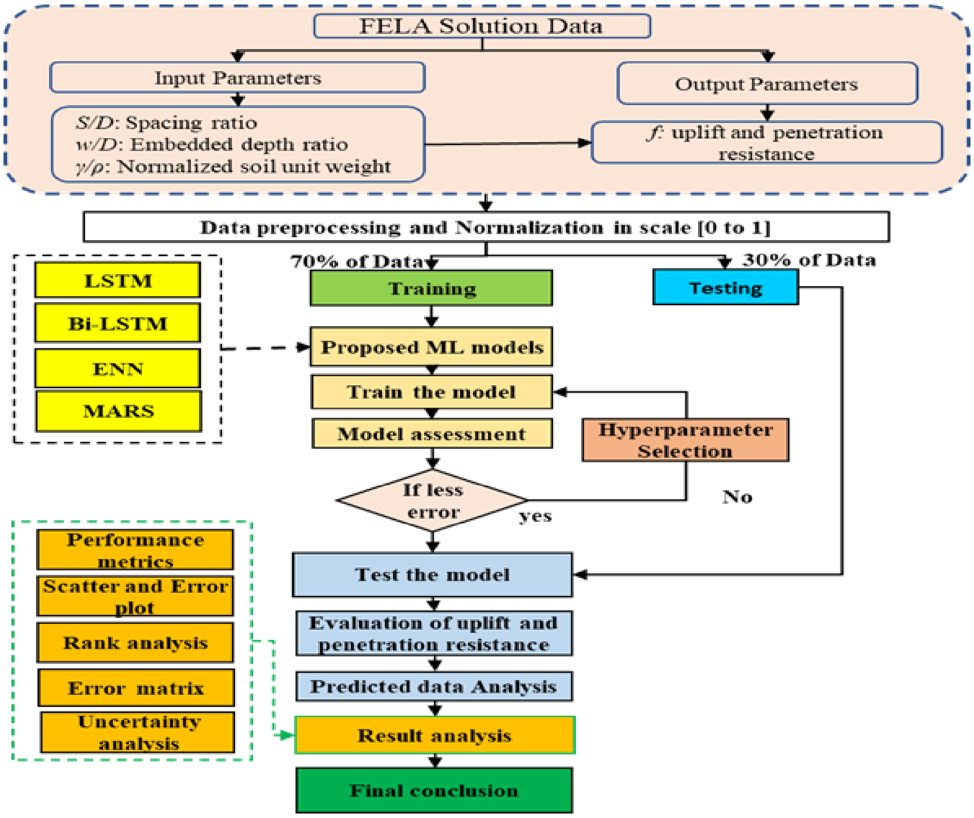


Fig. 5 Methodology flow chart.

formulations of LSTM recurrent units are presented as follows.

$$f(t) = \sigma(P_f[C_{t-1}, x_t] + b_f \quad (13)$$

$f(t)$ represents the output of the forget gate, and P_f and b_f represent the weight matrix and bias vector of the forget gates from the input layer. Thereafter, the new data are saved in the cell state using the following sigmoid layer of input gates, and then the tanh layer creates a vector to the cell state with a value of -1 to 1.

$$i(t) = \sigma(P_f[C_{t-1}, x_t] + b_i \quad (14)$$

$$\widetilde{C_t} = tanh(P_c[C_{t-1}, x_t] + b_c \quad (15)$$

The elementwise multiplication of the vector matrix C_{t-1} cell and f_t is used to update the old cell state to C_t, and then the term $A_t \otimes \widetilde{C_t}$ is added to remember the new information.

$$C_t = f_t \otimes C_{t-1} + A_t \otimes \widetilde{C_t} \quad (16)$$

$$O_t = \sigma(P_o \cdot [C_{t-1}, x_t] + b_o \quad (17)$$

$$h_t = O_t \otimes \tanh(C_t) \quad (18)$$

$$Sigmoid(x) = \frac{1}{1+e^{-N}} \quad (19)$$

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (20)$$

where x_t represents the input vector and b_o and b_i represent the bias vectors of the output gates and input gates, respectively. P_c and P_i represent the weight matrices of the output gates and input gates, respectively. $\otimes$ represents the elementwise multiplication of the vector matrix. C_t and h_t represent the past cell and its output vector of the hidden layer

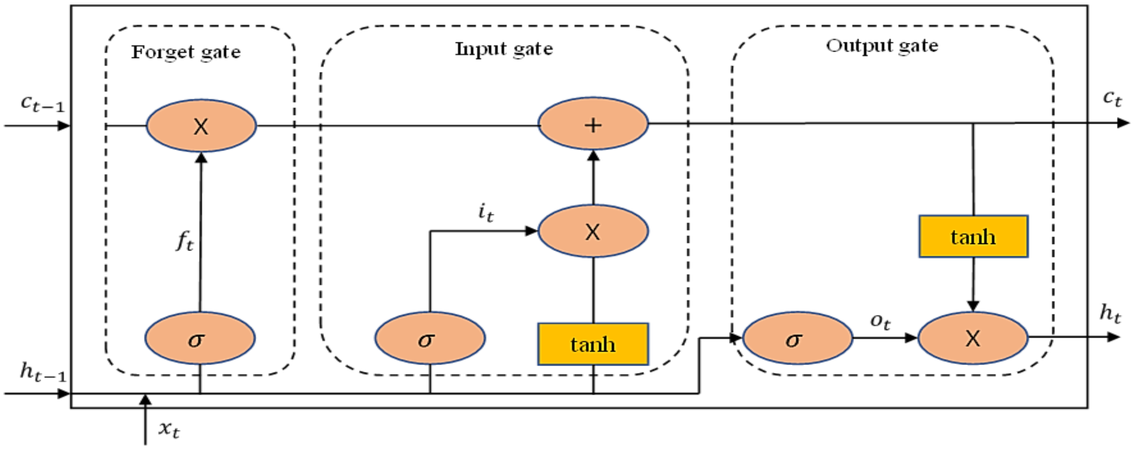


Fig. 6 Architecture of long-short term memory.

at the current moment, respectively. O_t represent the target of the output gate. By using these gates to regulate the flow of information, LSTMs are capable of learning and remembering long-term data dependencies. Overall, LSTMs are a powerful and flexible tool for dealing with complex engineering problems and are widely used in deep learning applications.

3.2 Bidirectional Long Short-Term Memory (Bi-LSTM)

Bidirectional long short-term memory (BI-LSTM) is a recurrent neural network (RNN) architecture that has acquired popularity in the past few years due to its capacity to capture both forward and backward dependencies in sequential data. The BLSTM architecture is an extension of Hochreiter and Schmidhuber's 1997 Long Short-Term Memory (LSTM) network.[46] BLSTM is based on the concept of processing input sequences in both forward and backward directions, enabling the network to capture information from both past and future contexts. In a conventional LSTM network, the output at each time step is computed based on the preceding input sequence. However, a BLSTM network calculates the output at each time step based on both the past and future contexts. The BLSTM network comprised two LSTM layers, one of which processes the input sequence in the forward direction and the other in the backward direction. Each layer has a hidden state vector, updated at each time step using the current input and the previous hidden state. The output of each layer is then concatenated to form the final output of the BLSTM network. The architecture of bidirectional long-short term memory is presented in Fig. 7. As shown in Fig. 7 after splicing two hidden layers together, Bi-LSTM calculates the final output $y_t$by combining the information from the forward hidden layer $h^{+(t)}$ and the backward hidden layer $h^{-(t)}$ at each time step using the following mathematical Eqs. (21-23).

$$y^t = W^{h^{+(t)}y} h^{+(t)} + W^{h^{-(t)}y} h^{-(t)} + b^y \tag{21}$$

$$h^{+(t)} = LSTM\left(x_t, h^{+(t-1)}\right) \tag{22}$$

$$h^{-(t)} = LSTM\left(x_t, h^{-(t-1)}\right) \tag{23}$$

where the terms $W^{h^{+(t)}y}$ and $W^{h^{-(t)}y}$ are the weight values of the forward and backward layers, respectively. $h^{+(t)}$ and $h^{-(t)}$ represent the forward and backward hidden layers respectively, and b^y represents the bias of the output layer.

BLSTM can capture long-term dependencies in the input sequence, which is one of its advantages over traditional LSTM. The network can collect information from both the past and the future by analyzing the sequence in both forward and backward directions. This is beneficial for speech recognition tasks, where the context of a particular phoneme may depend on both preceding and succeeding phonemes. BLSTM is a potent type of recurrent neural network that can detect long-term dependencies in input sequences by processing them in both forward and backward directions. It has become a popular choice for various applications in natural language processing, speech recognition, and image recognition due to its capacity to capture past and future contexts.

3.3 Emotional Neural Network (ENN)

An emotional neural network is a form of artificial neural network designed to simulate or model the emotional processes of humans or animals. These networks are typically based on the idea that emotions arise from the interactions between various neural circuits in the brain. They seek to replicate these interactions in a computational model. The goal of emotional neural networks is to better understand how

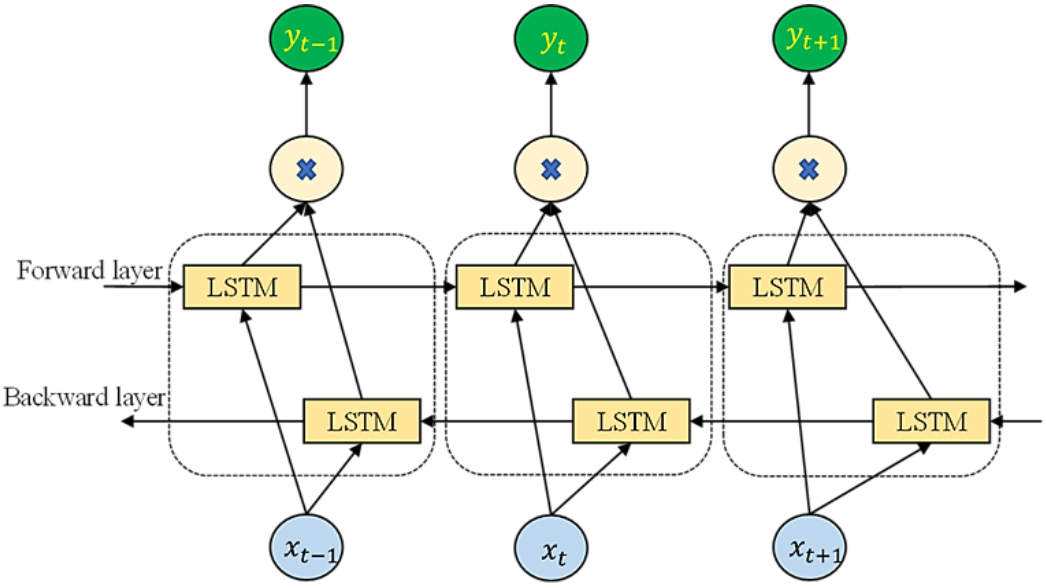


Fig. 7 Architecture of bidirectional long-short term memory.

emotions work, as well as to create more sophisticated and lifelike models of human or animal behavior. These networks can be used in a variety of applications, including artificial intelligence, robotics, and psychology research.[47,48]

Considering the input data path through the network, the ENN algorithm has three distinct layers: the input layer, the hidden layer, and the output layer. Each of the three layers of the ENN has its own set of neurons: input layers with *i* neurons, hidden layers with *h* neurons, and output layers with *j* neurons. A typical architecture of the ENN model is presented in Fig. 8. The processing of the input, output and hidden layers of the ENN are presented below.

(a) Input layer neurons

The neurons of the input layer are not processing any interaction, so for every input neuron, the output is defined as presented in Eq. (24):

$$Y_i = X_i \tag{24}$$

where X_i and Y_i represent the input value and output value of neuron *i* in the input layer, respectively.

(b) Hidden layer neurons

Since the neurons of the hidden layer are considered processing neurons, their activation is governed by a sigmoid activation function. In this case, we assume there is a single hidden layer, but the same procedure would apply regardless of the number of layers. The mathematical Eq. (25) for the output of each hidden neuron is presented as follows:

$$Y_h = \left(\frac{1}{1+\exp\left(-X_h\right)}\right) \tag{25}$$

where X_h represents the input value and Y_h represents the output value of neuron *h*, which represents the hidden layer values. The input value of hidden layer X_h is calculated using the summation of the byproduct of X_h and its respective weight value named the total potential (T_h).

