Asian Journal of Civil Engineering
https://doi.org/10.1007/s42107-023-00818-8

RESEARCH

Efficient hybrid machine learning model for calculating load-bearing capacity of driven piles

Trong-Ha Nguyen[1] · Kieu-Vinh Thi Nguyen[1] · Viet-Chuong Ho[1] · Duy-Duan Nguyen[1]

Received: 3 July 2023 / Accepted: 9 July 2023

Abstract
The aim of this study is to develop an efficient hybrid machine learning (ML) model, which combines the genetic algorithm (GA) and artificial neural network (ANN) for rapidly calculating the load-bearing capacity (LBC) reinforced concrete driven piles. An extensive database including 470 static tests is collected to train the hybrid ML model. The predicted results of the GA–ANN model in this study are compared to those of the pure ANN model. Statistical indicators containing the coefficient of determination (R^2), root-mean-squared error ($RMSE$), and $a20 - index$ are determined to assess the prediction performance of the ML models. The comparison emphasizes that the GA–ANN model predicts the LBC of the pile accurately with a very high R^2 value of 0.99 and small $RMSE$ of 49 kN. Furthermore, the effects of input variables on the predicted LBC are evaluated. Finally, to apply the ML model, a graphical user interface tool is developed for simplifying the LBC of reinforced concrete driven piles.

Keywords Reinforced concrete driven piles · Load-bearing capacity · GA–ANN · GUI tool

Introduction

Reinforced concrete (RC)-driven piles play a crucial role in load-bearing capacity of deep foundations of large civil engineering structures. Calculating the axial load-bearing capacity (LBC) of pile is an important step in designing deep foundations. Currently, there are many design code provisions and guidelines for estimating the capacity of piles. It is required to obtain soil mechanic properties before using equations in design standards. Some typical field methods for ground testing have been employed in construction projects such as static probing test, dynamic probing test, standard penetration test (SPT), cone penetration test, and field vane test. Additionally, engineers uses a combination of several field tests and laboratory tests. However, this approach is always time-consuming and costly (Kozłowski & Niemczynski, 2016; Pham & Tran, 2022).

So far, some studies have been conducted on applications of machine learning (ML) models for estimating LBC of RC driven piles. Pham et al. (2020) determined the bearing capacity of piles using evolution algorithms and deep learning neural networks. For that, they used 472 results of static load tests, and then concluded that performance of the ML models was accurate with goodness of fit (R^2) values of 0.9 and root-mean-squared error ($RMSE$) larger than 83 kN. Using the same data samples, Pham and Tran (2022) developed hybrid models with a combination of random forest and optimization algorithms. They obtained a high accuracy with R^2 of 0.987. Recently, Nguyen et al. (2023b) constructed an extreme gradient boosting model and then optimized by whale optimization technique for estimating LBC of RC piles. A higher R^2 value of 0.96 and small $RMSE$ of 64 kN were achieved. However, it is a challenge to apply those ML models since a practical tool was not developed for engineering design purposes.

Machine learning (ML) techniques have been employing popularly in civil and structural engineering (Kaveh, 2014; Kaveh & Bondarabady, 2004; Kaveh & Servati, 2001; Kaveh et al., 2008; Nguyen et al., 2022; Tran & Nguyen,

✉ Duy-Duan Nguyen
duan468@gmail.com

Trong-Ha Nguyen
trongha@vinhuni.edu.vn

Kieu-Vinh Thi Nguyen
kieuvinhkxd@vinhuni.edu.vn

Viet-Chuong Ho
vietchuongkxd@vinhuni.edu.vn

[1] Department of Civil Engineering, Vinh University, Vinh 461010, Vietnam

Published online: 20 July 2023

2022). It can be found that artificial neural network (ANN) is the most common ML models applying for estimating responses of RC and steel structures (Ahmed et al., 2019; Kaveh & Khavaninzadeh, 2023; Mai et al., 2022; Nguyen et al., 2021a, b, c, 2023a; Rönnholm et al., 2005; Selvan et al., 2018; Tran & Kim, 2020; Tran et al., 2022; Vakhshouri & Nejadi, 2018; Yang et al., 1992; Zorlu et al., 2008). Moreover, numerous studies have combined ANN and other optimization techniques for improving the predictive accuracy such as genetic algorithm (GA) (Bülbül et al., 2022; Chaabene & Nehdi, 2020; Congro et al., 2021; Rahami et al., 2008; Vijayakumar & Pannirselvam, 2022) and particle swarm optimization (PSO) (Barkhordari et al., 2022; Chatterjee et al., 2017; Chen et al., 2018; Huang et al., 2022; Naderpour et al., 2021; Nanda et al., 2014; Nguyen et al., 2020, 2023d). However, it is required to transform those efficient ML models into practical tools for solving engineering problems.

The purpose of this study is to develop an efficient hybrid ML model, which combines GA and ANN algorithms for improving the LCB prediction of RC driven piles. For this purpose, a large database containing 470 field static test results is utilized to build the ML models. The performance results of GA–ANN are compared to those of the pure ANN model (ANN-LM). Three statistical metrics including R^2, root RMS, and $a20-index$ are calculated to evaluate the prediction accuracy of those ML models. Furthermore, the influence of input variables on the LBC is evaluated using Shapley values. Lastly, a graphical user interface (GUI) program is constructed to rapidly estimate the LBC of driven piles in design practices (Fig. 1).

Database

A significant database is required to train ML models. In this study, a set of 470 tested results of driven piles are gathered from the literature (Pham & Tran, 2022; Pham et al., 2020). The result prediction is the load-bearing capacity (P_u), while all input parameters of the used database are as follows.

