Application of the neural network model in predicting the threedimensional response of suction caissons on clay

Yin, Xilin; Wang, Huan; Pisano, Federico; Gavin, Ken; Askarinejad, Amin; Zhou, Hongpeng

DOI
10.3723/LSLC8955

Publication date
2023

Document Version
Final published version

Published in
Offshore Site Investigation Geotechnics 9th International Conference Proceeding

doi:10.3723/LSLC8955

Application of the neural network model in predicting the three-dimensional response of suction caissons on clay

Xilin Yin
Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, 2628 CN, The Netherlands

Huan Wang
Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, 2628 CN, The Netherlands
Advanced Modelling Section, Norwegian Geotechnical Institute, Oslo, 0484, Norway

Federico Pisanò, Ken Gavin & Amin Askarinejad
Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, 2628 CN, The Netherlands

Hongpeng Zhou
Department of Computer Science, Faculty of Science and Engineering, Imaging and Data Sciences, The University of Manchester, Manchester, M13 9PL, United Kingdom

ABSTRACT: Predicting the nonlinear load response of caisson foundations is critical to the foundation design. Despite extensive studies aimed at developing models for predicting the combined V-H-M bearing capacity of suction caissons in clay, accurately predicting the three-dimensional (3D) deflection response of the foundation remains a significant challenge. In this paper, we present a novel solution by developing a fully connected (FC) neural network model that enables load-deflection prediction of suction caissons on clay. To train and evaluate the FC model, a series of 3D finite element simulations were performed covering caissons responses with an embedment ratio of up to 1. The effect of various model hyperparameters on the model's prediction accuracy and generalisation ability was systematically investigated. The results show that the proposed model achieves load-deflection response prediction with simplicity, efficiency and accuracy, demonstrating the significant potential of deep learning technology in the geotechnical design of foundations.

1 Introduction

Suction caissons are widely used as foundations or anchors in the oil and gas industry (Randolph and Gourvenec, 2017). Recently, the suction caisson has gained more popularity as the foundation for both bottom-fixed and floating wind turbines. Due to complex environmental actions such as wind, waves and currents, substructures of offshore facilities experience three-dimensional (3D) loads. It is therefore essential to understand the load-deflection response of foundations and develop accurate and efficient design strategies. Typically, for traditional foundation design, the primary consideration is the ultimate bearing capacity. The combined foundation capacity under complex vertical (V), horizontal (H) and moment (M) loads is normally represented by a failure envelope (Roscoe, 1956). The failure envelope of suction caisson foundations has been extensively studied in both drained sand and undrained clay (Houlsby et al., 2005; Bransby and Yun, 2009; Gourvenec and Barnett, 2011; Vulpe, 2015). It is important to note that the deflection of the foundation required to mobilize the bearing capacity is usually quite significant and exceeds the service limit condition. Moreover, external loads on offshore wind turbines are relatively small compared to the foundation's ultimate bearing capability. Therefore, instead of the ultimate limit state, foundation design is usually governed by the stiffness at small deflection (Byrne et al., 2002). As a result, it is crucial to predict the nonlinear load-deflection response of the foundation accurately.

Traditionally, macro-element models are usually used to model the nonlinear load-deflection response of a foundation under 3D loads. For example, Cassidy et al. (2006), Skau et al. (2018) and Yin et al. (2020) developed macro-element models based on plasticity theory, hypoplastic theory and multi-surface concept. However, the deflection response of suction caisson is strongly dependent on the geometrical configuration of the foundations (e.g., embedment ratio L/D, where D is foundation diameter and L is the foundation embedment length) (Zhang et al., 2014) and the geotechnical properties of the seabed (e.g. the stiffness and strength) (Cremer et al., 2002), making it very challenging to predict its response with a single macro-element model (Skau et al., 2018). Alternatively, finite element (FE) modelling can simulate explicitly the soil-foundation system and predict the foundation behaviour under complicated loads. However, given the requirement of expert knowledge and computational resource, FE analysis is less preferable in industry design (Szabo and Babuska, 2021; Houlsby, 2016). For example, in the design of offshore wind turbines, extensive simulations (e.g., thousands of cases) will be conducted to obtain the load inputs for foundation design. This process can be interactive for several rounds until getting the final

design. Running all these simulations using 3D FE is unfeasible (Feng and Shen, 2017). Therefore, a simpler and more effective model that retains the FE model's accuracy and flexibility needs to be developed.

An alternative technique for foundation design has recently emerged: the deep learning (DL), which has a powerful ability to handle nonlinear regression problems (Reimers and Requena-Mesa, 2020). Compared to traditional design approaches, such as the macro-element model, the DL technique stands out for its high degree of adaptability, flexibility, and lack of pre-geotechnical assumptions. Benefiting from its powerful nonlinear mapping ability, the DL technique has been successfully applied in many geotechnical problems (Baghbani, 2022). For example, Zhang et al. (2020) developed a surrogate model using the long short-term memory (LSTM) model to predict the two-dimensional load-deflection response of suction caisson foundations. In light of these premises, this study aims to develop an optimal DL-based model to predict the nonlinear response of suction caissons under 3D loads. The fully connected neural network (FC) is used for its simplicity. A series of 3D FE simulations were performed to provide the training database for the FC model. The influence of the selection of FC model hyperparameters on the model performance was investigated.