(c) Output layer neuron

Finally, the neurons of the output layer are activated with the help of the sigmoid activation function using Eq. (26):

$$Y_j = \left(\frac{1}{1-\exp\left(-X_j\right)}\right) \tag{26}$$

where the terms X_j and Y_j represent the neuron values of the input and output layers, respectively.

3.4 Multivariate Adaptive Regression Splines (MARS)

Multivariate adaptive regression splines (MARS) is a statistical modeling technique that involves fitting piecewise linear regressions to a set of predictor variables, which are combined using a series of “knots” or breakpoints in the data. It was introduced by Friedman[50] as an extension of the univariate adaptive regression splines (MARS) algorithm. The general equation of nonlinear and nonparametric regression is presented in Eq. (27).

$$y_i = f\left(x_{i1}, x_{i2}, \ldots . x_{ik}\right) + e_i \tag{27}$$

where e_i represents the error involved in the prediction of output. $f\left(x_{i1}, x_{i2}, \ldots . x_{ik}\right)$ represents the continuous regression function.

The penetration and uplift resistances of two interfering pipelines for given input are predicted using the MARS model by the following mathematical Eq. (28):

$$y = a_o + \sum_{n=1}^{N} a_n B_n(X) \tag{28}$$

where the term a_o represents the constant term, $B_n(X)$ represents the basic function (BF), X represents the input variable, and a_n represents the coefficient of the basic function.

The MARS algorithm builds a model that is a sum of individual basis functions, where each basis function is a product of univariate functions. These univariate functions are called “spline functions” and have the form max (0, $X - c$) or max (0, $X - c$), which is presented as follows in Eq. (29) and (30):

$$b_q^-(X-c) = [-(X-c)]_+^q = \begin{cases} (c-X)^q & if\ X < c \\ 0 & otherwise \end{cases} \tag{29}$$

$$b_q^+(X-c) = [+(X-c)]_+^q = \begin{cases} (X-c)^q & if\ X > c \\ 0 & otherwise \end{cases} \tag{30}$$

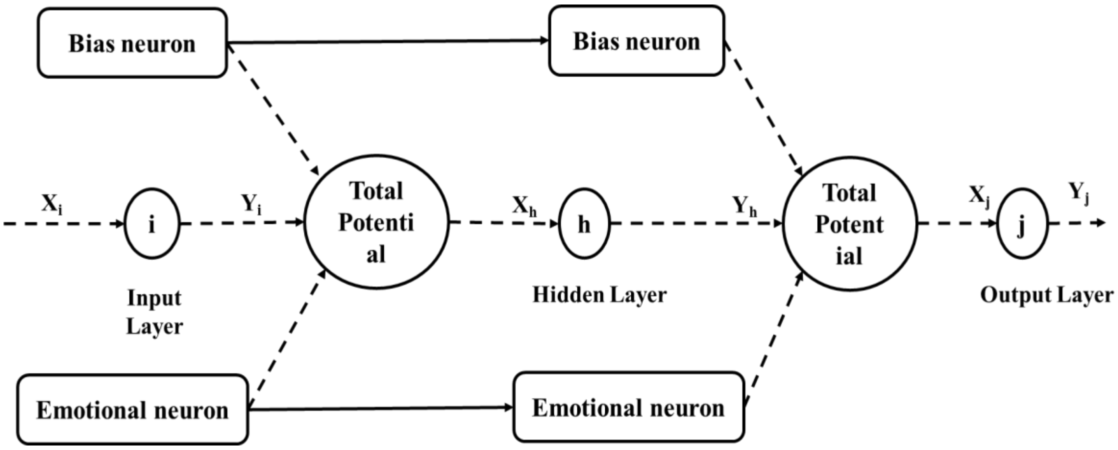


Fig. 8 A typical architecture of emotional neural network during feed forward calculation.[49]

where X is the predictor variable, c represents a "knot" or breakpoint, and $b_q^-(X-c)$ and $b_q^+(X-c)$ represent the left-side truncated and right-side truncated spline functions, respectively. Three basic steps involved in the process of constructing the multivariate adaptive regression spline (MARS) model.

(a) Constructing phase

The first step is also called the forward step or constructing phase. It can create an overfitted model with the help of the provided input parameters. In this step, BF plays an important role in building Eq. (28). The selection of BF in the constructing phase depends on the generalized cross validation (GCV).

(b) Backward/Pruning phase

The second step adopted a pruning algorithm to eliminate the least effective terms to attain the most accurate model based on the generalized cross-validation (GCV) value. The residual sum of squares of the input parameter value is assigned in the form of GCV. To limit the number of spline functions utilized in the model, the GCV parameter is employed as a complexity penalty. The mathematical equation used in the calculation of GCV is presented in Eq. (31) as follows:

$$GCV=\frac{RMSE}{\left[1-\frac{C(M)}{N}\right]^2} \tag{31}$$

where the penalty factor $C(M)$ is calculated as $C(M)=M+dM$, M represents the number of basic functions, d represents the penalty factor, and N represents the total number of data objects.

(c) Selection of the optimum MARS model

The cost penalty factor $C(M)$ represented in Eq. (31) is used to maximize each basic function. To minimize the overfitting of BF, they are eliminated during the backward phase. After completion of all the steps, the best MARS model is obtained. MARS is particularly useful when there is a nonlinear relationship between the predictor variables and the response variable and when a simple linear regression model does not easily capture the relationship. It is also useful when interactions between the predictor variables need to be accounted for. One advantage of MARS is that it can automatically identify the important predictor variables and the locations of the breakpoints in the data, which makes it useful for exploratory data analysis. It can also handle missing data and deal with continuous and categorical predictor variables. MARS has been applied to various applications, including finance, engineering, and medicine.[51] However, it has a few limitations, including the fact that it may overfit the data if too many basis functions are used and may not work well with high-dimensional data.

4. Results and discussion

4.1 Hyperparameter tuning

To construct the MARS model, MATLAB 2018a software was used. The prediction of the normalized uplift resistance MARS model started with 15 basic functions. After the pruning step, some basic functions are eliminated, and the best performing MARS model is obtained for twelve basic functions, presented in Table 4. The final normalized uplift resistance of two interfering pipelines buried in clay is calculated with the help of obtained correlation Eq. (32).

$$\frac{q_t}{\rho D}=0.381+0.534\times BF1-0.427\times BF2+0.235\times BF3-0.444\times BF4+0.581\times BF5-0.372\times BF6+0.0291\times BF7-0.412\times BF8-0.246\times BF9+0.971\times BF10-0.045\times BF11+0.513\times BF12 \tag{32}$$

Table 4. Equations of the basic functions used in the MARS model for uplift conditions.

SI. Number	Basic Function	Equation
1	BF1	max (0, w/D -0.857)
2	BF2	max (0, 0.857 - w/D)
3	BF3	max (0, γ/ρ -0.2)
4	BF4	max (0, 0.2 - γ/ρ)
5	BF5	BF3 × max (0, w/D -0.285)
6	BF6	BF3 × max (0, 0.285 - w/D)
7	BF7	max (0, S/D -0.181)
8	BF8	max (0, 0.181 - S/D)
9	BF9	BF4 × max (0, w/D -0.428)
10	BF10	BF4 × max (0, 0.428 - w/D)
11	BF11	BF2 × max (0, S/D -0.181)
12	BF12	BF2 × max (0, 0.181 - S/D)

On the other hand, for the prediction of normalized penetration resistance, the MARS model starts with 15 basic functions, and after the pruning step, some basic functions are eliminated. The best performing MARS model is obtained for thirteen basic functions, the equations of which are presented in Table 5. The final normalized penetration resistance of two interfering pipelines buried in clay is calculated with the help of the obtained correlation Eq. (33).

$$\frac{q_c}{\rho D}=0.405+0.930\times BF1-0.602\times BF2+0.195\times BF3-0.610\times BF4-0.209\times BF5-0.660\times BF6-0.074\times BF7+0.527\times BF8+0.075\times BF9+$$

$$0.352 \times BF10 - 0.168 \times BF11 - 0.173 \times BF12 - 0.921 \times BF13 \quad (33)$$

Table 5. Equations of the basic functions used in the MARS model for penetration conditions.

SI. Number	Basic Function	Equation
1	BF1	max (0, w/D -0.571)
2	BF2	max (0, 0.571 -w/D)
3	BF3	max (0, γ/ρ -0.2)
4	BF4	max (0, 0.2 - γ/ρ)
5	BF5	BF1 × max (0, γ/ρ -0.6)
6	BF6	BF1 × max (0, 0.6 - γ/ρ)
7	BF7	BF2 × max (0, γ/ρ -0.2)
8	BF8	BF2 × max (0, 0.2 -γ/ρ)
9	BF9	max (0, S/D -0.142)
10	BF10	max (0, 0.142 - S/D)
11	BF11	BF2 × max (0, S/D -0.314)
12	BF12	BF2 × max (0, 0.314 -S/D)
13	BF13	BF3 × max (0, 0.057 -S/D)

The ENN model selects the hidden layer's neurons as the processing neurons. Their activation is governed by a sigmoid activation function as presented in Eq. (13). The LSTM and Bi-LSTM achieved optimal hyperparameter values through trial and error by differing the hyperparameter value for the same training dataset for both the uplift and penetration resistance cases. The best configuration of hyperparameters to obtain accurate results for the LSTM and Bi-LSTM models are depicted in Table 6.

4.2 Performance evaluation

The most commonly used performance and error metrics stated previously have been used to evaluate the models used in this study. A detailed discussion of the obtained uplift and penetration results is presented below. The values of the calculated performance and error metrics for both uplift and penetration cases are presented in Tables 7 and 8.

Table 6. Hyperparametric configuration for LSTM and Bi-LSTM model.

Hyperparameters	LSTM	Bi-LSTM
Number of hidden layers	2	2
Batch Size	32	32
Activation function	ReLU	ReLU
Dense layer	64	64
Number of Epochs	500	500
Learning rate	0.0025	0.0025
Dropout rate	0.025	0.025
Loss function	mean_squared_error	mean_squared_error
optimizer	adam	adam

The R^2 value for MARS (0.999 for training and 0.998 for testing) attained the highest value, followed by the Bi-LSTM (0.995 for training and 0.997 for testing), LSTM (0.973 for training and 0.974 for testing), and ENN (0.963 for training and 0.951 for testing) models in the case of uplift, and in the case of penetration, the R^2 value for MARS (0.999 for training and 0.998 for testing) is again found to be more than that of BI-LSTM (0.998 for both training and testing), LSTM (0.988 for training and 0.989 for testing), and ENN (0.979 for training and 0.981 for testing). The $AdjR^2$ value obtained for MARS (0.999 for training and 0.998 for testing) is more followed by the Bi-LSTM (0.995 for training and 0.997 for testing), LSTM (0.973 for both training and testing), and ENN (0.962 for training and 0.949 for testing) models in the case of uplift

Table 7. Performance metrics of all the models for the uplift condition.

	LSTM		BILSTM		ENN		MARS		Ideal values
	TR	TS	TR	TS	TR	TS	TR	TS	
R^2	0.973	0.974	0.995	0.997	0.963	0.951	0.999	0.998	1
$Adj\ R^2$	0.973	0.973	0.995	0.997	0.962	0.949	0.999	0.998	1
$WMAPE$	0.095	0.108	0.036	0.038	0.127	0.151	0.019	0.026	0
NS	0.973	0.973	0.995	0.997	0.950	0.944	0.999	0.998	1
$RMSE$	0.040	0.039	0.018	0.014	0.055	0.056	0.008	0.010	0
VAF	97.31	97.43	99.49	99.66	95.30	94.77	99.89	99.82	100
PI	1.906	1.908	1.972	1.979	1.860	1.841	1.990	1.986	2
RSR	0.164	0.164	0.072	0.058	0.225	0.236	0.033	0.042	0
WI	0.993	0.993	0.999	0.999	0.986	0.985	1.000	1.000	1
MAE	0.032	0.031	0.012	0.011	0.043	0.043	0.006	0.007	0

Table 8. Performance metrics of all the models for the penetration condition.