- Diameter of driven pile: D (mm)
- Thickness of the first soil layer where pile embedded: X_1 (mm)
- Thickness of the second soil layer where pile embedded: X_2 (mm)
- Thickness of the third soil layer where pile embedded: X_3 (mm)
- Elevation of pile top: X_p (mm)
- Elevation of natural ground: X_g (mm)
- Elevation of the extra steel pile segment: X_t (mm)
- Elevation of the pile top: X_m (mm)
- Average SPT blow along the pile shaft: N_s (mm)

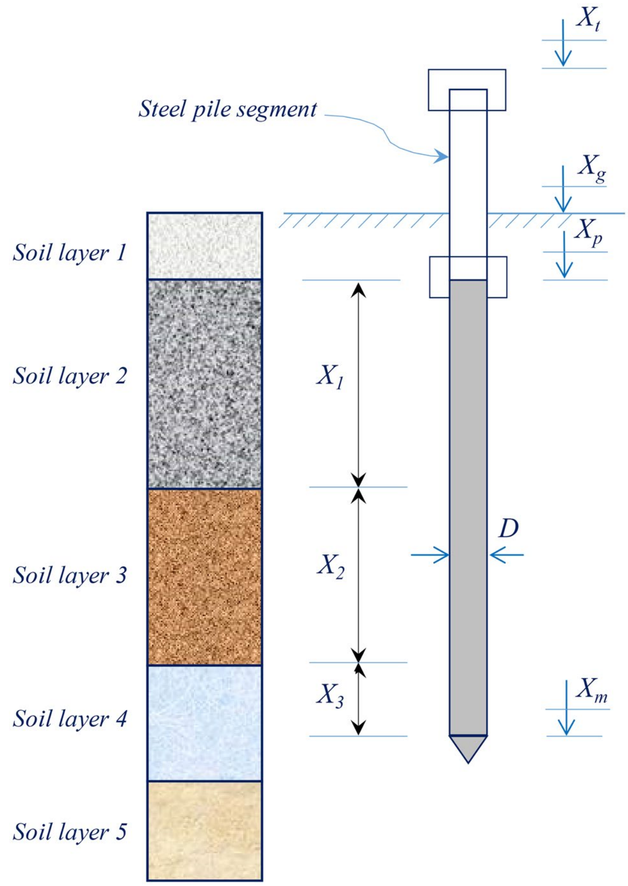


Fig. 1 Schematic soil stratigraphy and pile dimensions

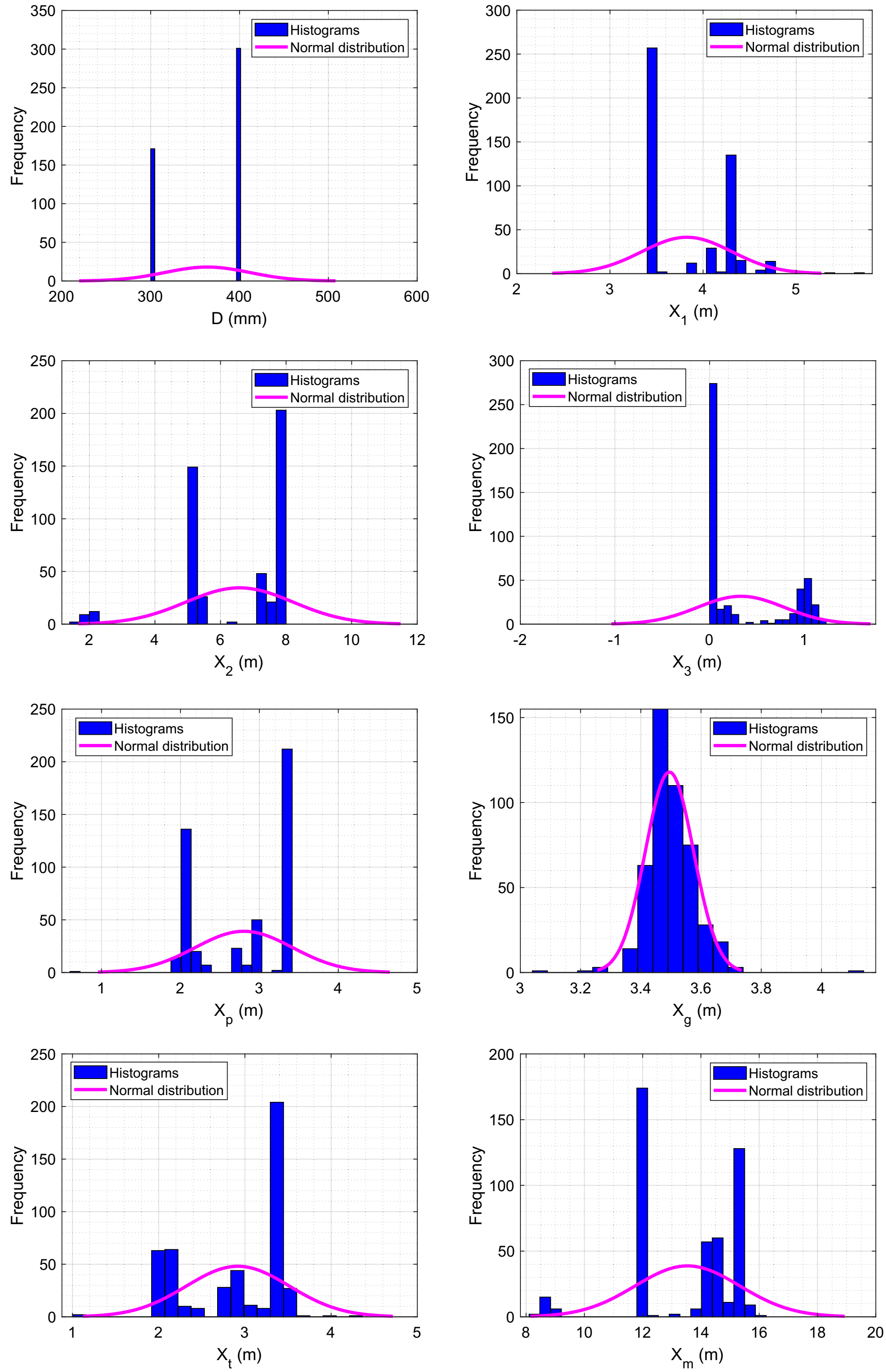
- Average SPT blow at the pile tip: N_t (mm)

The histograms of input and output parameters of 470 data sets are shown in Fig. 2. The statistical properties of the data samples are summarized in Table 1. It should be noted that this database considered the diameter of the piles from 300 to 400 mm, the maximum elevation of the pile tip was 16 m, and the maximum average SPT blow along the pile shaft was 15.4.