2 Database preparation

This study focuses on the response of suction caissons in undrained clay under combined loads in the same plane. The finite element software Abaqus 6.14 was used to obtain the foundation deflection response under combined loads (Systemes, 2014). The suction caissons are thin-walled, large-diameter steel cylinders with an open bottom and closed top. In this study, a fixed foundation diameter and wall thickness of 10 m and 0.1 m, respectively, were adopted. (Villalobos, 2006). Since all the calculated data will be normalized with the foundation dimensions and soil undrained strength, the absolute value of the diameter is expected not to affect the results (Gourvenec and Barnett, 2011). A total of five embedment depth-to-diameter ratios (L/D = 0.2, 0.4, 0.6, 0.8 and 1.0) were considered to cover the suction caisson foundations used in the field for bottom fixed offshore wind turbines (Fu et al., 2020). Considering the symmetry of the problem, only half of the soil-foundation system was modelled to save computation time. Roller supports were applied around the mesh circumference 7D away from the caisson walls, while the base boundary was fully fixed at 4D from the caisson bottom. A fine mesh domain and a coarse mesh domain were applied around the foundation skirts and in the far field, respectively, to save computational costs. The boundary conditions, mesh distributions and foundation configurations in this study are illustrated in Figure 1. A fully rough interface with no separation between the foundation and the soil was used in all simulations. The suction caisson and clay soil were modelled with eight-node linear strain brick elements with reduced integration (i.e., 'C3D8R' in Abaqus terminology) and hybrid eight-node linear strain brick elements (i.e., 'C3D8H' in Abaqus terminology), respectively (Gourvenec and Barnett, 2011).

In this study, the clay soil was assumed to be a linearly elastic and perfectly plastic material that satisfies the Tresca criterion. The soil modulus, denoted by G, was determined from the undrained strength (s_u) and was equal to 500s_u (Hu and Randolph, 1998; Jeanjean et al., 2017). The effective unit weight of the soil was $\gamma' = 6\text{kN/m}^3$. To simulate the undrained loading conditions, Poisson's ratio υ of 0.495 and a dilation angle of 0.01° were used. A homogeneous shear strength soil profile with s_u = 10kPa, representing over-consolidated clay, was considered in this study. The suction caisson foundations were assumed to have a fully elastic response, whose mechanical properties were defined by Young's modulus E and Poisson's ratio υ of steel. Table 1 summarizes all the mechanical properties used in the FE modelling.

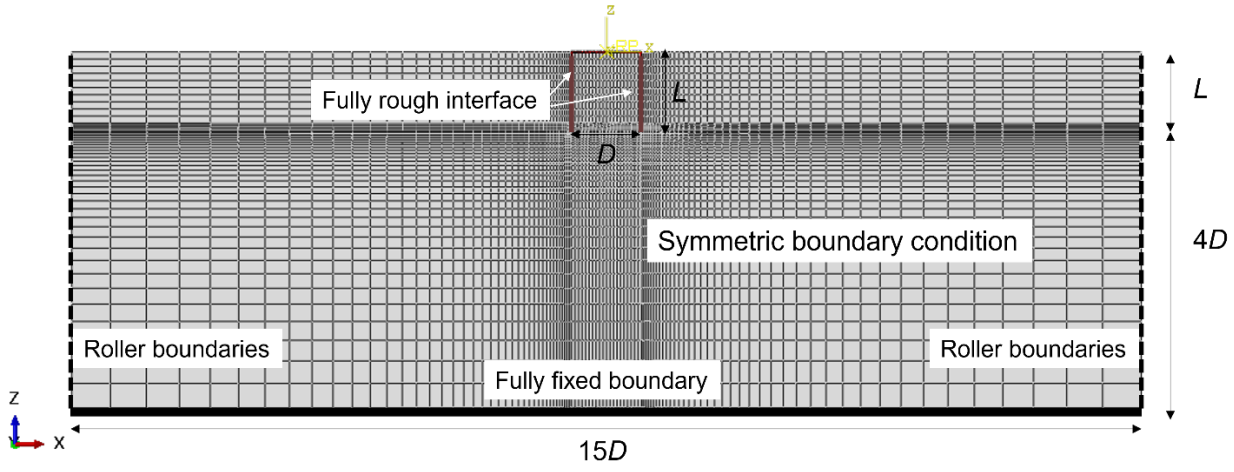


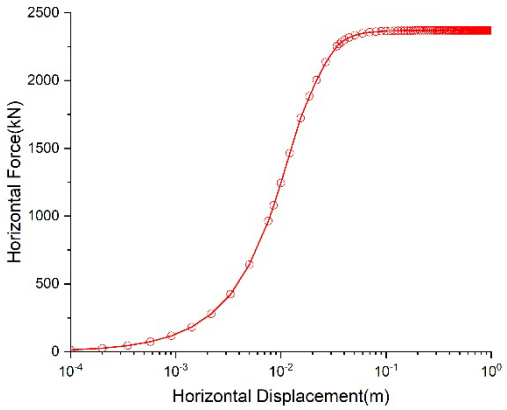
Figure 1 Boundary conditions of the FE model