	LSTM		BILSTM		ENN		MARS		Ideal Values
	TR	TS	TR	TS	TR	TS	TR	TS	
R^2	0.988	0.989	0.998	0.998	0.979	0.981	0.999	0.998	1
$Adj\ R^2$	0.988	0.989	0.998	0.997	0.979	0.980	0.999	0.998	1
WMAPE	0.050	0.047	0.019	0.020	0.066	0.060	0.017	0.017	0
NS	0.988	0.989	0.998	0.997	0.979	0.980	0.999	0.998	1
RMSE	0.027	0.026	0.011	0.013	0.036	0.035	0.009	0.010	0
VAF	98.806	98.90	99.81	99.74	97.91	98.05	99.86	99.84	100
PI	1.949	1.951	1.985	1.982	1.922	1.926	1.988	1.987	2
RSR	0.110	0.106	0.044	0.051	0.145	0.140	0.037	0.040	0
WI	0.997	0.997	1.000	0.999	0.995	0.995	1.000	1.000	1
MAE	0.021	0.021	0.008	0.009	0.027	0.027	0.007	0.008	0

and for the penetration case again MARS (0.999 for training and 0.998 for testing) achieves the highest $AdjR^2$ compared to Bi-LSTM (0.998 for training and 0.997 for testing), LSTM (0.988 for training and 0.989 for testing) and ENN (0.979 for training and 0.980 for testing) models.

The other performance parameters, such as NS, VAF, PI, and WI value for MARS, are more than Bi-LSTM, followed by LSTM and ENN model for both uplift and penetration conditions presented in Tables 7 and 8. Similarly, the error metrics WMAPE, RMSE, RSR, and MAE value for MARS is found to be less than Bi-LSTM, followed by LSTM and ENN model for both uplift and penetration conditions presented in Tables 7 and 8. From the performance metrics of all the models presented for both the uplift and penetration conditions (see Tables 3 and 4), it can be concluded that the MARS model is considered the best. The other models, such as Bi-LSTM, LSTM, and ENN, are in the second, third, and fourth positions, respectively.

4.3 Scatter and error plot

A scatter plot is a data visualization tool used in statistics and machine learning to display the relationship between actual and predicted output. A scatter plot helps to illustrate the relationship between the variables by drawing the actual and predicted data points along the line ($y = x$). The actual and predicted data points close to the line $y = x$ show better prediction. This study plots scatter and error plots for the training and testing phases for both the uplift and penetration conditions presented in Figs. 9 to 16. The scatter plots presented in Figs. 9 and 13 show the performance of the MARS model for uplift and penetration conditions, respectively, which conclude that the MARS model attained a high ability to predict uplift and penetration resistance followed by the Bi-LSTM, LSTM, and ENN models. The MARS model results exhibit good accuracy following these findings and less error. Therefore, the MARS model is a helpful uplift resistance prediction tool when considering the datasets used in this study. However, other models such as Bi-LSTM, LSTM, and ENN can also efficiently predict uplift and penetration resistance.

4.4 Error matrix

An error heatmap matrix is a graphical representation of the model's predicted error. It is a color-coded representation of the error of the used machine learning models, which helps visualize the model's accuracy across different categories.[43] The error matrix is created based on the ideal value of the performance metrics. This study uses ten performance metrics for error matrix development, namely, *R^2, Adj R^2, WMAPE, NS, RMSE, VAF, PI, RSR, WI, and MAE*. The error matrices for uplift and penetration conditions are presented separately in Figs. 17 and 18, respectively, where TR stands for training and TS stands for testing. As per the error matrix of the uplift and penetration condition, the MARS model predicted error is lower than the other model. However, the average error for all models for uplift condition is 4% and for penetration condition is 2%.

4.5 Rank analysis

Rank analysis can also be used in model selection, which involves choosing the best statistical model that fits a given dataset. In this context, rank analysis is often used to compare the performance of used models and to identify the best model based on its ability to predict the desired output. The score was allotted to each model based on the performance metric values for the training and testing phases of the uplift and penetration conditions separately. Finally, the total score is calculated by adding the training and testing scores, and rank is allotted based on the total score obtained for each model.[52] The mathematical Eq. (34) used for the total score calculation is

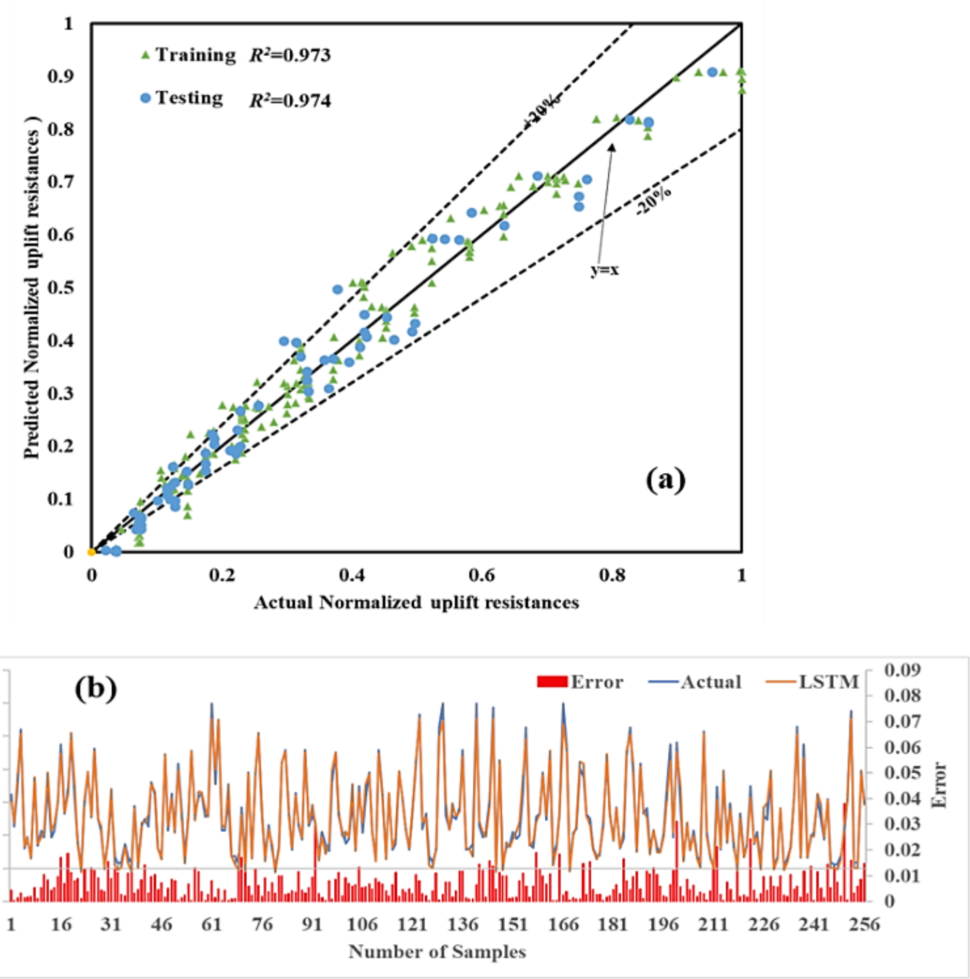


Fig. 9 Performance of LSTM model for uplift condition (a) Scatter plot (b) Error plot.

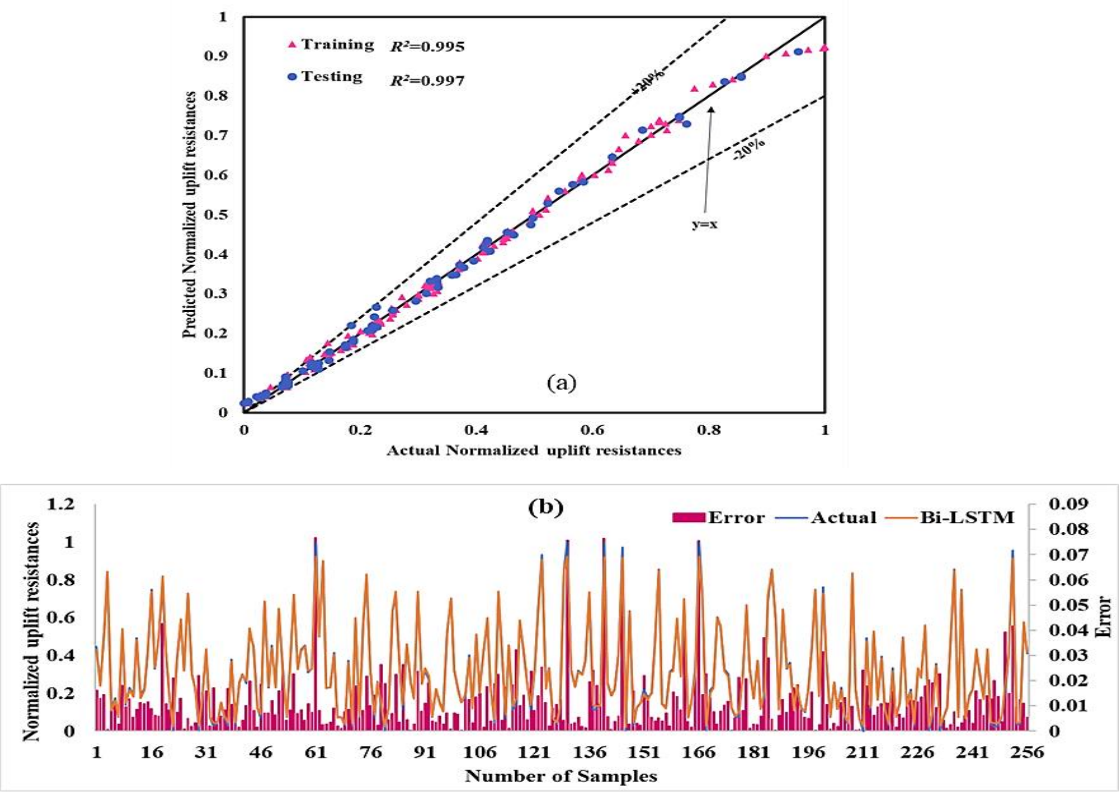
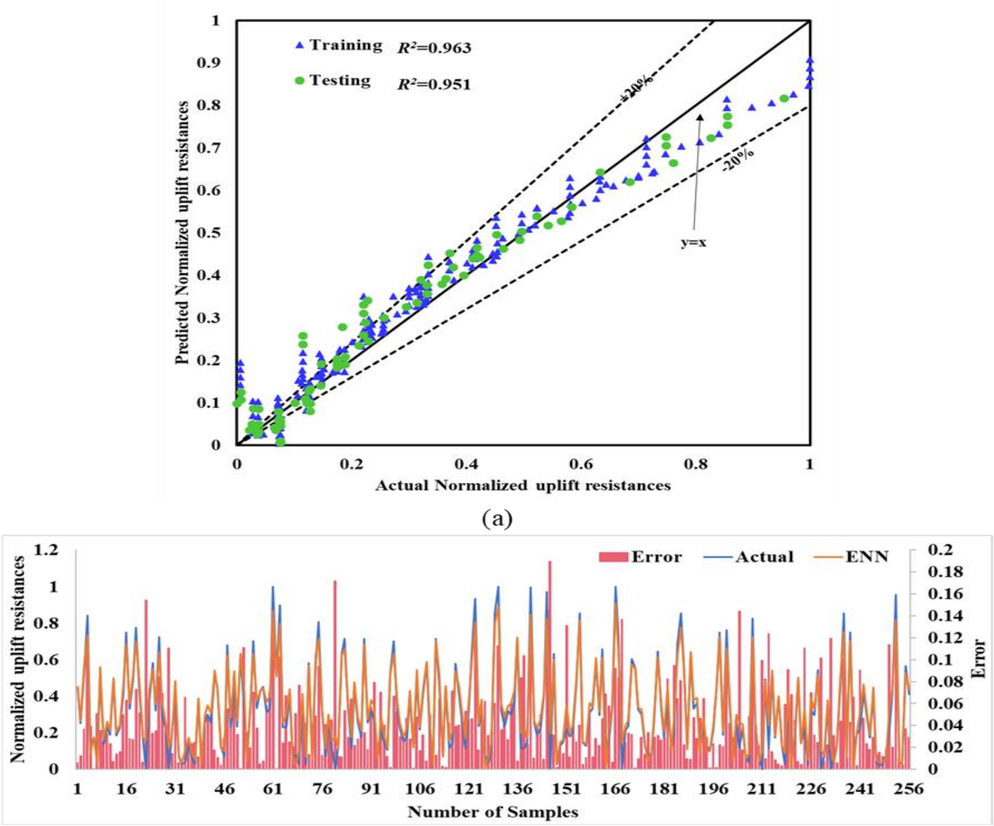