Hybrid GA–ANN model

So far, ANNs have been widely used to resolve different civil engineering situations (Nguyen et al., 2021a; Nguyen et al., 2021b, c; Tran et al., 2019, 2021; Zorlu et al., 2008). An ANN is a computational model inspired by the structure and functioning of biological neural networks in the human brain. It is a type of machine learning algorithm used for various tasks, including classification, regression, pattern recognition, and more. Neurons (nodes) are the basic building blocks of an ANN are artificial neurons, also known as nodes. These nodes receive inputs, perform computations,

Fig. 2 Histograms of input and output parameters in the database



and produce outputs. They are organized into layers. The three main types of layers in an ANN are:

- Input layer: The input layer receives the initial input data. Each input node represents a feature or attribute of the data.
- Hidden layers: Hidden layers are intermediate layers between the input and output layers. They perform computations on the inputs and pass the results to the next layer. ANN models can have multiple hidden layers, each with varying numbers of neurons.

Fig. 2 (continued)

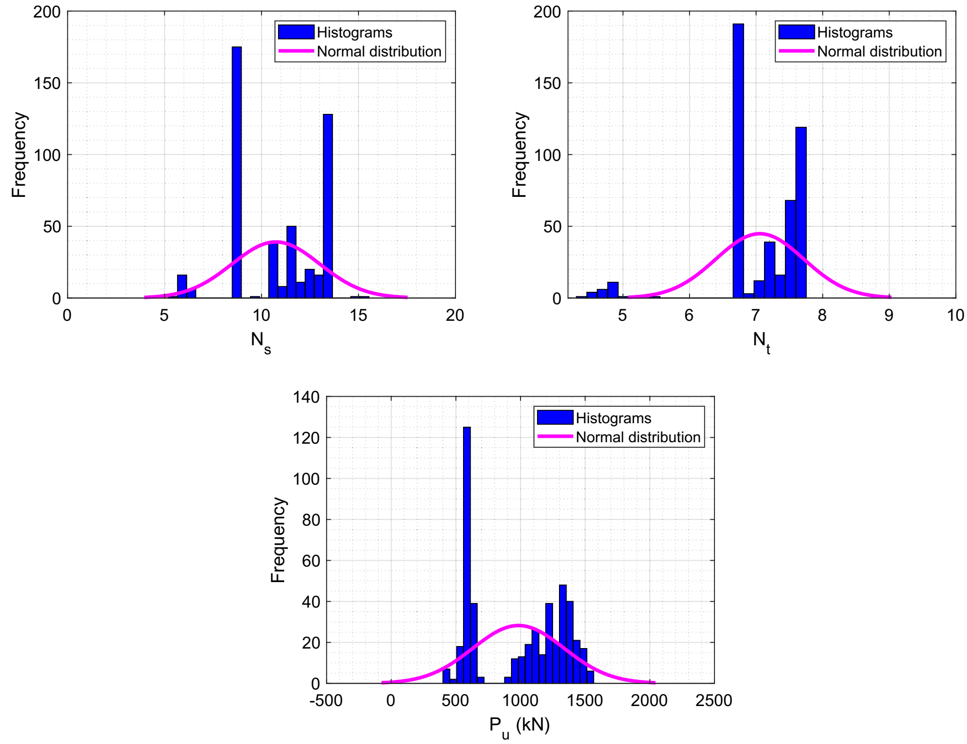


Table 1 Summary of used database

Parameter	Unit	Min	Mean	Max	SD	CoV
D	mm	300	363.77	400	48.06	0.13
*X*1	mm	3.4	3.82	5.72	0.48	0.12
*X*2	mm	1.5	6.58	8.0	1.63	0.25
*X*3	mm	0	0.33	1.69	0.45	1.37
Xp	mm	0.68	2.80	3.4	0.61	0.22
Xg	mm	3.04	3.49	4.12	0.08	0.02
Xt	mm	1.03	2.92	4.35	0.60	0.20
Xm	mm	8.3	13.53	16.09	1.80	0.13
Ns	–	5.6	10.74	15.41	2.26	0.21
Nt	–	4.38	7.05	7.75	0.66	0.09
Pu	kN	407.2	984.20	1551	352.83	0.36

- Output layer: The output layer produces the final output or prediction. The number of nodes in the output layer depends on the type of problem being solved. For example, for binary classification, there will be one output node, whereas for multi-class classification, there will be multiple output nodes.

Connections (i.e., weights and biases) between neurons represent the strength of the relationship between them. Each connection is associated with a weight and a bias, which determine the impact of the input on the output of the neuron. During training, these weights biases are adjusted to optimize the performance of the network. The mathematical expressions are shown as follows.

$$f : X \in R^D \rightarrow Y \in R^1,$$
$$f(X) = f_0\left(b_2 + W_2\left(f_h\left(b_1 + W_1 X\right)\right)\right), \tag{1}$$

where b_1, W_1, and f_h are the vector of biases, the weight matrix, and the activation function of the hidden layer, respectively. Meanwhile, b_2, W_2, and f_0 are the biases vector, the weight matrix, and the activation function of the hidden layer output layer, respectively.

Besides, each neuron applies an activation function to the weighted sum of its inputs. The activation function introduces non-linearity into the network, allowing it to model complex relationships between inputs and outputs. Common activation functions include sigmoid, tanh, ReLU, and

SoftMax. In this study, we use *tansig* and *purelin* functions for hidden and output layers, respectively. These functions are expressed by Eqs. (2) and (3).

$$tansig(x) = \frac{2}{(1 + \mathrm{epx}(-2\mathrm{x}))} - 1, \tag{2}$$

$$purelin(x) = x. \tag{3}$$

Moreover, four crucial steps are required in performing ANN models.

- Setting hyperparameters: ANN models have various hyperparameters that need to be set before training, such as the number of layers, number of neurons in each layer, learning rate, batch size, and regularization parameters. These hyperparameters impact the model's performance and need to be tuned for optimal results.
- Forward propagation: In the forward propagation step, the inputs are passed through the network layer by layer, and the weighted sum and activation function are applied at each neuron. This process continues until the output layer produces the final prediction.
- Training (backpropagation): During training, the network learns from labeled training data. Backpropagation is used to update the weights of the connections by minimizing a loss function. This process involves propagating the error backward through the network, adjusting the weights and biases to reduce the difference between predicted and actual outputs.
- Quantifying error: The loss function quantifies the discrepancy between the predicted output and the actual output. It provides a measure of how well the model is performing. Common loss functions include mean squared error (*MSE*), categorical cross-entropy, and binary cross-entropy.