Table 1 Mechanical properties in FE modelling

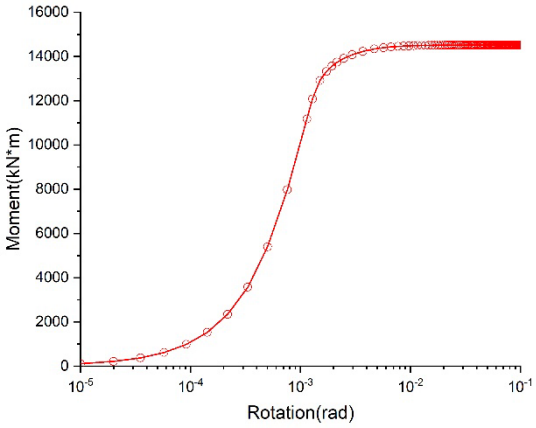
Suction	Outer diameter (D)	10 m
	Length (L)	1–10 m
Clay	Angle of internal friction	0.01°
	Angle of dilation (ψ)	0.01°
	Shear modulus to undrained	500
	Poisson's ratio (v_s)	0.495
	Shear strength in OC clay	10 kPa
Steel	Young's modulus (E_s)	210 GPa
	Poisson's ratio (v_s)	0.25

To obtain the complete response of the foundation from the small to larger deflection, displacement-controlled simulations were performed. As shown in Figure 2, under the given loading direction, the foundation has reached the horizontal and moment bearing capacity, while the vertical resistance can still increase. A final vertical displacement of 1 m (10%D) is adopted in this study.

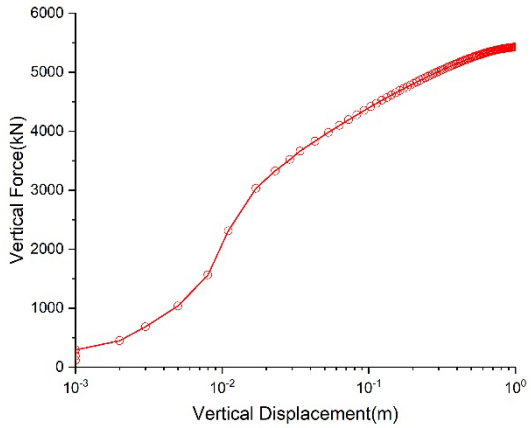
To further obtain the foundation response in a three-dimensional VHM space, 96 distinct displacement loading paths were simulated for each foundation, with 100 data points captured in each direction. This resulted in 9600 data points for each foundation, with each data point comprising displacement information in three directions (v, u, θ) corresponding to three load components (V, H, M) and the foundation configuration (L/D). In the subsequent training, the embedment ratio L/D was used instead of the absolute values of D and L to accommodate as many combinations of practical embedment depths as possible. This approach has the advantage of disregarding the influence of the absolute foundation dimensions. The displacement information of the skirted foundation was stored directly in comma-separated values (CSV) files without any processing. The deflection and load information of the skirted foundation were represented by the dimensionless factors (v/D, u/D, θ) and (V/As_{u0}, H/As_{u0}, M/ADs_{u0}, $A = \pi D^2/4$ is the foundation area) in the files, respectively. The numerically generated data was then used to train the FC model in Section 3. To further investigate model generalisation ability, twelve new displacement loading paths were simulated for three new embedment depths foundations (L/D = 0.23, 0.45 and 0.79), with 100 data points captured in each direction as a backup dataset.



(a) Horizontal response



(b) Rotational response



(c) Vertical response

Figure 2 Typical foundation load-deflection response (L/D = 1)

3 Deep learning based methodology

3.1 *Fully connected neural network*

In this paper, the fully connected neural network (FC) is used for the prediction of nonlinear regression (Bishop and Nasrabadi, 2006). The FC model consists of an input layer, one or more hidden layers and an output layer. The weights, biases and activation functions are used to transfer information between these layers (e.g., Figure 3). FC models are powerful tools

for the nonlinear regression problem, but designing an appropriate structure and learning the parameters can be difficult. Specifically, the model structure (e.g., size and complexity), controlled by hyperparameters (e.g., the number of neurons and hidden layers), is an innate determinant of the model's ability. An under-parameterized model can lead to underfitting, while an over-parameterized model can lead to overfitting. The performance of the model is also affected by other hyperparameters such as the learning rate, batch size and training iterations. Therefore, the final performance (e.g., convergence, training efficiency) of a model is highly dependent on the setting of the hyperparameters and we will elaborate on the ways to tune hyper-parameter in the next section.