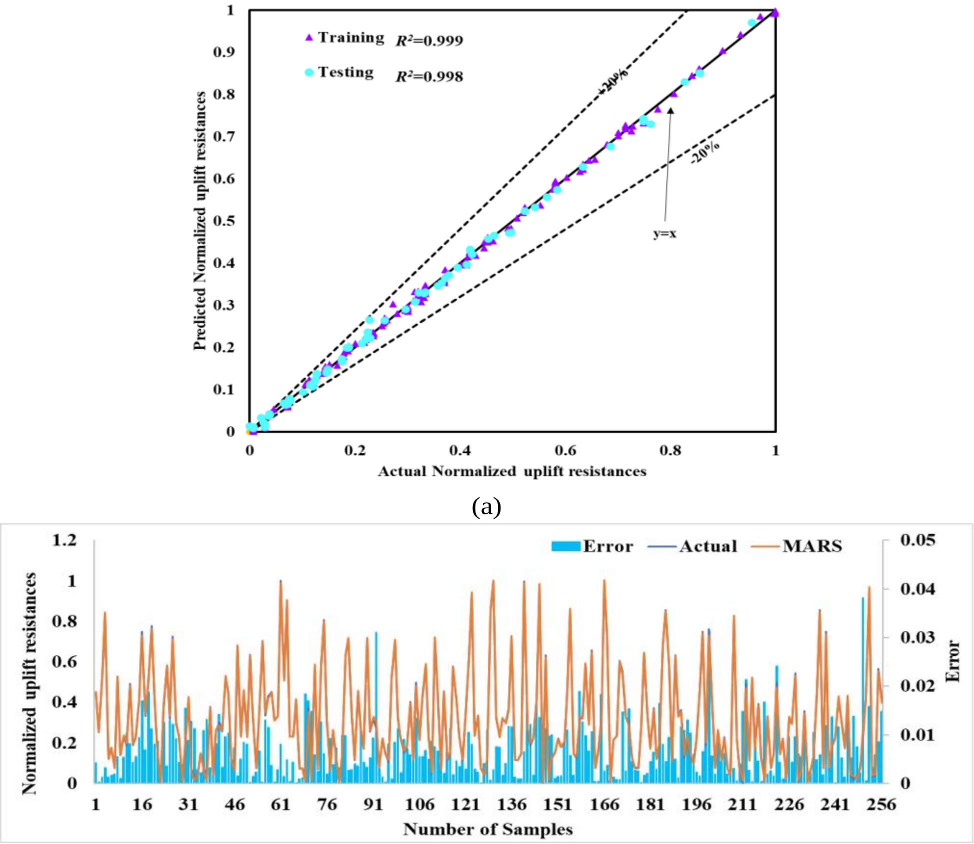
Fig. 10 Performance of Bi-LSTM model for uplift condition (a) Scatter plot (b) Error plot.



(a)

(b)

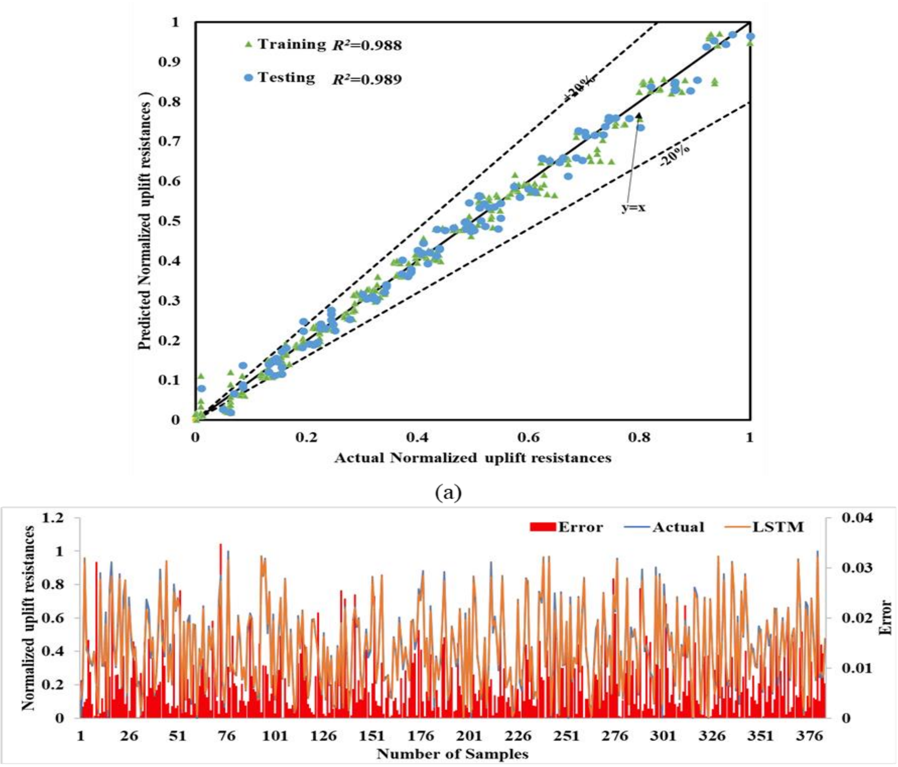
Fig. 11 Performance of ENN model for uplift condition (a) Scatter plot (b) Error plot.



(a)

(b)

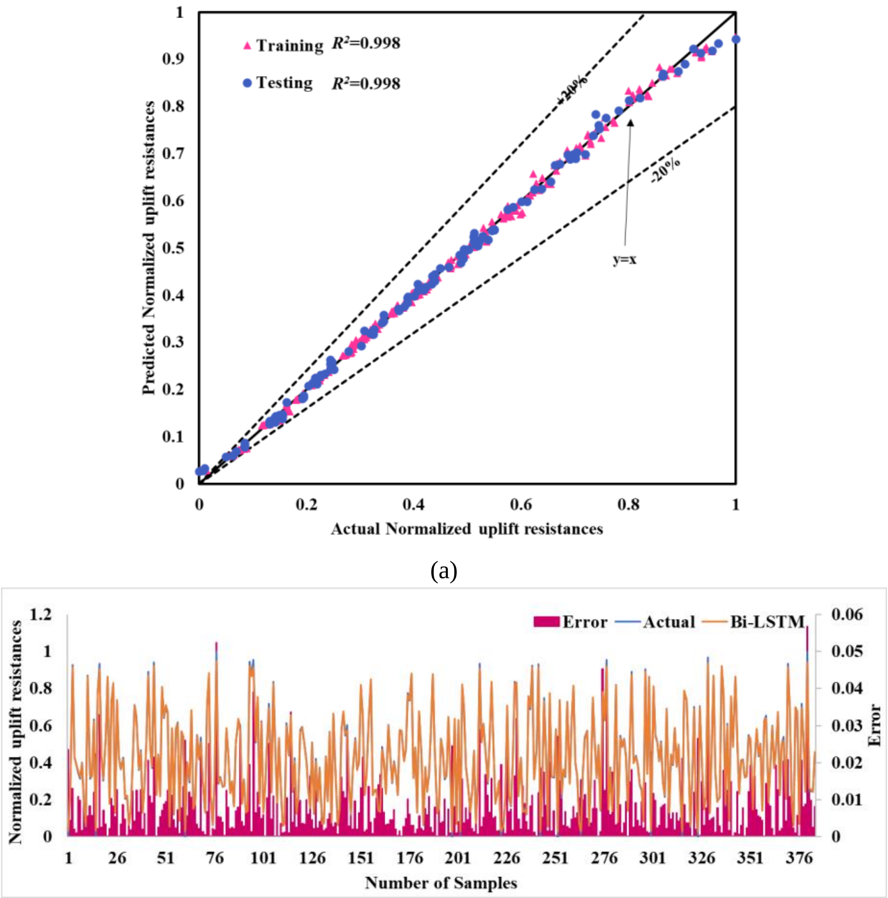
Fig. 12 Performance of MARS model for uplift condition (a) Scatter plot (b) Error plot.



(a)

(b)

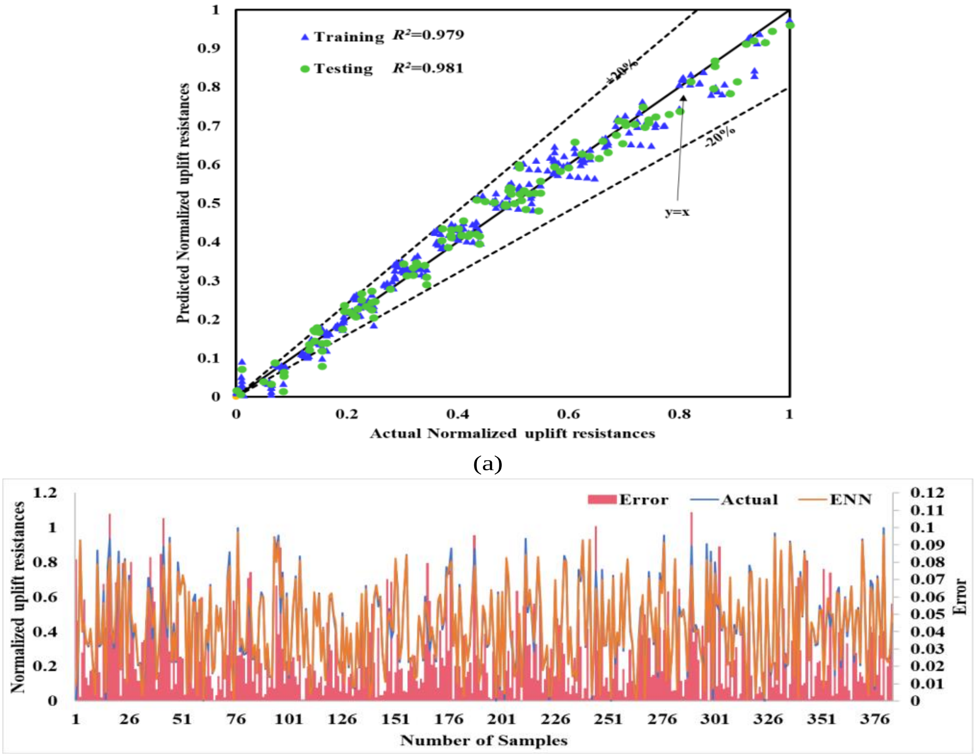
Fig. 13 Performance of LSTM model for penetration condition (a) Scatter plot (b) Error plot.