Genetic algorithm (GA) (Holland, 1992) is one of the most efficient models in optimizing engineering problems (Chou & Ghaboussi, 2001; Kaveh & Kalatjari, 2002; Marasco et al., 2022). Therefore, GA is used for optimizing the ANN model for improving its prediction performance. GA–ANN is a type of ML algorithm that combines the power of ANNs with the optimization capabilities of genetic algorithms. In GA–ANN, a genetic algorithm is used to optimize the weights and biases of an artificial neural network, which is used to make predictions based on input data. The genetic algorithm works by creating a population of potential solutions, each with its own set of weights and biases. The algorithm then evaluates each solution's fitness, or how well it performs on a given task. The fittest solutions are then selected for breeding, which involves combining the weights and biases of two or more solutions to create a new generation. This process is repeated over multiple generations, with the hope that the population will eventually converge on a set of weights and biases that result in the best possible performance on the given task. In other words, the GA–ANN model is an iterative process that involves training and optimizing the ANN using GA techniques. The goal is to find the best set of weights and biases that minimize the error or maximize the accuracy of the ANN for a given problem. The basic steps to perform the GA–ANN model are as follows.

(1) Define the problem: Identify the problem to be solved and the data to be used.
(2) Build the ANN: Develop an ANN architecture that is appropriate for the problem at hand. This includes selecting the number of input and output neurons, hidden layers, and activation functions.
(3) Define the fitness function: The fitness function is used to evaluate the performance of the ANN. It is typically based on a measure of accuracy or error.
(4) Initialize the GA population: Create an initial population of solutions (ANNs) using random weights and biases.
(5) Evaluate the fitness of each solution: Apply the fitness function to each solution in the population to determine its fitness score.
(6) Select the fittest solutions: Use selection techniques (e.g., tournament selection) to choose the fittest solutions from the population.
(7) Apply genetic operators: Use genetic operators (e.g., crossover and mutation) to create new solutions from the selected fittest solutions.
(8) Evaluate the fitness of the new solutions: Apply the fitness function to the new solutions to determine their fitness scores.
(9) Repeat steps (6)–(8) until a stopping criterion is met: The stopping criterion may be a maximum number of iterations, a convergence threshold, or other criteria.
(10) Select the best solution: Choose the solution with the highest fitness score as the final solution.
(11) Test the final solution: Evaluate the performance of the final solution on a separate test set of data to assess its generalization ability.
(12) Tune the model: If necessary, fine-tune the model parameters (e.g., learning rate, activation function) to further improve performance.

Figure 3 depicts the flowchart of the GA–ANN technique.

In this study, statistical indicators, which are R^2, $RMSE$, and $a20 - index$ were employed to assess the performance of the ML model. It should be noted that the R^2 is a statistical concept that measures how well a calculated set of data matches an experimental result. In other words, it

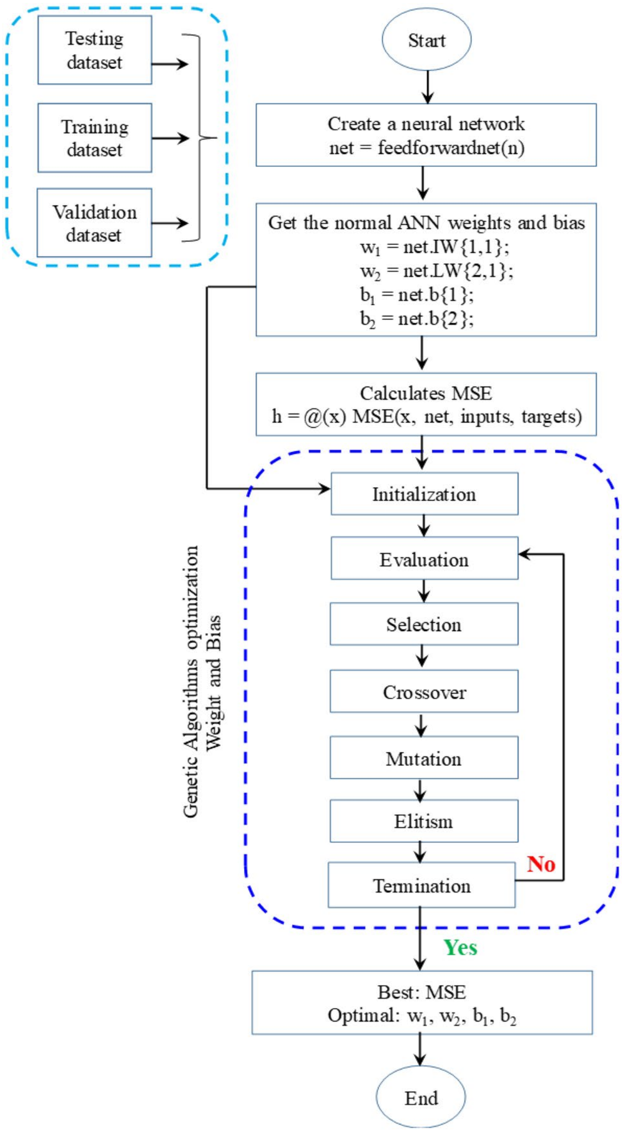


Fig. 3 Flowchart for GA–ANN model (Nguyen et al., 2023c)

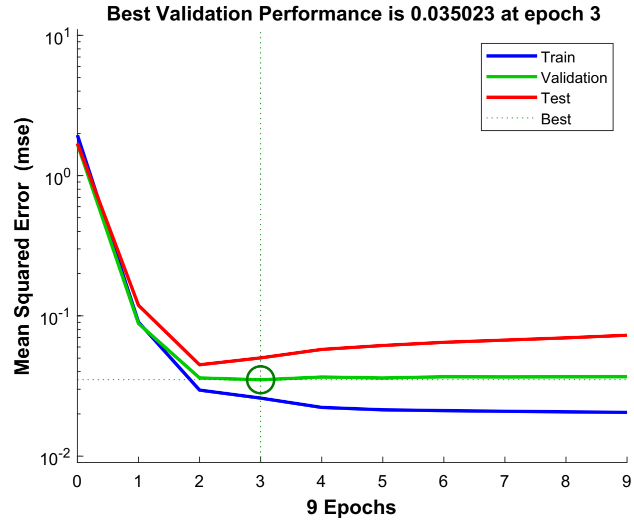


Fig. 4 Convergence of GA–ANN

evaluates how well the data fit the empirical model being used to predict the experiment. The higher the R^2, the better is the performance of the predicted model. Meanwhile, *RMSE* is a commonly used metric to evaluate the performance of a regression model. It measures the average magnitude of the differences between the predicted and actual values, providing an indication of how well the model's predictions align with the true values. *RMSE* is often used to evaluate the accuracy of a predictive model, such as in regression analysis, and is a measure of how well the model fits the data. The lower the *RMSE*, the better the model is at predicting the outcome variable. The expressions of R^2 and *RMSE* are described in the following equations.