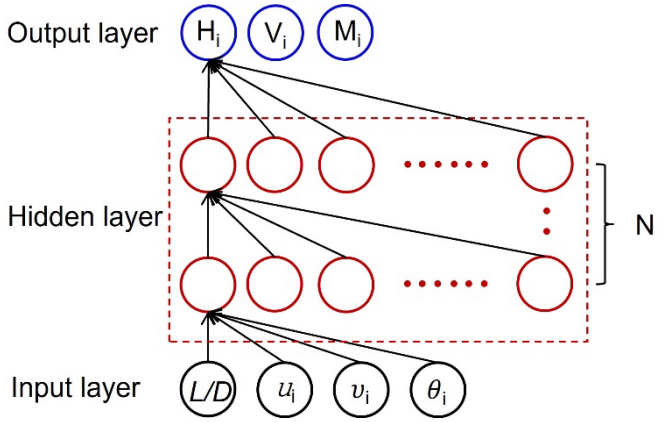


Figure 3 Fully connected neural network

In this study, the RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) are used as the indication of regression results error. A smaller RMSE or MAE represents a better prediction. The coefficient of determination (R^2) represents the goodness of fit of the model predictions. A closer R^2 to 1 means a better fit. Three evaluation metrics are calculated as follows:

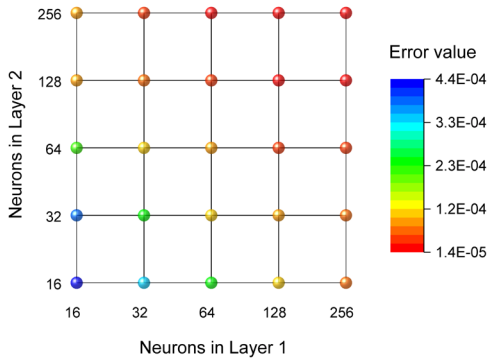
$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (1)$$

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m}|(y_i - \hat{y}_i)| \quad (2)$$

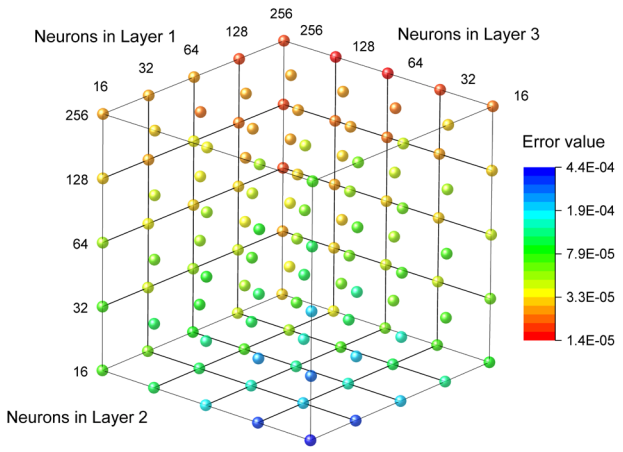
$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2} \quad (3)$$

3.2 *Accuracy driven hyper-parameters tuning*

To implement this hyper-parameter tuning, Keras framework, a Python-based deep learning toolkit (Gulli and Pal, 2017), is adopted. The grid search technique is employed to draft the model structure hyper-parameters. As per the universal approximation theorem (Hornik et al., 1989), the neural network can be considered as a 'universal' function and a one-hidden-layer fully connected (FC) neural network model can provide satisfactory prediction performance. We evaluate a more complex neural network with multiple hidden layers to enhance prediction stability and robustness. Furthermore, to approximate the nonlinear regression model, the nonlinear activation function will be introduced after each hidden layer. In this study, the rectified linear unit (ReLU) activation function is introduced after each hidden layer to overcome the gradient vanishing issue and speed up the gradient descent convergence. It should be noted that despite increasing the model's depth can result in a more accurate function fit and a broader solution space, an overparameterized model can lead to overfitting. Therefore, this study aims to strike a balance between model complexity and accuracy. The FC neural network models with two and three hidden layers, combining 16, 32, 64, 128 and 256 neurons per layer, were tested. The minimum mean squared errors in each combination were extracted and shown in Figure 4.



(a) Predictions error with two hidden layers model



(b) Predictions error with three hidden layers model

Figure 4 The minimum loss of the model with different combinations of neurons

Analysis of the results in Figure 4(a) reveals that the model performance improves as the number of neurons increases. However, it should be noted that beyond a certain point, the increase in the number of neurons does not contribute significantly to the

model's accuracy improvement, as observed in the marginal diminishing effect. Figure 4(b) presents a similar pattern of variation in the three hidden layers model, where increasing the number of neurons or multiplying the weight parameters does not lead to a significant accuracy gain. Furthermore, it is essential to note that increasing the number of neurons may decrease the model's accuracy, as observed in the case of the (32, 256, 128) neuron combination, which outperforms the (64, 256, 128) neuron combination. This indicates redundancy in the model's nodes and increasing neurons may weaken the model generalisation ability, leading to overfitting and trapping the model in a local optimum. Therefore, it is concluded that the optimal model structure has two hidden layers, and each layer should have 256 neurons.