(a)

(b)

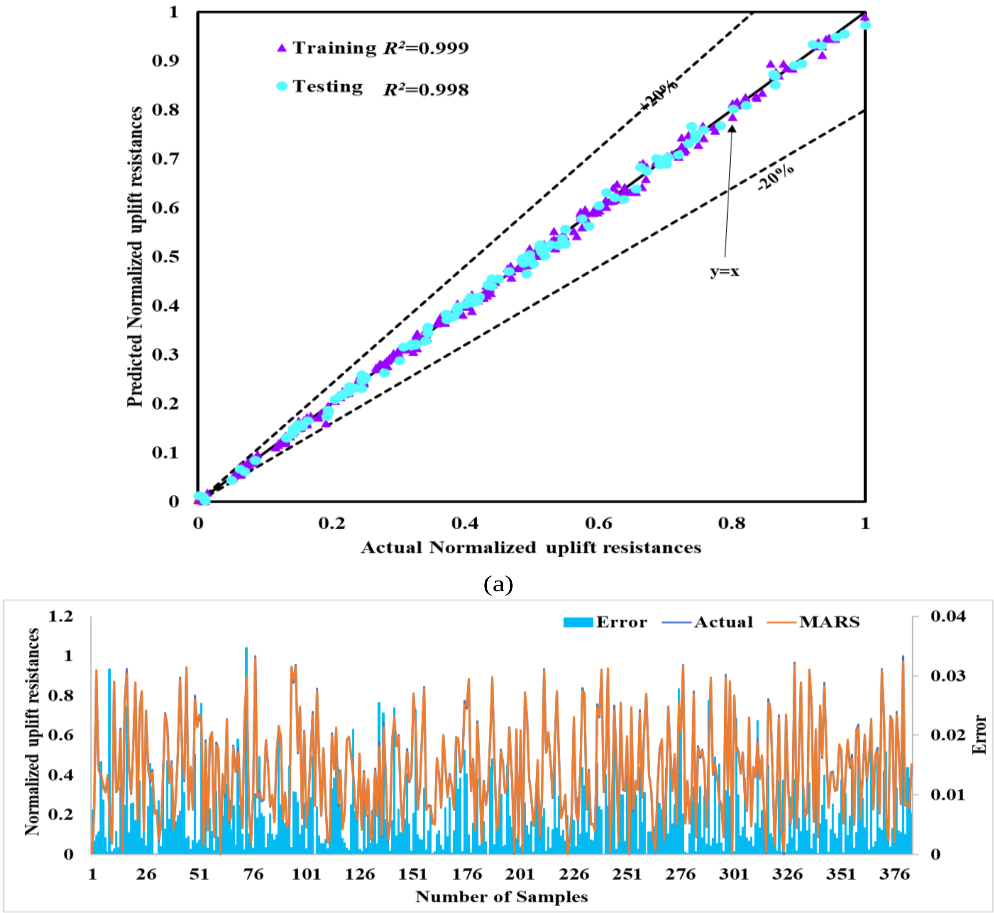
Fig. 14 Performance of Bi-LSTM model for penetration condition (a) Scatter plot (b) Error plot.



(a)

(b)

Fig. 15 Performance of ENN model for penetration condition (a) Scatter plot (b) Error plot.



(a)

(b)

Fig. 16 Performance of MARS model for penetration condition (a) Scatter plot (b) Error plot.

presented as follows:

$$Total\ score = \sum_{i=1}^{m} S_i + \sum_{j=1}^{n} S_j \tag{34}$$

where S_i and S_j represent the score of the individual statistical parameters for the training and testing phases, respectively. m and n represent the number of statistical parameters used for the rank analysis.

The best model has allotted the first rank, and the worst model has the last position. Overall, rank analysis provides a useful tool for model selection by enabling researchers to compare the performance of different models and identify the best model based on its ability to predict uplift and penetration resistance.

In the case of the uplift condition, the MARS model achieved the first rank with a score value of 80, followed by the BI-LSTM (60), LSTM (40), and ENN (20) models, which achieved second, third, and fourth ranks, respectively. Similarly, for penetration conditions, the MARS model again achieved the first rank with a score value of 80, followed by the BI-LSTM (60), LSTM (40), and ENN (20) models, achieving second, third, and fourth ranks, respectively. It can be concluded that MARS performs best when compared to the other models. However, the Bi-LSTM, LSTM, and ENN models can also predict the normalized uplift and penetration resistance. Tables 9 and 10 display the rank of the proposed models for uplift and penetration conditions, respectively.

4.6 Uncertainty analysis

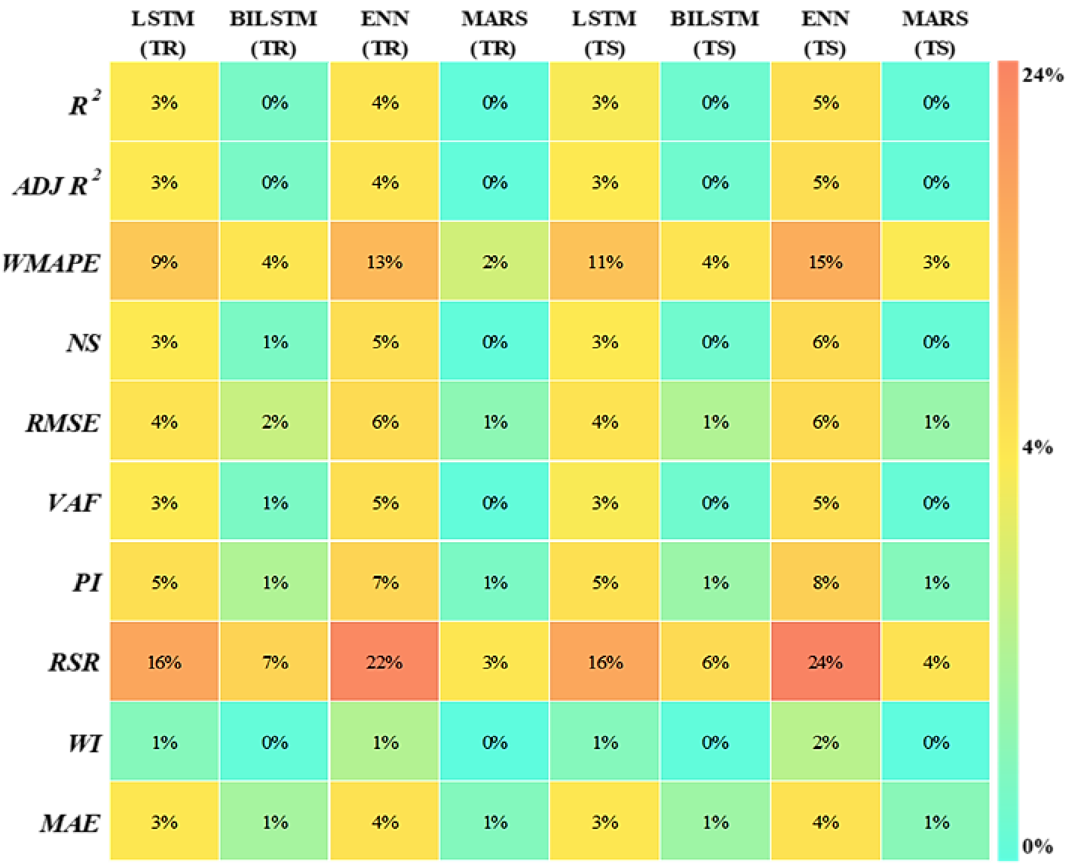
Uncertainty analysis is a process used to assess the potential errors and uncertainties of predictive models. This analysis is performed for 256 cases of uplift conditions and 384 cases of penetration conditions. Several steps are involved in uncertainty analysis. The first step is to characterize the uncertainty. This step consists of quantifying the uncertainty associated with each source. This may include calculating standard deviations (S_e), mean error ($\hat{e}$), confidence intervals, or other statistical measures. The mathematical formulations used for the calculation of standard deviations (S_e) and mean error ($\hat{e}$) are presented as follows in Eqs. (35) and (36):

$$S_e = \sqrt{\frac{\sum_{i=1}^{n}(e_i - \hat{e})}{n-1}} \tag{35}$$

$$\hat{e} = \frac{\sum_{i=1}^{n} e_i}{n-1} \tag{36}$$

Table 9. Rank analysis for uplift condition.

Model		LSTM		Bi-LSTM		ENN		MARS	
Parameters		TR	TS	TR	TS	TR	TS	TR	TS
R^2	Value	0.973	0.974	0.995	0.997	0.963	0.951	0.999	0.998
	Score	2	2	3	3	1	1	4	4
Adj R^2	Value	0.973	0.973	0.995	0.997	0.962	0.949	0.999	0.998
	Score	2	2	3	3	1	1	4	4
WMAPE	Value	0.095	0.108	0.036	0.038	0.127	0.151	0.019	0.026
	Score	2	2	3	3	1	1	4	4
NS	Value	0.973	0.973	0.995	0.997	0.950	0.944	0.999	0.998
	Score	2	2	3	3	1	1	4	4
RMSE	Value	0.040	0.039	0.018	0.014	0.055	0.056	0.008	0.010
	Score	2	2	3	3	1	1	4	4
VAF	Value	97.31	97.43	99.49	99.66	95.30	94.77	99.89	99.82
	Score	2	2	3	3	1	1	4	4
PI	Value	1.906	1.908	1.972	1.979	1.860	1.841	1.990	1.986
	Score	2	2	3	3	1	1	4	4
RSR	Value	0.164	0.164	0.072	0.058	0.225	0.236	0.033	0.042
	Score	2	2	3	3	1	1	4	4
WI	Value	0.993	0.993	0.999	0.999	0.986	0.985	1.000	1.000
	Score	2	2	3	3	1	1	4	4
MAE	Value	0.032	0.031	0.012	0.011	0.043	0.043	0.006	0.007
	Score	2	2	3	3	1	1	4	4
Sub Total		20	20	30	30	10	10	40	40
Total Score		40		60		20		80	
Rank		Third		Second		Forth		First	

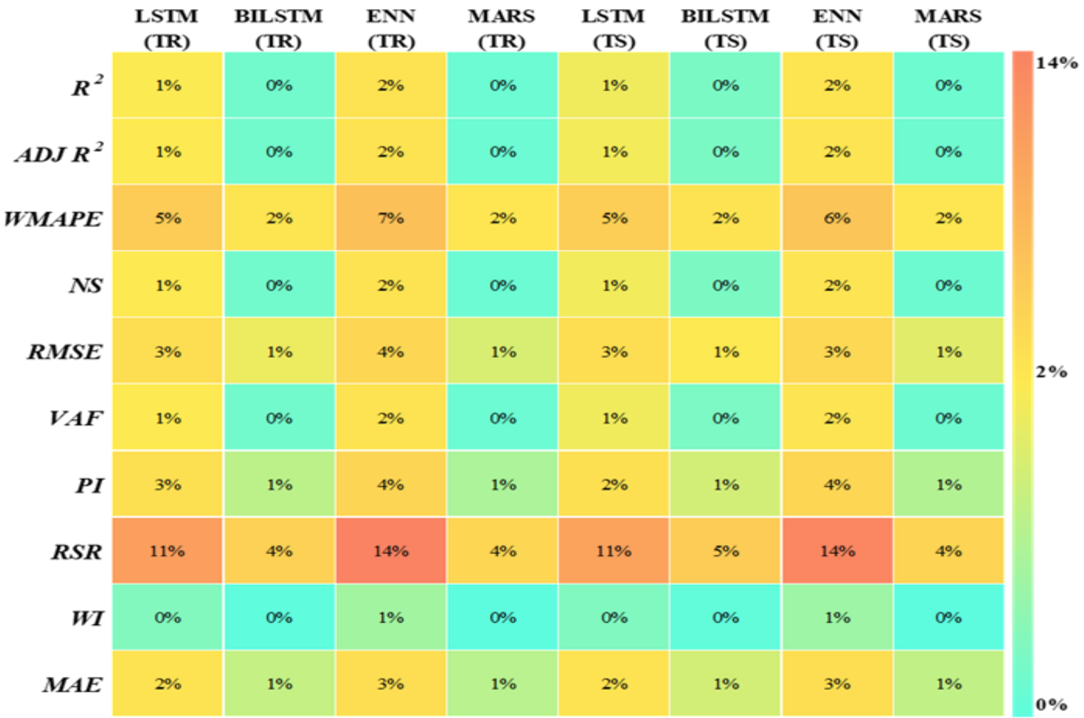


	LSTM (TR)	BILSTM (TR)	ENN (TR)	MARS (TR)	LSTM (TS)	BILSTM (TS)	ENN (TS)	MARS (TS)
R^2	3%	0%	4%	0%	3%	0%	5%	0%
$ADJ\ R^2$	3%	0%	4%	0%	3%	0%	5%	0%
WMAPE	9%	4%	13%	2%	11%	4%	15%	3%
NS	3%	1%	5%	0%	3%	0%	6%	0%
RMSE	4%	2%	6%	1%	4%	1%	6%	1%
VAF	3%	1%	5%	0%	3%	0%	5%	0%
PI	5%	1%	7%	1%	5%	1%	8%	1%
RSR	16%	7%	22%	3%	16%	6%	24%	4%
WI	1%	0%	1%	0%	1%	0%	2%	0%
MAE	3%	1%	4%	1%	3%	1%	4%	1%

Fig. 17 Error matrix for uplift condition.

