

Research Paper

Transfer learning on lateral capacity prediction of monopile foundations with limited data

Mohammed Alduais^a, Qipei (Gavin) Mei^a, Amir Hosein Taherkhani^b, Fei Han^{b,*}

^a University of Alberta, Department of Civil and Environmental Engineering, Edmonton, AB T6G 2H5, USA

^b University of New Hampshire, Department of Civil and Environmental Engineering, Durham, NH 03824, USA

A B S T R A C T

This study investigates the application of transfer learning to improve predictions of lateral capacity for monopile foundations in offshore wind farms. As the demand for renewable energy grows, cost-effective and accurate monopile designs are essential to support offshore wind turbines. Although deep learning models excel in capturing complex interactions among soil conditions, pile parameters, and loading scenarios, their performance often suffers due to limited training data. Building on a hybrid deep learning architecture combining convolutional and fully connected layers, this research explores the use of transfer learning to address data scarcity by adapting pre-trained models to smaller, task-specific datasets. The study highlights the pivotal role of selecting optimal layer-freezing configurations in transfer learning, demonstrating their significant impact on model performance. Furthermore, an analysis of dataset size reveals that transfer learning consistently outperforms models trained from scratch, particularly when data availability is limited. These findings underscore the potential of transfer learning to enhance predictive accuracy and efficiency in geotechnical engineering, contributing to the advancement of offshore wind energy infrastructure.

1. Introduction

Given the pressing concerns surrounding global climate change, energy supply, and energy costs, there is an increasing emphasis from governments and societies on transitioning towards renewable energy sources (Díaz & Guedes Soares, 2020; Henderson et al., 2003). This shift is motivated by the imperative to reduce CO₂ emissions and address the adverse effects of global warming (Lozano-Minguez et al., 2011). Wind energy emerges as a viable solution, with onshore and offshore locations serving as its primary sources. In particular, offshore wind farms offer significant advantages, such as minimal visual and noise impact due to their distance from the shore and higher wind speed ratios, resulting in increased energy production (Díaz & Guedes Soares, 2020; Oh et al., 2018). As the cost of foundation accounts to up to 36 % of the total cost depending on the water depth, the foundation type and optimal design plays a significant role in reducing the total cost (Oh et al., 2018).

The foundation system for offshore wind turbines includes gravity, monopile, tripod, braced frame, tension leg with suction buckets, and buoy with suction anchor, while monopiles are by far the most widely adopted foundation option in shallow water (<= 40 m) (Arshad & O'Kelly, 2016). The process of installing monopiles entails using hydraulic hammering to drive steel tubes into the seabed. These steel tubes usually have a weight ranging from 500 to 800 tons, a diameter of 4 to

10 m, and a length of up to 50 to 60 m (Díaz & Guedes Soares, 2020). In 2020, the monopile foundation type was adopted by eighty percent of the installed offshore wind farms in Europe (Vázquez et al., 2022). This preference can be attributed to the favorable soil conditions in the North Sea, which requires relatively less effort in pile installation (Díaz & Guedes Soares, 2020).

For monopiles, lateral loading typically governs the design process (Arshad & O'Kelly, 2016). One of the popular methods used in monopile lateral design is the p-y method (DNV, 2014). The method employs a one-dimensional beam model for the pile and simplifies the surrounding soil as nonlinear springs, providing reaction to lateral loading of the pile (Reese et al., 1975). Nevertheless, the predictive accuracy of this method is limited due to its reliance on p-y curves derived from specific field pile load tests (Choi et al., 2014; Han et al., 2015, 2017), and its original development for long, slender piles raises doubts about its suitability for large diameter monopiles (Doherty & Gavin, 2012; Suryasentana & Lehane, 2016). Due to the perceived challenges in utilizing the p-y method for monopile design, endeavors have been undertaken to create new design methods that are more suitable for monopiles. As an example, Hu et al. (2021) performed a series of 3D finite element analyses, leading to the development of design equations and a web application that facilitates the assessment of load-rotation response for monopiles having diameters up to 10 m in uniform sand profiles. Based

* Corresponding author.