Then we focus on choosing the batch size and optimizer, specifically determining the learning rate and training epoch. To achieve this, a preliminary experiment is performed, and it is found that smaller batch sizes resulted in better prediction accuracy but a weaker model generalisation. Considering acceptable loss values and generalisation capabilities, a batch size of 128 data points is selected. For the optimizer, the Adam algorithm (adaptive moment estimation algorithm) (Kingma and Ba, 2014), which combines the momentum method and the RMSprop algorithm, is adopted. The Adam algorithm uses momentum as the direction of parameter update and adaptively changes the learning rate. A comparison experiment is carried out then to determine the initial learning rate. Figure 5 shows the loss values during the training process at three different learning rates. It is clear that the training loss remained stable at all learning rates and converged to a constant value within 50 epochs. The higher the learning rate, the faster the model convergence (within 5 epochs), but at the cost of stability (with significant fluctuations). In the end, a learning rate of 0.001 is set in the model training, as it could balance rapid convergence and stability.

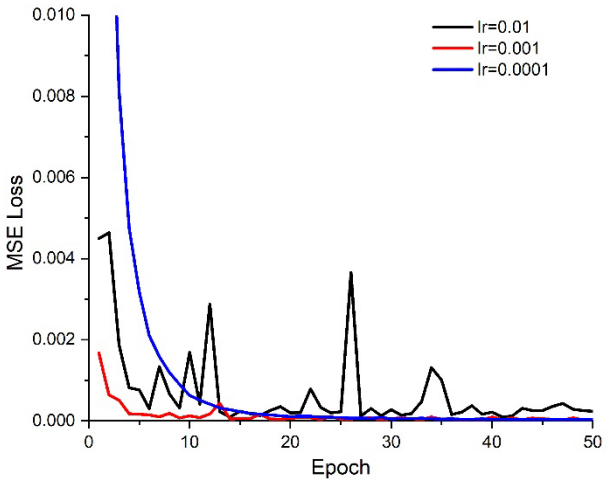


Figure 5 Training loss for the DNN model with different learning rates

3.3 *Generalisation validation*

Regularisation is a crucial technique to improve the generalisation performance of neural networks by reducing model complexity and preventing overfitting (Tian and Zhang, 2022). Commonly used regularisation methods for deep learning models include data augmentation, early stopping (Prechelt, 2012), dropout (Srivastava et al., 2014), among others. In the FC model, we have employed regularisation methods such as early stopping (limiting training epoch to 50) and weight decay using the Adam optimizer (Hanson and Pratt, 1988).

This section aims to further investigate whether additional regularisation methods, such as dropout, can enhance the model's generalisation ability. To evaluate the generalisation performance and model complexity, three new datasets with caisson embedment ratios of L/D = 0.23, 0.45 and 0.79 were generated. Each dataset consists of 1200 data points captured in twelve different loading directions on the new embedment ratio caisson.

The dropout method is a technique used to prevent overfitting by randomly dropping some neurons during training. In this study, we experimented with different dropout rates ranging from 0 to 0.9 to find the optimal rate for improving model generalisation. Figure 6 shows the prediction errors of the model on three datasets with different dropout rates (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9). The model performed consistently across all three datasets, with prediction error increasing as the dropout rate increased. This result indicates that the model can effectively overcome overfitting without the need for additional dropout layers between the hidden layers.

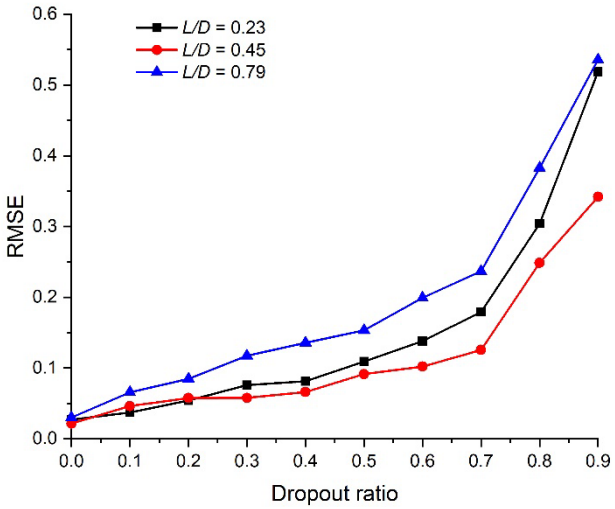


Figure 6 Generalisation ability under different dropout ratios

3.4 *Model complexity validation*

Variance and bias are crucial components of error that indicate the generalisation and fitting ability of the

model, respectively (Fortmann-Roe, 2012). As the model complexity increases, the prediction results become less biased and have more variance, as illustrated in Figure 7. This means that the model's ability to fit the data gradually increases at the cost of generalisation. However, having too high or too low a model complexity can result in a large prediction error. Therefore, it is necessary to validate whether the model complexity is in the appropriate range, where the model can balance its generalisation and fitting abilities. It is important to note that we are not interested in acquiring a model that is solely good at fitting the given training data.