Table 10. Rank analysis for penetration condition.

Model		LSTM		Bi-LSTM		ENN		MARS	
Parameters		TR	TS	TR	TS	TR	TS	TR	TS
R^2	Value	0.988	0.989	0.998	0.998	0.979	0.981	0.999	0.998
	Score	2	2	3	3	1	1	4	4
Adj R^2	Value	0.988	0.989	0.998	0.997	0.979	0.980	0.999	0.998
	Score	2	2	3	3	1	1	4	4
WMAPE	Value	0.050	0.047	0.019	0.020	0.066	0.060	0.017	0.017
	Score	2	2	3	3	1	1	4	4
NS	Value	0.988	0.989	0.998	0.997	0.979	0.980	0.999	0.998
	Score	2	2	3	3	1	1	4	4
RMSE	Value	0.027	0.026	0.011	0.013	0.036	0.035	0.009	0.010
	Score	2	2	3	3	1	1	4	4
VAF	Value	98.806	98.90	99.81	99.74	97.91	98.05	99.86	99.84
	Score	2	2	3	3	1	1	4	4
PI	Value	1.949	1.951	1.985	1.982	1.922	1.926	1.988	1.987
	Score	2	2	3	3	1	1	4	4
RSR	Value	0.110	0.106	0.044	0.051	0.145	0.140	0.037	0.040
	Score	2	2	3	3	1	1	4	4
WI	Value	0.997	0.997	1.000	0.999	0.995	0.995	1.000	1.000
	Score	2	2	3	3	1	1	4	4
MAE	Value	0.021	0.021	0.008	0.009	0.027	0.027	0.007	0.008
	Score	2	2	3	3	1	1	4	4
Sub Total		20	20	30	30	10	10	40	40
Total Score		40		60		20		80	
Rank		Third		Second		First		First	



	LSTM (TR)	BILSTM (TR)	ENN (TR)	MARS (TR)	LSTM (TS)	BILSTM (TS)	ENN (TS)	MARS (TS)
R^2	1%	0%	2%	0%	1%	0%	2%	0%
$ADJ\ R^2$	1%	0%	2%	0%	1%	0%	2%	0%
WMAPE	5%	2%	7%	2%	5%	2%	6%	2%
NS	1%	0%	2%	0%	1%	0%	2%	0%
RMSE	3%	1%	4%	1%	3%	1%	3%	1%
VAF	1%	0%	2%	0%	1%	0%	2%	0%
PI	3%	1%	4%	1%	2%	1%	4%	1%
RSR	11%	4%	14%	4%	11%	5%	14%	4%
WI	0%	0%	1%	0%	0%	0%	1%	0%
MAE	2%	1%	3%	1%	2%	1%	3%	1%

Fig. 18 Error matrix for penetration condition.

Where e_i represents the mean error of individual feature parameters, which can be calculated with the help of Eq. (37).

$$e_i = y_i - d_i \tag{37}$$

where d_i and y_i represent the actual and predicted output values, respectively.

The second step is to combine the uncertainties. Once the individual uncertainties are quantified, they can be combined to determine the overall uncertainty associated with the measurement or calculation. The third step is to communicate the uncertainty: It is important to communicate the uncertainty related to the results of a measurement or calculation to ensure that others can correctly interpret the data. Uncertainty analysis is a crucial part of scientific research as it helps to ensure that results are accurate and reliable. The Wilson score method is applied to calculate the confidence band for the mean error with the help of calculated standard deviations (S_e) and mean error ($\hat{e}$) values.[53,54] The results of the uncertainty analysis for both uplift and penetration conditions are presented in Tables 11 and 12, respectively. Additionally, the comparison of the uncertainty bandwidths of all the models for both uplift and penetration conditions is shown in Fig. 19. The results are presented in Tables 11 and 12. The MARS model represents the minimum mean error value in uplift and penetration conditions, followed by the Bi-LSTM, LSTM, and ENN models. Conversely, the MARS model represents a lesser uncertainty bandwidth than the BI-LSTM, LSTM, and ENN models, which means that the MARS model provides the best result in uplift and penetration conditions. However, the Bi-LSTM, LSTM, and ENN models can also accurately predict the uplift and penetration resistance.

Table 11. Results of uncertainty analysis for the uplift condition.

Model	Uplift Condition	Mean Error	SD E	Band width	95% confidence interval error	
LSTM	Training	-0.0009	0.0405	0.0794	-0.0803	0.0785
	Testing	-0.0080	0.0387	0.0759	-0.0838	0.0679
BI-LSTM	Training	0.0007	0.0177	0.0346	-0.0339	0.0354
	Testing	0.0015	0.0141	0.0276	-0.0260	0.0291
ENN	Training	0.0144	0.0536	0.1051	-0.0906	0.1195
	Testing	0.0136	0.0552	0.1083	-0.0947	0.1218
MARS	Training	3.37E-06	0.0081	0.0161	-0.0160	0.0161
	Testing	-0.0005	0.0091	0.0178	-0.0184	0.0173

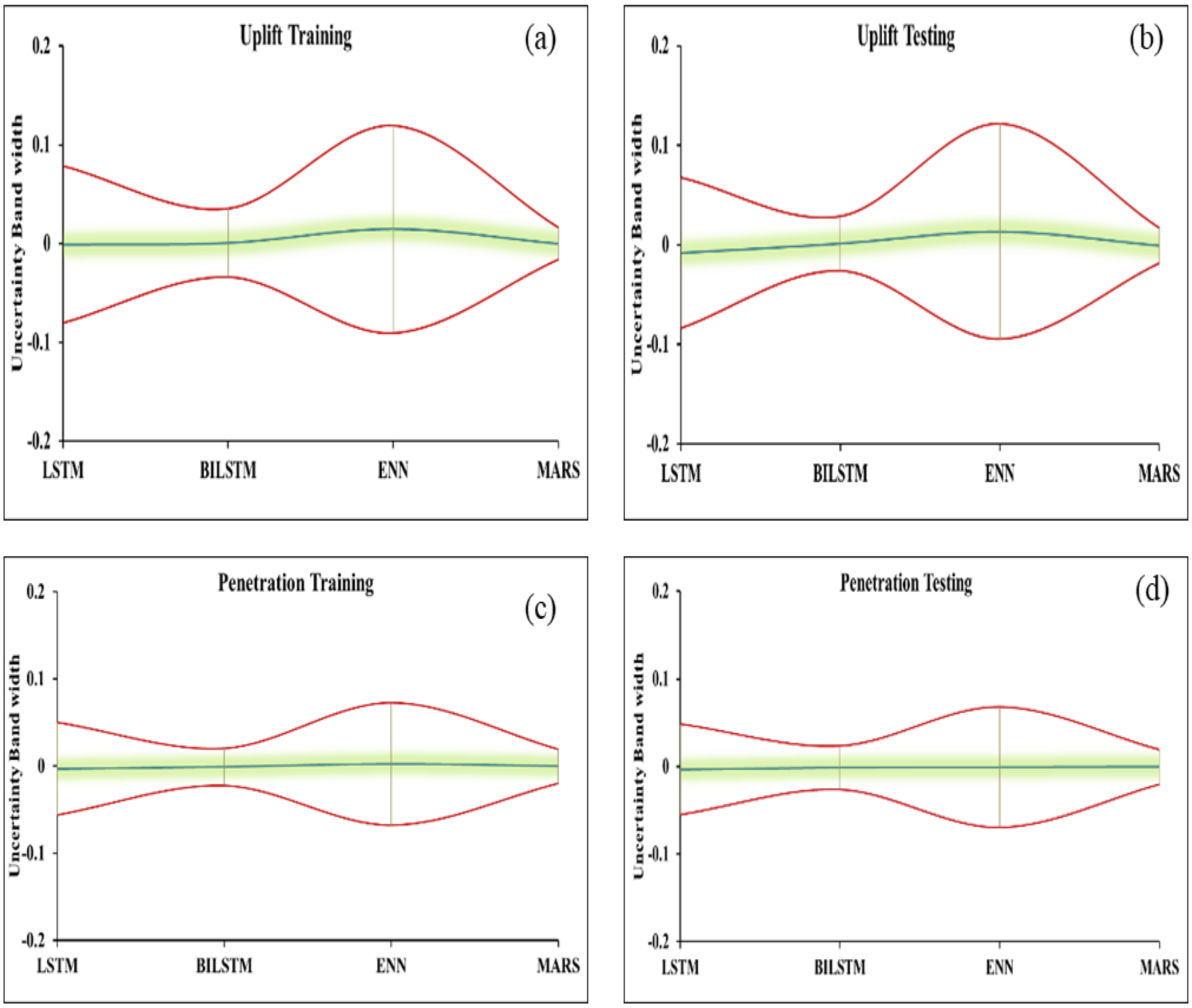


Fig. 19 Uncertainty band width for uplift and penetration condition, a) Uncertainty band width of uplift condition in training, b) Uncertainty band width of uplift condition in testing, c) Uncertainty band width of penetration condition in training, d) Uncertainty band width of penetration condition in testing.

Table 12. Results of uncertainty analysis for the penetration condition.

Model	Uplift Condition	Mean Error	SD E	Band width	95% confidence interval error	
LSTM	Training	-0.0029	0.0271	0.0531	-0.0560	0.0502
	Testing	-0.0032	0.0264	0.0517	-0.0549	0.0486
BI-LSTM	Training	-0.0007	0.0108	0.0213	-0.0219	0.0206
	Testing	-0.0012	0.0127	0.0249	-0.0261	0.0237
ENN	Training	0.0024	0.0358	0.0702	-0.0678	0.0726
	Testing	-0.0008	0.0351	0.0688	-0.0696	0.0680
MARS	Training	0.0001	0.0099	0.0195	-0.0194	0.0196
	Testing	-0.0004	0.0101	0.0197	-0.0201	0.0194

4.7 Sensitivity analysis

Sensitivity analysis is a method used to evaluate the sensitivity of a model's output to changes in its input parameters. It involves changing one or more input variables and observing the resulting changes in the output variable. This analysis is useful in understanding how different input parameters affect the model's output and identifying which input variables have the most significant impact on the output and which input variables have the least significant impact on the output. Therefore, to analyze the impact of the input parameters on the output parameters (*i.e.*, the uplift resistance ($q_t/\rho D$) for the pull-out case and penetration resistance ($q_c/\rho D$) for the bearing case), the cosine amplitude method of Jong & Lee (2004) is used in this paper. The dataset is presented in the form of a data array, X in the following form in Eq. (38):

$$X = \{x_1, x_2, \dots x_i \dots x_n\} \quad (38)$$

where the variable x_i represents a length vector of m as follows in Eq. (39):

$$x_i = \{x_{i1}, x_{i2}, \dots \dots x_{im}\} \quad (39)$$

To determine the correlation between the strength of relation (R_{ij}) and the input datasets of x_i and x_j is calculated using the mathematical Eq. (40).

$$R_{ij} = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ik}^2 \sum_{k=1}^{m} x_{jk}^2}} \tag{40}$$

The obtained values of R_{ij} for both uplift and penetration conditions are presented in Figs. 20 and 21, respectively, indicating the relation between the input parameters and uplift resistance and penetration resistance. The results of sensitivity analysis presented in Figs. 20 and 21 reveal that the pipe embedded depth ratio (*w/D*) has the greatest influence on the uplift and penetration resistance with strength values of 0.91 and 0.99, respectively. The spacing ratio (*S/D*) has the minimum effect on the uplift and penetration resistance with strength values of 0.79 and 0.80, respectively. All the three input parameters have a higher influence on the output of both uplift and penetration conditions; hence, all three parameters are considered in predicting the uplift and penetration resistance.