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{n} \left(t_i - o_i\right)^2}{\sum_{i=1}^{n} \left(t_i - \overline{o}\right)^2} \right), \tag{5}$$

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^{n} \left(t_i - o_i\right)^2}, \tag{6}$$

$$a20 - index = \frac{N20}{N}, \tag{7}$$

where t_i and o_i are the actual and predicted results of the i sample; N is number of database; $\overline{o}$ is the mean of calculated results.

Results and discussion

Performance of ML model

The converged test of GA–ANN achieved the 3-epoch, and the *MSE* value was 0.035 (very close to zero), as shown in Fig. 4. Additionally, predicted regressions of GA–ANN are depicted in Fig. 5. It was found that R^2 values in training, test, validation, and all datasets were 0.99, 0.98, 0.98, and 0.99, respectively. Furthermore, the linear regression trends were very identical with the target line. This result implies that GA–ANN predicted LBC of driven piles accurately.

Tables 2 and 3 show the performance metrics (R^2, $RMSE$, and $a20 - index$) and statistical properties of the ratio $P_{Exp.}/P_{predict}$ of the hybrid GA–ANN and pure ANN (i.e., ANN-LM) models. Figure 6 compares the results of performance metrics between GA–ANN and

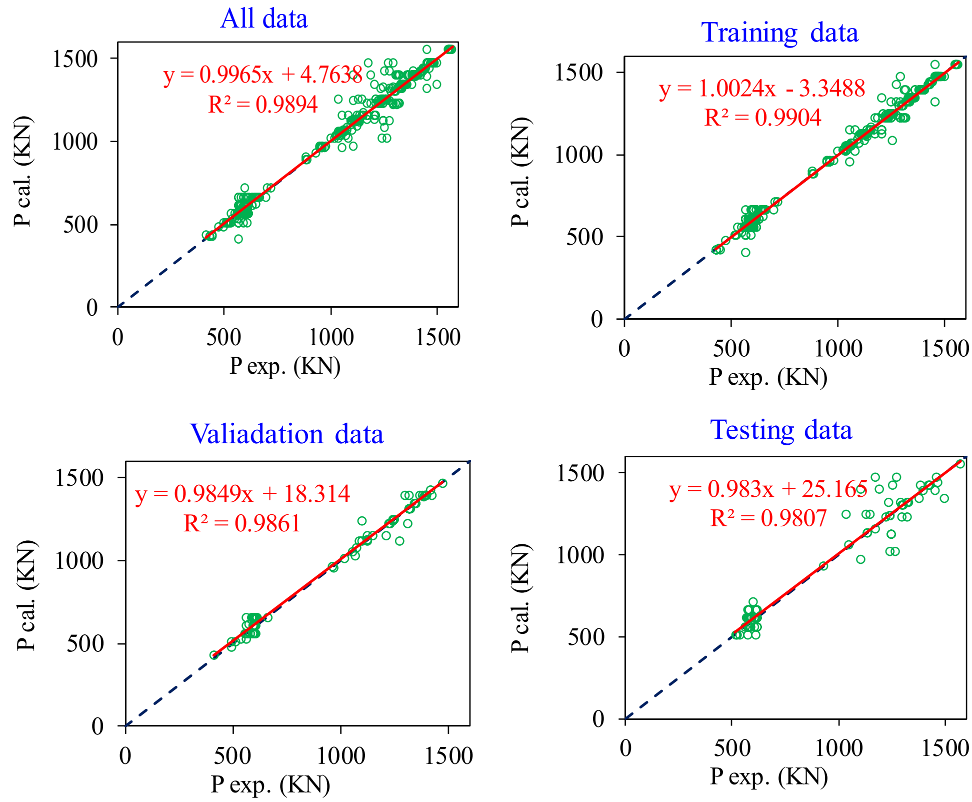


Fig. 5 Performance of GA–ANN model

Table 2 Performance of GA–ANN model for load-bearing capacity of driven piles

	R^2	$RMSE$ (kN)	$a20$ $index$	$P_{Exp.}/P_{GA-ANN}$		
				Mean	SD	CoV
All data	0.989	49.464	0.989	1.001	0.053	0.053
Training data	0.990	34.488	1.000	1.003	0.044	0.043
Validation data	0.986	41.521	1.000	1.003	0.043	0.042
Testing data	0.980	94.940	0.983	0.996	0.088	0.087

Table 3 Performance of ANN-LM for load-bearing capacity of driven piles

	R^2	$RMSE(kN)$	$a20$ $index$	P_{Exp}/P_{ANN-LM}		
				Mean	SD	CoV
All data	0.932	152.864	0.567	1.357	0.125	0.245
Training data	0.919	145.045	0.560	1.356	0.137	0.126
Validation data	0.921	151.237	0.571	1.234	0.134	0.250
Testing data	0.938	156.305	0.501	1.807	0.079	0.185

ANN-LM models. It is observed that the values of R^2 and $20-$ *index* obtained from GA–ANN were very close to unity, significantly higher than those from the ANN-LM model. Additionally, *RMSE* values of the hybrid GA–ANN model were approximately one third compared to those of ANN-LM model. Moreover, the mean values of the ratio $P_{Exp.}/P_{GA-ANN}$ were mostly close to 1.0, whereas that for the ratio $P_{Exp.}/P_{ANN-LM}$ were larger than 1.3. Once again, it can be confirmed that the GA–ANN model outperformed the pure ANN model, and the predicted results of GA–ANN were highly accurate.