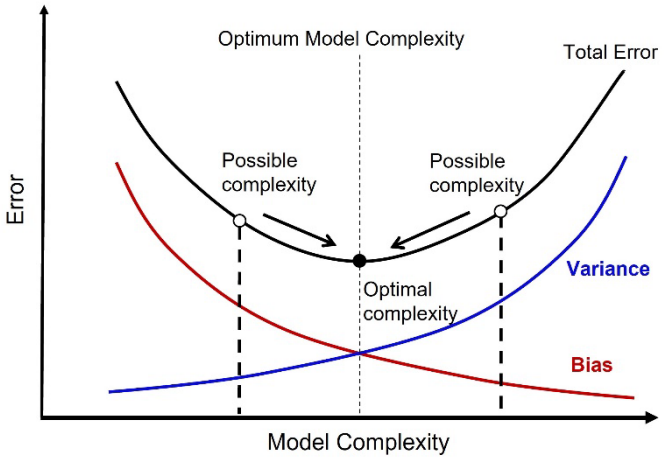
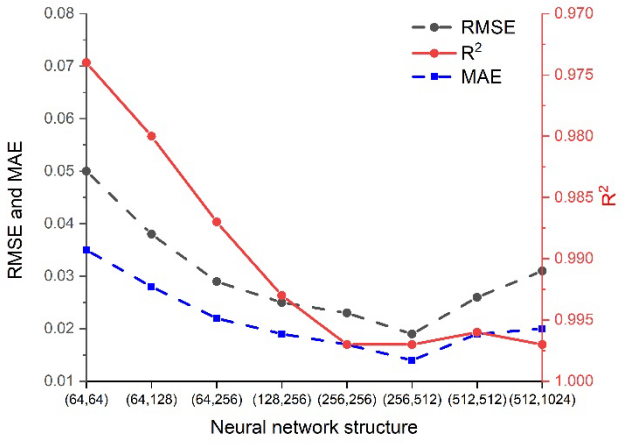
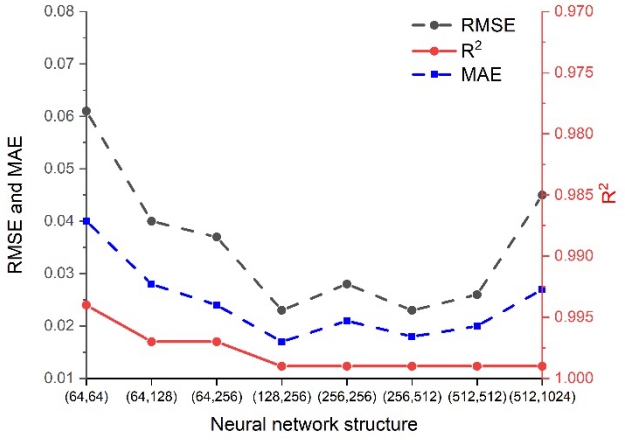


Figure 7 Effect of model complexity on prediction error

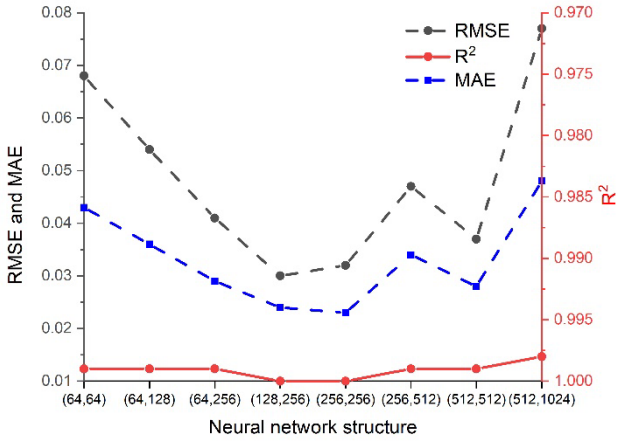
Therefore, eight experiments are conducted to control model complexity by increasing or decreasing the number of neurons. In these experiments, an FC neural network with 256 neurons in each hidden layer is denoted as (256, 256). The simplest neural network structure is represented by (64, 64) and the most complex by (512, 1024). The prediction errors are illustrated in Figure 8. The three error evaluation metrics show a similar trend of first decreasing and then increasing as the model complexity increases. The current model complexity of (256, 256) is located at the error curve trough. However, it should be noted that the optimal model complexity is not constant, which decreases as the embedment ratio increases. This phenomenon implies that the underlying mechanism of data is more complex and variable at a smaller embedment ratio, requiring a more complex model.