4.8 Taylor diagram

A Taylor diagram is a graphical tool used to assess the similarity between two or more sets of data or models. It was introduced by Karl E. Taylor[55] as a way to visually represent the correlation, root mean square error (RMSE), and standard deviation of different proposed machine learning models in a concise and intuitive manner. The Taylor diagram consists of a polar coordinate system where the reference dataset or model is represented by a point at the origin. Other datasets or models are represented by points on the diagram, with their distance from the origin indicating their correlation with the actual dataset. The angle between the radial line connecting the point to the origin and the horizontal axis represents the standard deviation ratio, while the distance from the reference dataset to the point represents the RMSE ratio. By plotting multiple datasets or models on a Taylor diagram, it becomes easy to compare their correlation, RMSE, and standard deviation with respect to a reference dataset. Points that are close to the reference point indicate a high correlation and low RMSE, while points farther away indicate lower correlation and higher RMSE. The angle of the point also provides information about the standard deviation. Figs. 22 and 23 represent the Taylor diagrams for uplift conditions in the training and testing phases, respectively, and Figs. 24 and 25 represent the Taylor diagrams for penetration conditions in the training and testing phases, respectively. It can be observed from the presented figures that the MARS model is close to the reference point and has a high correlation coefficient and low RMSE under both uplift and penetration conditions, while the ENN model lies further away from the reference point and has a low correlation coefficient and high RMSE under both uplift and penetration conditions.

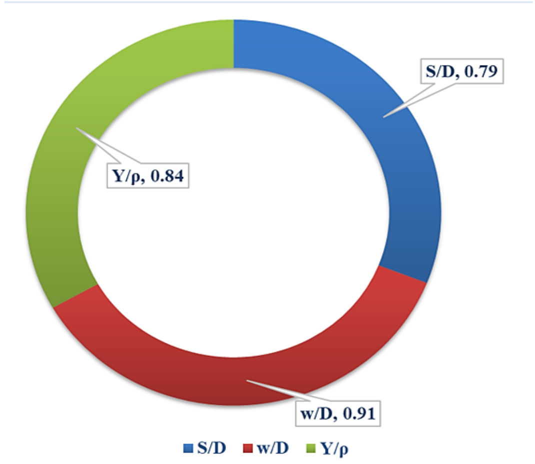


Fig. 20 Doughnut diagram of sensitivity analysis for uplift condition.

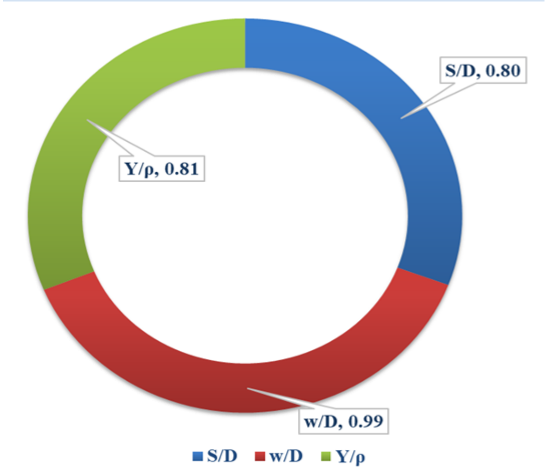


Fig. 21 Doughnut diagram of sensitivity analysis for penetration condition.

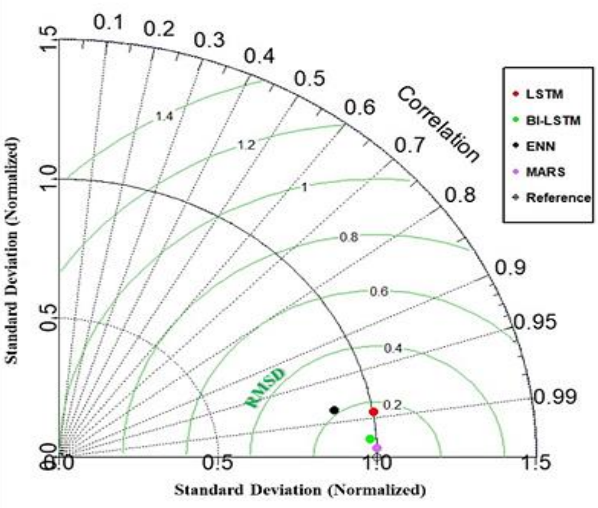


Fig. 22 Taylor diagram of uplift condition in training.

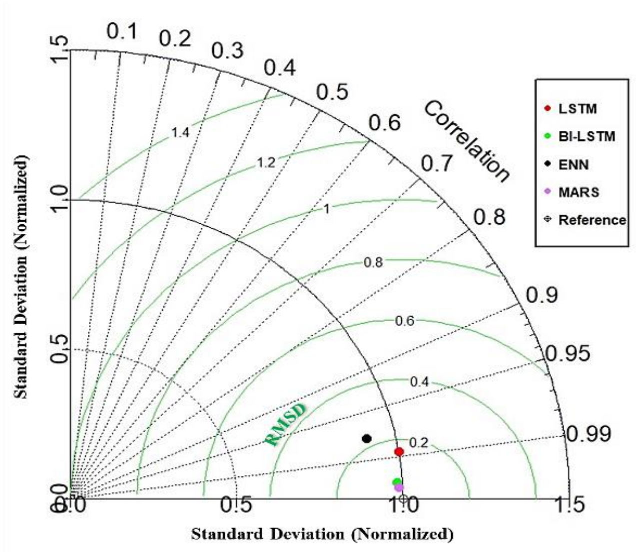


Fig. 23 Taylor diagram of uplift condition in testing.

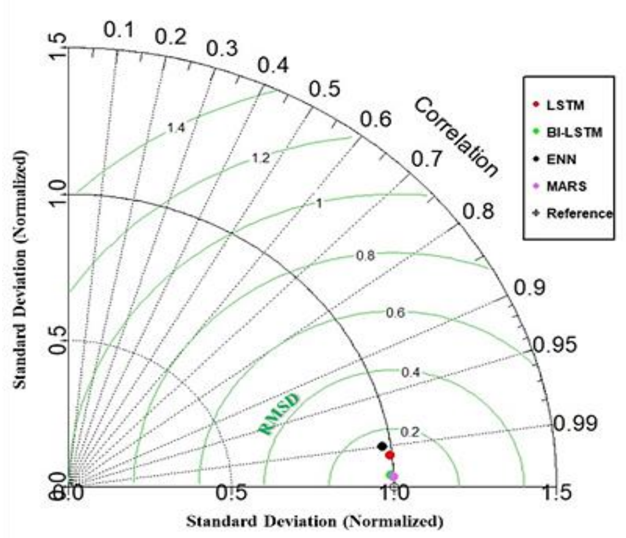


Fig. 24 Taylor diagram of penetration condition in training.

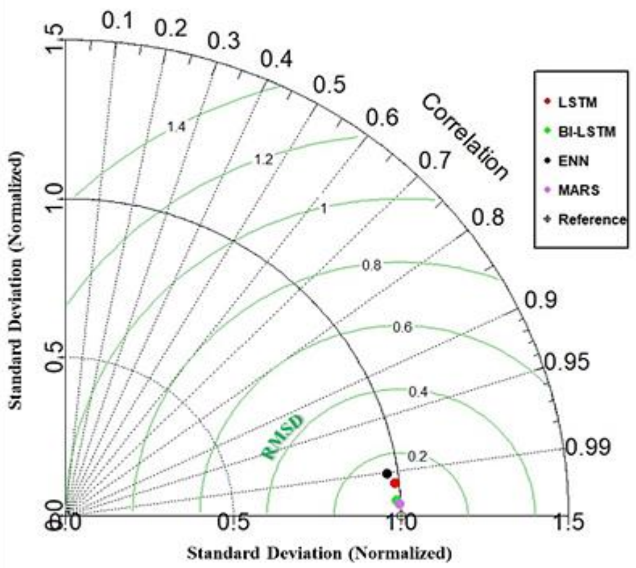


Fig. 25 Taylor diagram of penetration condition in testing.

5. Conclusion and future scope

This paper presents AI-based regression models for evaluating the normalized uplift and penetration resistance of two interfering pipelines buried in clay, which increase in strength linearly based on the FELA solution. Three input parameters, namely, the spacing ratio (*S/D*), the pipe embedded depth ratio (*w/D*), and the normalized soil unit weight (γ/ρ), are considered as input parameters, while the normalized uplift resistance ($q_t/\rho D$) and normalized penetration resistance ($q_c/\rho D$) are considered as output parameters during training of the regression models. To develop regression models such as LSTM, Bi-LSTM, ENN, and MARS models, 254 and 384 FELA solutions are used for uplift and penetration conditions, respectively. The predictive performance of the models used was examined and compared using various indicators of performance metrics, including scatter and error plots, rank analysis, error matrices, Taylor diagram and uncertainty analysis. Sensitivity analysis is also performed to determine the impact of the most and least significant input parameters on the output parameters for both uplift and penetration condition. The results of the sensitivity analysis indicate that the pipe embedded depth ratio (*w/D*) has the most significant effect and the spacing ratio (*S/D*) has the least significant effect in both uplift and penetration conditions. According to the obtained results of all models presented in Tables 7 and 8, it can be concluded that the MARS model outperforms overall in both uplift and penetration conditions. All used models can predict the normalized uplift and penetration resistance, as presented by the performed study. However, the predictive equations (32) and (33) of MARS can be employed as a straightforward formula to estimate the normalized uplift and penetration resistance of two interfering pipelines buried in clay. The uncertainty bandwidth in the MARS model case is narrower than that in the BI-LSTM, LSTM, and ENN models for uplift and penetration conditions. The results of uncertainty analysis have shown that the developed MARS model is robust and can generalize over the given data range. Importantly, once the necessary database is established, practical engineers can implement the proposed models in additional engineering applications.

The proposed advanced regression models have several benefits, such as being inexpensive, easy to connect to FELA and experimental data, and having high prediction accuracy. The used models predicting the uplift and penetration resistance of two interfering pipelines are valid for the abovementioned dimensionless input and output parameters. On the other hand, once more recent data become accessible, the proposed models can be easily updated.

Acknowledgment

This work was supported by Thammasat University Research Unit in Data Science and Digital Transformation.

Conflict of Interest

There is no conflict of interest.

Supporting Information

Not applicable.

References

[1] C. H. Trautmann, T. D. O'Rourfce, F. H. Kulhawy, Uplift force-displacement response of buried pipe, *Journal of Geotechnical Engineering*, 1985, **111**, 1061-1076, doi: 10.1061/(asce)0733-9410(1985)111: 9(1061).

[2] E. A. Dickin, Uplift resistance of buried pipelines in sand, *Soils and Foundations*, 1994, **34**, 41-48, doi: 10.3208/sandf1972.34.2_41.

[3] H. I. Ling, Y. Mohri, T. Kawabata, H. Liu, C. Burke, L. Sun, Centrifugal modeling of seismic behavior of large-diameter pipe in liquefiable soil, *Journal of Geotechnical and Geoenvironmental Engineering*, 2003, **129**, 1092-1101, doi: 10.1061/(asce)1090-0241(2003)129: 12(1092).

[4] F. Calvetti, C. di Prisco, R. Nova, Experimental and numerical analysis of soil-pipe interaction, *Journal of Geotechnical and Geoenvironmental Engineering*, 2004, **130**, 1292-1299, doi: 10.1061/(asce)1090-0241(2004)130: 12(1292).

[5] C. Y. Cheuk, D. J. White, M. D. Bolton, Uplift mechanisms of pipes buried in sand, *Journal of Geotechnical and Geoenvironmental Engineering*, 2008, **134**, 154-163, doi: 10.1061/(asce)1090-0241(2008)134: 2(154).

[6] D. J. White, C. Y. Cheuk, M. D. Bolton, The uplift resistance of pipes and plate anchors buried in sand, *Géotechnique*, 2008, **58**, 771-779, doi: 10.1680/geot.2008.3692.

[7] J. Kumar, Uplift Response of Buried Pipes in Sand Using FEM, *Indian Geotechnical Journal*, 2002, **32**, 146-160.