Important features

The Shapley value is a concept from cooperative game theory that measures the contribution of each player in a cooperative game (Roth, 1988; Winter, 2002). It is used

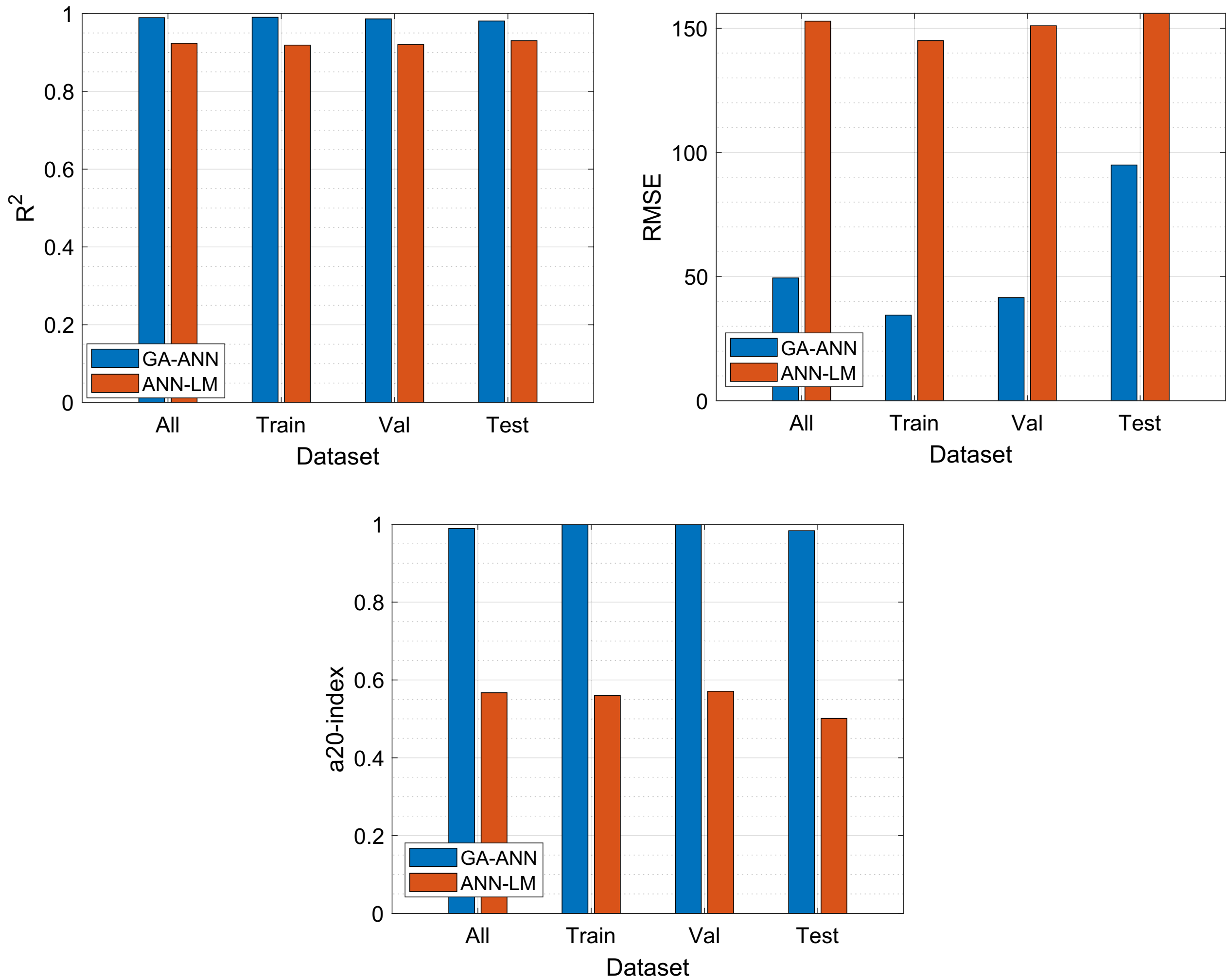


Fig. 6 Comparison of performance metrics between GA–ANN and ANN-LM models

in various fields, including economics, political science, and machine learning, to assess the importance or contribution of individual players or features. In the context of ML, the Shapley value is used to explain the importance of input features in predicting the output of a model. It helps to understand the contribution and impact of each feature on the model's predictions. By assigning a value to each feature, the Shapley value provides a fair allocation of importance among the features.

Figure 7 shows important features in calculating the LBC of driven piles using the Shapley value method. It can be found that the elevation of pile tip (X_m), diameter of piles (D), and the thicknesses of the soil layers where pile embedded (X_1 and X_2) showed to be the most influential parameters on the LBC of driven piles. Meanwhile, the elevation of natural ground (X_g) and the elevation of the top pile (X_p) had a negative influence on the LBC of the pile. Furthermore, the average SPT blow (N_t) at the tip of piles and the third soil layer where piles embedded (X_3) were shown to be less influential on the predicted LBC result.