(a) Impact of model complexity on generalisations at $L/D = 0.23$



(b) Impact of model complexity on generalisations at $L/D = 0.45$



(c) Impact of model complexity on generalisations at $L/D = 0.79$

Figure 8 Impact of model complexity on generalisations at three new embedment depths

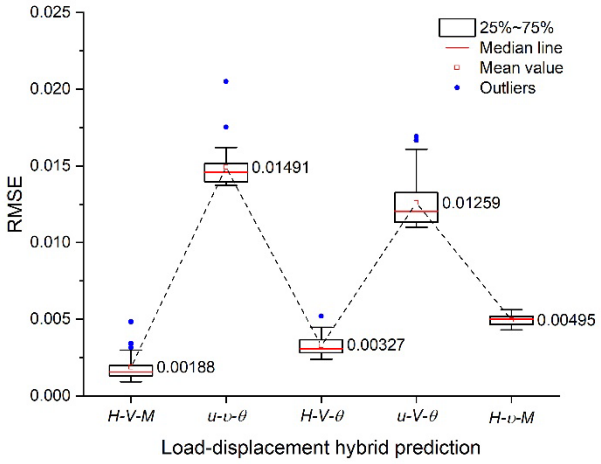
In summary, the optimal model complexity does not change significantly with the embedment ratio and remains around (256, 256). This demonstrates that a model of this complexity generalises well and is capable of handling different embedment depths in the caisson. The final hyperparameters utilized for FC neural network training are summarized in Table 2.

Table 2 Hyper-parameters used for training the neural network model

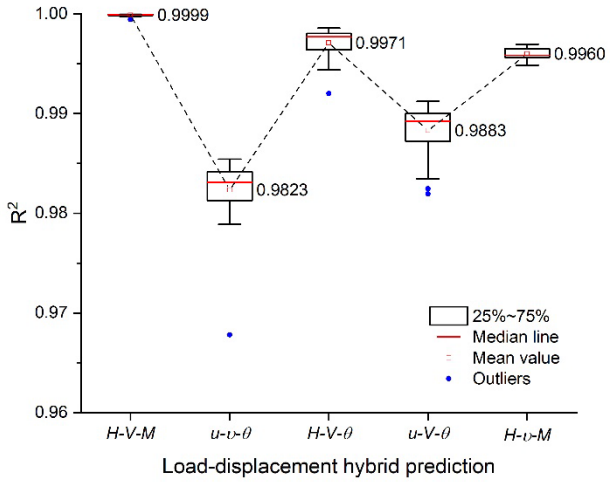
Hyperparameter	Description	Value
N_h	Number of hidden layers	2
N_n	Number of nodes in the hidden layer	256,256
η	Learning rate in the optimizer	0.001
Batch size	Number of training samples	128
Epoch	Number of iterations during training	50
validation split	Proportion of validation set in total training	0.2

3.5 *Evaluation of model performance*

The previous section presents a thorough discussion on the basis of the selected model hyperparameters (listed in Table 2), based on which the model will be constructed. The training set comprises 80% of the total dataset with 48000 data points from 5 embedment ratios described in Section 2. While the test set consists of the left 20% dataset which are 9600 randomly selected data points from all embedment depths of the caisson response. After training, the interpolation prediction error on the test set was found to be RMSE = 0.032, R^2= 1.000 and MAE = 0.022. The FC neural network model was then tested for its ability to generalise and predict the caisson response for a specific embedment ratio, where L/D = 0.7. The prediction accuracy was only slightly reduced to RMSE = 0.036, R^2= 0.999 and MAE = 0.026, demonstrating the outstanding interpolation generalisation performance of the model. The good generalisation and minor computational errors in the prediction attest to the promise of this model. Meanwhile, model performance in load-displacement hybrid predictions is tested subsequently. Four typical load-displacement combinations were fed into the model compared with only displacement input. Box plots demonstrate the variance of model performance over 30 repeatable experiments in Figure 9, where the horizontal axis represents the expected predicted parameter combination. The model takes the force or/and displacement corresponding to the predicted parameter as input and shows the most accurate prediction when predict (H, V, M) combination (i.e. pure displacement input with embedment ratio). By contrast, the model does not perform as well as the (H, V, M) combination when input mechanical parameters and output displacement (i.e. (v, u, θ) combination). The reasonable explanation of this phenomenon is that the loading path under displacement control captures the ultimate state response, which implies that the displacement is one-to-one mapping to the force, but not vice versa. The predicted performance of other combinations falls somewhere in between, also with relatively high accuracy. Comparative experiments demonstrate that the model is capable of handling various combinations of force-displacement inputs and exhibits high accuracy in the corresponding prediction.



(a) RMSE in five hybrid load-displacement response predictions



(b) R^2 in five hybrid load-displacement response predictions

Figure 9 Box plots showing model prediction error distributions in five hybrid load-displacement responses

4 Conclusion

This paper presents a FC neural network model to predict the 3D deflection response of caisson foundations with various geometry configurations in clay. The model is trained on the basis of the foundation response of five embedment ratios and can effectively predict the load-deflection response of the provided embedment ratio within seconds. After hyperparameter tuning and complexity validation, the model has the capacity to accurately capture the complete response of the foundation at a specific embedment ratio with negligible prediction error. This successful application demonstrates the efficiency and competitiveness of the FC neural network model in caisson design. Furthermore, the

model can be further developed to consider more complex soil properties than just the uniform soil layer. The model applicability will therefore expand as more comprehensive datasets become available.