[8] P. Guo, Numerical modeling of pipe-soil interaction under oblique loading, *Journal of Geotechnical and Geoenvironmental Engineering*, 2005, **131**, 260-268, doi: 10.1061/(asce)1090-0241(2005)131: 2(260).

[9] P. J. Guo, D. F. E. Stolle, Lateral pipe-soil interaction in sand with reference to scale effect, *Journal of Geotechnical and Geoenvironmental Engineering*, 2005, **131**, 338-349, doi: 10.1061/(asce)1090-0241(2005)131: 3(338).

[10] S. Chatterjee, M. F. Randolph, D. J. White, The effects of penetration rate and strain softening on the vertical penetration resistance of seabed pipelines, *Géotechnique*, 2012, **62**, 573-582, doi: 10.1680/geot.10.p.075.

[11] S. Chatterjee, D. J. White, M. F. Randolph, Numerical simulations of pipe-soil interaction during large lateral movements on clay, *Géotechnique*, 2012, **62**, 693-705, doi: 10.1680/geot.10.p.107.

[12] S. Chatterjee, D. J. White, M. F. Randolph, Coupled consolidation analysis of pipe-soil interactions, *Canadian Geotechnical Journal*, 2013, **50**, 609-619, doi: 10.1139/cgj-2012-0307.

[13] J. K. Jung, T. D. O'Rourke, N. A. Olson, Lateral soil-pipe interaction in dry and partially saturated sand, *Journal of Geotechnical and Geoenvironmental Engineering*, 2013, **139**, 2028-2036, doi: 10.1061/(asce)gt.1943-5606.0000960.

[14] J. K. Jung, T. D. O'Rourke, N. A. Olson, Uplift soil-pipe interaction in granular soil, *Canadian Geotechnical Journal*, 2013, **50**, 744-753, doi: 10.1139/cgj-2012-0357.

[15] S. Maitra, S. Chatterjee, D. Choudhury, Generalized framework to predict undrained uplift capacity of buried offshore pipelines, *Canadian Geotechnical Journal*, 2016, **53**, 1841-1852, doi: 10.1139/cgj-2016-0153.

[16] S. W. Sloan, Geotechnical stability analysis, *Géotechnique*, 2013, **63**, 531-571, doi: 10.1680/geot.12.rl.001.

[17] C. M. Martin, D. J. White, Limit analysis of the undrained bearing capacity of offshore pipelines, *Géotechnique*, 2012, **62**, 847-863, doi: 10.1680/geot.12.og.016.

[18] D. Chakraborty, J. Kumar, Vertical uplift resistance of pipes buried in sand, *Journal of Pipeline Systems Engineering and Practice*, 2014, **5**, 4013009, doi: 10.1061/(asce)ps.1949-1204.0000149.

[19] D. Chakraborty, J. Kumar, Uplift resistance of long pipelines in the presence of seismic forces, *Journal of Pipeline Systems Engineering and Practice*, 2014, **5**, 6014003, doi: 10.1061/(asce)ps.1949-1204.0000172.

[20] D. Chakraborty, J. Kumar, Uplift resistance of interfering pipelines buried in sand, *Journal of Pipeline Systems Engineering and Practice*, 2016, **7**, 6015002, doi: 10.1061/(asce)ps.1949-1204.0000217.

[21] D. Chakraborty, Lateral resistance of buried pipeline in c-ϕ soil, *Journal of Pipeline Systems Engineering and Practice*, 2018, **9**, 6017006, doi: 10.1061/(asce)ps.1949-1204.0000289.

[22] P. Bhattacharya, S. Sahoo, Pullout resistance of buried pipeline in cohesionless soil nearby sloping ground, *Journal of Pipeline Systems Engineering and Practice*, 2020, **11**, 4020008, doi: 10.1061/(asce)ps.1949-1204.0000450.

[23] J. Shiau, B. Chudal, K. Mahalingasivam, S. Keawsawasvong, Pipeline burst-related ground stability in blowout condition, Transportation Geotechnics, 2021, **29**, 100587, doi: 10.1016/j.trgeo.2021.100587.

[24] J. Shiau, S. Keawsawasvong, B. Chudal, K. Mahalingasivam, S. Seehavong, Sinkhole stability in elliptical cavity under collapse and blowout conditions, *Geosciences*, 2021, **11**, 421, doi: 10.3390/geosciences11100421.

[25] J. Shiau, K. Mahalingasivam, B. Chudal, S. Keawsawasvong, Pipeline burst-related soil stability in collapse condition, *Journal of Pipeline Systems Engineering and Practice*, 2022, **13**, 4022019, doi: 10.1061/(asce)ps.1949-1204.0000657.

[26] J. Shiau, B. Chudal, S. Keawsawasvong, Three-dimensional sinkhole stability of spherical cavity, *Acta Geotechnica*, 2022, **17**, 3947-3958, doi: 10.1007/s11440-022-01522-8.

[27] S. Seehavong, S. Keawsawasvong, Penetration and uplift resistances of two interfering pipelines buried in clays, *International Journal of Computational Materials Science and*

Engineering, 2021, **10**, 2150020, doi: 10.1142/s2047684121500202.

[28] M. Najafzadeh, F. Saberi-Movahed, S. Sarkamaryan, NF-GMDH-Based self-organized systems to predict bridge pier scour depth under debris flow effects, *Marine Georesources & Geotechnology*, 2018, **36**, 589-602, doi: 10.1080/1064119X.2017.1355944.

[29] M. Najafzadeh, F. Saberi-Movahed, GMDH-GEP to predict free span expansion rates below pipelines under waves, *Marine Georesources & Geotechnology*, 2019, **37**, 375-392, doi: 10.1080/1064119x.2018.1443355.

[30] F. Saberi-Movahed, M. Najafzadeh, A. Mehrpooya, Receiving more accurate predictions for longitudinal dispersion coefficients in water pipelines: training group method of data handling using extreme learning machine conceptions, *Water Resources Management*, 2020, **34**, 529-561, doi: 10.1007/s11269-019-02463-w.

[31] S. Krabbenhoft, L. Damkilde, K. Krabbenhoft, Lower-bound calculations of the bearing capacity of eccentrically loaded footings in cohesionless soil, *Canadian Geotechnical Journal*, 2012, **49**, 298-310, doi: 10.1139/t11-103.

[32] H. Ciria, J. Peraire, J. Bonet, Mesh adaptive computation of upper and lower bounds in limit analysis, *International Journal for Numerical Methods in Engineering*, 2008, **75**, 899-944, doi: 10.1002/nme.2275.

[33] S. Keawsawasvong, B. Ukritchon, Undrained stability of a spherical cavity in cohesive soils using finite element limit analysis, *Journal of Rock Mechanics and Geotechnical Engineering*, 2019, **11**, 1274-1285, doi: 10.1016/j.jrmge.2019.07.001.

[34] S. Keawsawasvong, K. Yoonirundorn, T. Senjuntichai, Pullout capacity factor for cylindrical suction caissons in anisotropic clays based on anisotropic undrained shear failure criterion, *Transportation Infrastructure Geotechnology*, 2021, **8**, 629-644, doi: 10.1007/s40515-021-00154-x.

[35] S. Keawsawasvong, C. Thongchom, S. Likitlersuang, Bearing capacity of strip footing on hoek-brown rock mass subjected to eccentric and inclined loading, *Transportation Infrastructure Geotechnology*, 2021, **8**, 189-202, doi: 10.1007/s40515-020-00133-8.

[36] S. Keawsawasvong, J. Shiau, C. Ngamkhanong, V. Qui Lai, C. Thongchom, Undrained stability of ring foundations: axisymmetry, anisotropy, and nonhomogeneity, *International Journal of Geomechanics*, 2022, **22**, 04021253, doi: 10.1061/(asce)gm.1943-5622.0002229.

[37] W. Yodsomjai, S. Keawsawasvong, V. Q. Lai, Limit analysis solutions for bearing capacity of ring foundations on rocks using hoek-brown failure criterion, *International Journal of Geosynthetics and Ground Engineering*, 2021, **7**, 1-10, doi: 10.1007/s40891-021-00281-y.

[38] W. Yodsomjai, S. Keawsawasvong, S. Likitlersuang, Stability of unsupported conical slopes in hoek-brown rock masses, *Transportation Infrastructure Geotechnology*, 2021, **8**, 279-295, doi: 10.1007/s40515-020-00137-4.

[39] J. E. Nash, J. V. Sutcliffe, River flow forecasting through conceptual models part I—a discussion of principles, *Journal of Hydrology*, 1970, **10**, 282-290, doi: 10.1016/0022-1694(70)90255-6.

[40] T. Chai, R. R. Draxler, Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature, *Geoscientific Model Development*, 2014, **7**, 1247-1250, doi: 10.5194/gmd-7-1247-2014.

[41] M. Kumar, R. Biswas, D. R. Kumar, T. Pradeep, P. Samui, Metaheuristic models for the prediction of bearing capacity of pile foundation, *Geomechanics and Geoengineering*, 2022, **31**, 129–147, doi: 10.12989/gae.2022.31.2.129.

[42] D. R. Kumar, P. Samui, A. Burman, Prediction of probability of liquefaction using hybrid ANN with optimization techniques, *Arabian Journal of Geosciences*, 2022, **15**, 1-21, doi: 10.1007/s12517-022-10855-3.

[43] D. R. Kumar, P. Samui, A. Burman, Prediction of probability of liquefaction using soft computing techniques, *Journal of the Institution of Engineers (India): Series A*, 2022, **103**, 1195-1208, doi: 10.1007/s40030-022-00683-9.

[44] R. Kumar, B. Rai, P. Samui, A comparative study of prediction of compressive strength of ultra-high performance concrete using soft computing technique, *Structural Concrete*, 2023, doi: 10.1002/suco.202200850.

[45] P. Kumar, P. Samui, Design of an energy pile based on CPT data using soft computing techniques, *Infrastructures*, 2022, **7**, 169, doi: 10.3390/infrastructures7120169.

[46] S. Hochreiter, J. Schmidhuber, Long short-term memory, *Neural Computation*, 1997, **9**, 1735-1780, doi: 10.1162/neco.1997.9.8.1735.

[47] T. Babaie, R. Karimizandi, C. Lucas, Learning based brain emotional intelligence as a new aspect for development of an alarm system, *Soft Computing*, 2008, **12**, 857-873, doi: 10.1007/s00500-007-0258-8.

[48] M. Kumar, A. Bardhan, P. Samui, J. W. Hu, M. R. Kaloop, Reliability analysis of pile foundation using soft computing techniques: a comparative study, *Processes*, 2021, **9**, 486, doi: 10.3390/pr9030486.

[49] E. Lotfi, M.-R. Akbarzadeh-T, Practical emotional neural networks, *Neural Networks*, 2014, **59**, 61-72, doi: 10.1016/j.neunet.2014.06.012.

[50] J. H. Friedman, C. B. Roosen, An introduction to multivariate adaptive regression splines, *Statistical Methods in Medical Research*, 1995, **4**, 197-217, doi: 10.1177/096228029500400303.

[51] M. Kumar, P. Samui, Reliability analysis of pile foundation using ELM and MARS, *Geotechnical and Geological Engineering*, 2019, **37**, 3447-3457, doi: 10.1007/s10706-018-00777-x.

[52] D. R. Kumar, P. Samui, A. Burman, Suitability assessment of the best liquefaction analysis procedure based on SPT data, Multiscale and Multidisciplinary Modeling, *Experiments and Design*, 2023, **6**, 319-329, doi: 10.1007/s41939-023-00148-x.

[53] R. G. Newcombe, Two-sided confidence intervals for the single proportion: comparison of seven methods, *Statistics in Medicine*, 1998, **17**, 857-872, doi: 10.1002/(SICI)1097-

0258(19980430)17: 8857: AID-SIM777>3.0.CO;2-E.
[54] A. M. A. Sattar, Gene expression models for the prediction of longitudinal dispersion coefficients in transitional and turbulent pipe flow, *Journal of Pipeline Systems Engineering and Practice*, 2014, **5**, 4013011, doi: 10.1061/(asce)ps.1949-1204.0000153.