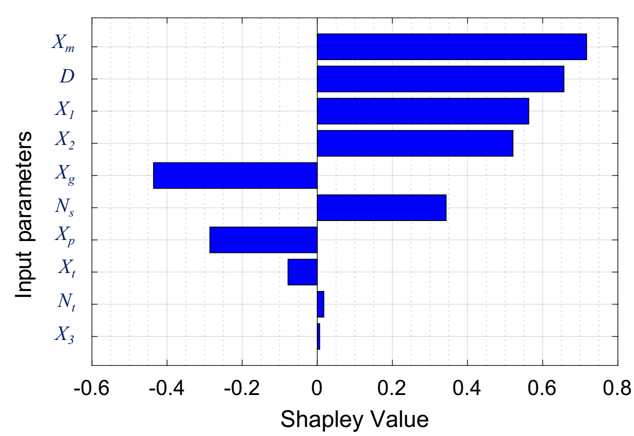


Fig. 7 Important features

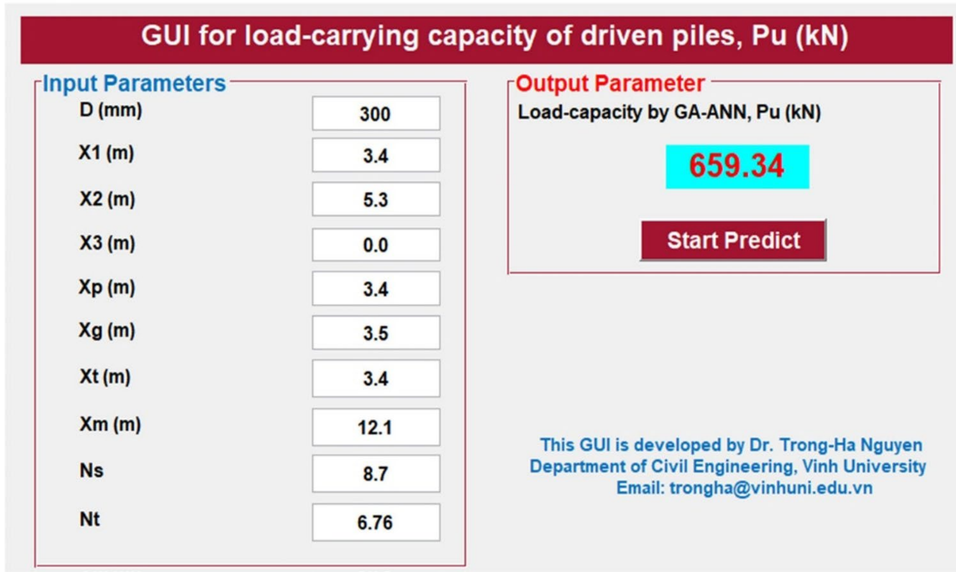


Fig. 8 GUI for calculating axial load-carrying capacity of driven piles

Practical GUI tool

For applying the ML model in design practices, it is important to transform the algorithm into practical tools (e.g., equations or GUI). Figure 8 shows the developed GUI program for rapidly predicting LBC of driven piles. This GUI can be useful for not only designers but also for technical managers since it is very easy to use. It should be noted that the verification of the algorithm was conducted and presented in Section "Performance of ML model"; therefore, the accuracy of the GUI is confirmed. Additionally, the GUI can estimate the LBC within the data provided in Table 1, a re-training process should be performed if using datasets outside the range.

Conclusions

This study developed the hybrid GA–ANN model for predicting the axial load-bearing capacity of reinforced concrete driven piles. An extensive database including 470 onsite tests was collected. The performance accuracy of the GA–ANN model was evaluated using statistical indicators, which are including the goodness of fit (R^2), root-mean-squared error ($RMSE$), $a20-index$, and mean value of the ratio $P_{experiment}$ to $P_{predict}$. The main conclusions are drawn as follows.

- GA–ANN model predicts load-bearing capacity of driven piles accurately with a very high R^2 value of 0.99 and small *RMSE* of 49 kN.
- The elevation of pile tip (X_m), diameter of piles (D), and the thicknesses of the soil layers where pile embedded (X_1 and X_2) showed to be the most influential parameters on the LBC of driven piles.
- A practical GUI was developed to rapidly predict LBC of driven piles.

Author contributions T-HN: Methodology, formal analysis, writing—review and editing; K-VTN: visualization, validation; V-CH: visualization, validation; D-DN: conceptualization, methodology, writing—original draft, writing—review and editing, supervision.

Funding No funding was used in this study.

Data availability Data request is considered by the corresponding author.

Declarations

Competing interests The authors declare no competing interests.

References

Ahmed, A., Elkatatny, S., Ali, A., Mahmoud, M., & Abdulraheem, A. (2019). New model for pore pressure prediction while drilling using artificial neural networks. *Arabian Journal for Science and Engineering, 44*, 6079–6088. https://doi.org/10.1007/s13369-018-3574-7

Barkhordari, M. S., Feng, D.-C., & Tehranizadeh, M. (2022). Efficiency of hybrid algorithms for estimating the shear strength of deep reinforced concrete beams. *Periodica Polytechnica Civil Engineering, 66*, 398–410.

Bülbül, M. A., Harirchian, E., Işık, M. F., Aghakouchaki Hosseini, S. E., & Işık, E. (2022). A hybrid ANN-GA model for an automated rapid vulnerability assessment of existing RC buildings. *Applied Sciences, 12*, 5138.

Chaabene, W. B., & Nehdi, M. L. (2020). Novel soft computing hybrid model for predicting shear strength and failure mode of SFRC beams with superior accuracy. *Composites Part C: Open Access, 3*, 100070.

Chatterjee, S., Sarkar, S., Hore, S., Dey, N., Ashour, A. S., & Balas, V. E. (2017). Particle swarm optimization trained neural network for structural failure prediction of multistoried RC buildings. *Neural Computing and Applications, 28*, 2005–2016.

Chen, X., Fu, J., Yao, J., & Gan, J. (2018). Prediction of shear strength for squat RC walls using a hybrid ANN–PSO model. *Engineering with Computers, 34*, 367–383.

Chou, J.-H., & Ghaboussi, J. (2001). Genetic algorithm in structural damage detection. *Computers & Structures, 79*, 1335–1353.

Congro, M., de Alencar Monteiro, V. M., Brandão, A. L., dos Santos, B. F., Roehl, D., & de Andrade Silva, F. (2021). Prediction of the residual flexural strength of fiber reinforced concrete using artificial neural networks. *Construction and Building Materials, 303*, 124502.

Holland, J. H. (1992). Genetic Algorithms. *Scientific American, 267*, 66–73.

Huang, J., Zhou, M., Zhang, J., Ren, J., Vatin, N. I., & Sabri, M. M. S. (2022). The use of ga and pso in evaluating the shear strength of steel fiber reinforced concrete beams. *KSCE Journal of Civil Engineering, 26*, 3918–3931.

Kaveh, A. (2014). *Advances in metaheuristic algorithms for optimal design of structures*. Springer.

Kaveh, A., & Bondarabady, H. R. (2004). Wavefront reduction using graphs, neural networks and genetic algorithm. *International Journal for Numerical Methods in Engineering, 60*, 1803–1815.

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

Kaveh, A., & Kalatjari, V. (2002). Genetic algorithm for discrete-sizing optimal design of trusses using the force method. *International Journal for Numerical Methods in Engineering, 55*, 55–72.

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

Kaveh, A., & Servati, H. (2001). Design of double layer grids using backpropagation neural networks. *Computers & Structures, 79*, 1561–1568.

Kozłowski, W., & Niemczynski, D. (2016). Methods for estimating the load bearing capacity of pile foundation using the results of penetration tests-case study of road viaduct foundation. *Procedia Engineering, 161*, 1001–1006.

Mai, S. H., Tran, V.-L., Nguyen, D.-D., Nguyen, V. T., & Thai, D.-K. (2022). Patch loading resistance prediction of steel plate girders using a deep artificial neural network and an interior-point algorithm. *Steel and Composite Structures, 45*, 159.

Marasco, G., Piana, G., Chiaia, B., & Ventura, G. (2022). Genetic algorithm supported by influence lines and a neural network for bridge health monitoring. *Journal of Structural Engineering, 148*, 04022123.

Naderpour, H., Parsa, P., & Mirrashid, M. (2021). Innovative approach for moment capacity estimation of spirally reinforced concrete columns using swarm intelligence-based algorithms and neural network. *Practice Periodical on Structural Design and Construction, 26*, 04021043.