5 References

Baghbani, A. (2022). Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. page 26.

Bishop, C. M. and Nasrabadi, N. M. (2006). Pattern recognition and machine learning, volume 4. Springer.

Bransby, M. and Yun, G.-J. (2009). The undrained capacity of skirted strip foundations under combined loading. Geotechnique, 59(2):115—125.

Byrne, B., Houlsby, G., Martin, C., and Fish, P. (2002). Suction caisson foundations for offshore wind turbines. Wind Engineering, 26(3):145-155.

Cassidy, M., Randolph, M., and Byrne, B. (2006). A plasticity model describing caisson behaviour in clay. Applied Ocean Research, 28(5):345–358.

Cremer, C., Pecker, A., and Davenne, L. (2002). Modelling of nonlinear dynamic behaviour of a shallow strip foundation with macro-element. Journal of Earthquake Engineering, 6(02):175–211.

Feng, J. and Shen, W. Z. (2017). Design optimization of offshore wind farms with multiple types of wind turbines. Applied Energy, 205:1283–1297.

Fortmann-Roe, S. (2012). Understanding the bias-variance tradeoff. URL: http://scott. fortmann-roe. com/docs/BiasVariance. html (hamtad 201903-27).

Fu, D., Zhang, Y., and Yan, Y. (2020). Bearing capacity of a side-rounded suction caisson foundation under general loading in clay. Computers and Geotechnics, 123:103543.

Gourvenec, S. and Barnett, S. (2011). Undrained failure envelope for skirted foundations under general loading. Geotechnique, 61(3):263–270.

Gulli, A. and Pal, S. (2017). Deep learning with Keras. Packt Publishing Ltd.

Hanson, S. and Pratt, L. (1988). Comparing biases for minimal network construction with back-propagation. Advances in neural information processing systems, 1.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. Neural networks, 2(5):359–366.

Houlsby, G. (2016). Interactions in offshore foundation design. Geotechnique, 66(10):791–825.

Houlsby, G. T., Ibsen, L. B., and Byrne, B. W. (2005). Suction caissons for wind turbines. Frontiers in Offshore Geotechnics: ISFOG, Perth, WA, Australia, pages 75–93.

Hu, Y. and Randolph, M. (1998). A practical numerical approach for large deformation problems in soil. International Journal for Numerical and Analytical Methods in Geomechanics, 22(5):327–350.

Jeanjean, P., Zhang, Y., Zakeri, A., Andersen, K., Gilbert, R., and Senanayake, A. (2017). A framework for monotonic py curves in clays. In Offshore Site Investigation Geotechnics 8th International Conference Proceeding, volume 108, pages 108–141. Society for Underwater Technology.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Prechelt, L. (2012). Early stopping — but when? Neural networks: tricks of the trade: second edition, pages 53–67.

Randolph, M. and Gourvenec, S. (2017). Offshore geotechnical engineering. CRC press.

Reimers, C. and Requena-Mesa, C. (2020). Deep learning-an opportunity and a challenge for geo-and astrophysics. In Knowledge Discovery in Big Data from Astronomy and Earth Observation, pages 251–265. Elsevier.

Roscoe, K. (1956). The stability of short pier foundations in sand. British Welding Journal, pages 343–354.

Skau, K. S., Grimstad, G., Page, A. M., Eiksund, G. R., and Jostad, H. P. (2018). A macro-element for integrated time domain analyses representing bucket foundations for offshore wind turbines. Marine Structures, 59:158–178.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from over fitting. The journal of machine learning research, 15(1):1929–1958.

Systemes, D. (2014). Abaqus 6.14 — abaqus analysis user's guide. Dassault Systemes: Velizy-Villacoublay, France.

Szabo, B. and Babuska, I. (2021). Finite element analysis: Method, verification and validation.

Tian, Y. and Zhang, Y. (2022). A comprehensive survey on regularization strategies in machine learning. Information Fusion, 80:146–166.

Vulpe, C. (2015). Design method for the undrained capacity of skirted circular foundations under combined loading: effect of deformable soil plug. Geotechnique, 65(8):669–683.

Villalobos Jara, F. A. (2006). Model testing of foundations for offshore wind turbines, PhD thesis, University of Oxford.

Yin, Z.-Y., Teng, J.-C., Li, Z., and Zheng, Y.-Y. (2020). Modelling of suction bucket foundation in clay: From finite element analyses to macroelements. Ocean Engineering, 210:107577.

Zhang, P., Yin, Z.-Y., Zheng, Y., and Gao, F.-P. (2020). A lstm surrogate modelling approach for caisson foundations. Ocean Engineering, 204:107263.

Zhang, Y., Cassidy, M. J., and Bienen, B. (2014). A plasticity model for spudcan foundations in soft clay. Canadian Geotechnical Journal, 51(6):629–646.