Nanda, B., Maity, D., & Maiti, D. K. (2014). Damage assessment from curvature mode shape using unified particle swarm optimization. *Structural Engineering and Mechanics, 52*, 307–322.

Nguyen, D.-D., Tran, N.-L., & Nguyen, T.-H. (2023a). ANN-based model for predicting the axial load capacity of the cold-formed steel semi-oval hollow section column. *Asian Journal of Civil Engineering, 24*, 1165–1179.

Nguyen, D.-D., Tran, V.-L., Ha, D.-H., Nguyen, V.-Q., & Lee, T.-H. (2021). *A machine learning-based formulation for predicting shear capacity of squat flanged RC walls* (pp. 1734–1747). Elsevier.

Nguyen, H., Cao, M.-T., Tran, X.-L., Tran, T.-H., & Hoang, N.-D. (2023b). A novel whale optimization algorithm optimized XGBoost regression for estimating bearing capacity of concrete piles. *Neural Computing and Applications, 35*, 3825–3852.

Nguyen, H., Moayedi, H., Foong, L. K., Al Najjar, H. A. H., Jusoh, W. A. W., Rashid, A. S. A., & Jamali, J. (2020). Optimizing ANN models with PSO for predicting short building seismic response. *Engineering with Computers, 36*, 823–837.

Nguyen, T.-H., Tran, N.-L., & Nguyen, D.-D. (2021). Prediction of axial compression capacity of cold-formed steel oval hollow section columns using ANN and ANFIS models. *International Journal of Steel Structures*. https://doi.org/10.1007/s13296-021-00557-z

Nguyen, T.-H., Tran, N.-L., & Nguyen, D.-D. (2021). Prediction of critical buckling load of web tapered I-section steel columns using artificial neural networks. *International Journal of Steel Structures, 21*, 1–23.

Nguyen, T.-H., Tran, N.-L., Phan, V.-T., & Nguyen, D.-D. (2023c). Improving axial load-carrying capacity prediction of concrete columns reinforced with longitudinal FRP bars using hybrid GA-ANN model. *Asian Journal of Civil Engineering*. https://doi.org/10.1007/s42107-023-00695-1

Nguyen, T.-H., Tran, N.-L., Phan, V.-T., & Nguyen, D.-D. (2023d). Prediction of shear capacity of RC beams strengthened with FRCM composite using hybrid ANN-PSO model. *Case Studies in Construction Materials, 18*, e02183.

Nguyen, V.-Q., Tran, V.-L., Nguyen, D.-D., Sadiq, S., & Park, D. (2022). Novel hybrid MFO-XGBoost model for predicting the racking ratio of the rectangular tunnels subjected to seismic loading. *Transportation Geotechnics, 37*, 100878.

Pham, T. A., & Tran, V. Q. (2022). Developing random forest hybridization models for estimating the axial bearing capacity of pile. *PLoS One, 17*, e0265747.

Pham, T. A., Tran, V. Q., Vu, H.-L.T., & Ly, H.-B. (2020). Design deep neural network architecture using a genetic algorithm for estimation of pile bearing capacity. *PLoS One, 15*, e0243030.

Rahami, H., Kaveh, A., & Gholipour, Y. (2008). Sizing, geometry and topology optimization of trusses via force method and genetic algorithm. *Engineering Structures, 30*, 2360–2369.

Rönnholm, M., Arve, K., Eränen, K., Klingstedt, F., Salmi, T., & Saxén, H. (2005). ANN modeling applied to NO X reduction with octane. Ann future in personal vehicles. *Adaptive and Natural computing algorithms* (pp. 100–103). Springer. https://doi.org/10.1007/3-211-27389-1_24

Roth, A. E. (1988). Introduction to the Shapley value. *the shapley value* (pp. 1–27). Cambridge University Press.

Selvan, S. S., Pandian, P. S., Subathira, A., & Saravanan, S. (2018). Comparison of response surface methodology (RSM) and artificial neural network (ANN) in optimization of aegle marmelos oil extraction for biodiesel production. *Arabian Journal for Science and Engineering, 43*, 6119–6131. https://doi.org/10.1007/s13369-018-3272-5

Tran, N.-L., Nguyen, D.-D., & Nguyen, T.-H. (2022). Prediction of speed limit of cars moving on corroded steel girder bridges using artificial neural networks. *Sādhanā, 47*, 1–14.

Tran, N.-L., Nguyen, T.-H., Phan, V.-T., & Nguyen, D.-D. (2021). A machine learning-based model for predicting atmospheric corrosion rate of carbon steel. *Advances in Materials Science and Engineering, 2021*, 1–25.

Tran, V.-L., & Kim, S.-E. (2020). Efficiency of three advanced data-driven models for predicting axial compression capacity of

CFDST columns. *Thin-Walled Structures, 152*, 106744. https://doi.org/10.1016/j.tws.2020.106744

Tran, V.-L., & Nguyen, D.-D. (2022). Novel hybrid WOA-GBM model for patch loading resistance prediction of longitudinally stiffened steel plate girders. *Thin-Walled Structures, 177*, 109424.

Tran, V.-L., Thai, D.-K., & Kim, S.-E. (2019). Application of ANN in predicting ACC of SCFST column. *Composite Structures, 228*, 111332. https://doi.org/10.1016/j.compstruct.2019.111332

Vakhshouri, B., & Nejadi, S. (2018). Prediction of compressive strength of self-compacting concrete by ANFIS models. *Neurocomputing, 280*, 13–22. https://doi.org/10.1016/j.neucom.2017.09.099

Vijayakumar, R., & Pannirselvam, N. (2022). Multi-objective optimisation of mild steel embossed plate shear connector using artificial neural network-integrated genetic algorithm. *Case Studies in Construction Materials, 17*, e01560.

Winter, E. (2002). The shapley value. *Handbook of Game Theory with Economic Applications, 3*, 2025–2054.

Yang, H., Akiyama, T., and Sasaki, T. (1992). A neural network approach to the identification of real time origin-destination flows from traffic counts

Zorlu, K., Gokceoglu, C., Ocakoglu, F., Nefeslioglu, H., & Acikalin, S. (2008). Prediction of uniaxial compressive strength of sandstones using petrography-based models. *Engineering Geology, 96*, 141–158. https://doi.org/10.1016/j.enggeo.2007.10.009