E-mail addresses: alduais@ualberta.ca (M. Alduais), Qipei.Mei@ualberta.ca (Q.(G. Mei)), AmirHosein.Taherkhani@unh.edu (A.H. Taherkhani), Fei.Han@unh.edu (F. Han).

<https://doi.org/10.1016/j.compgeo.2025.107359>

Received 20 January 2025; Received in revised form 6 May 2025; Accepted 16 May 2025

Available online 23 May 2025

0266-352X/© 2025 Published by Elsevier Ltd.

on the data derived from their analyses, the authors formulated design equations to accurately estimate the load-rotation response of monopiles in such profiles. However, 3D finite element analyses rely on the proficiency of the engineer involved in the modeling process and the availability of computational resources, which may affect the accuracy of the results (Taherkhani et al., 2023a; Xu et al., 2013).

Deep Learning (DL) has emerged as a transformative approach to solving nonlinear geotechnical problems. Unlike traditional methods, DL algorithms excel in autonomously learning complex relationships between input and output data without the need for explicit formation (Goh et al., 2018; Pham et al., 2020; Zhang et al., 2021; Benbouras et al., 2021). The remarkable popularity of DL can be credited to several advancements: the increasing availability of data, improved network architectures, the development of accessible open-source code libraries, and the computational power of modern Graphics Processing Units (GPUs). These advancements have enabled DL to tackle intricate geotechnical challenges with unprecedented efficiency and accuracy.

Deep Learning (DL) models often require large datasets (e.g., from hundreds to millions of data points) to train, which is a significant challenge in geotechnical engineering where such datasets are rarely available. Collecting high-quality geotechnical data involves costly and time-intensive processes such as in-situ testing or laboratory experiments, making it difficult to assemble large datasets. To address this limitation, we propose a novel transfer learning framework. Our approach begins with training a foundation DL model on a large synthetic dataset generated through 3D finite-element analyses, which encompass a broad range of design parameters such as pile geometry, cone penetration test (CPT) data, and loading conditions. This foundation model is then fine-tuned using transfer learning, enabling it to adapt to project- or site-specific conditions with as few as five new data points, thereby bridging the data scarcity gap in geotechnical applications.

The paper is structured as follows: Section 2 introduces the methodology, encompassing dataset generation, model architecture, transfer learning strategies, and the training process. Section 3 presents the results, accompanied by a discussion on the influence of various transfer learning strategies and the effects of using small dataset sizes on model performance. Section 4 describes the training and testing of a general model on a case study utilizing real-world data from the PISA project. Finally, Section 5 concludes the paper by summarizing the key findings.

2. Methodology

2.1. Datasets

The datasets used in this paper were generated based on the results presented by Hu et al. (2021, 2022), where they performed 3D Finite element (FE) analyses to model monopiles laterally loaded in multilayered sandy soil using Abaqus/Explicit (ABAQUS, 2014). The analyses were validated against pile load test results obtained from the PISA project (Byrne et al., 2019). Leveraging their FE analysis results, we created a comprehensive dataset named DS1, comprising 100,000 data points. This dataset serves as the foundation for the foundation model. Each data point in DS1 includes input parameters such as moment of inertia I_p , pile length L , cone penetration test (CPT) cone resistance q_c profile as a function of depth, and the load eccentricity h . The outputs for each data point are the lateral pile capacities $H_{0.5}$ and H_1 , corresponding to mudline pile rotation levels of $\theta = 0.5^\circ$ and $\theta = 1^\circ$, respectively. To train a general foundation model, DS1 spans a wide range of design parameters: pile length (6–150 m), pile diameter (2–10 m), and load eccentricity (15–30 m). To explore the generalization and transfer learning capabilities of our model, we included two additional datasets, DS2 and DS3, with extended parameter ranges:

- **DS2:** Simulates over-consolidated soil by expanding the over-consolidation ratio (OCR) range from 1 to the 5–20 range. The ranges for pile length, diameter, and eccentricity were narrowed

Table 1

Characteristics of generated datasets for monopile response prediction with varied parameters and loading conditions.

Data Set	Pile Diameter (m)	Pile length (m)	Load Eccentricity (m)	OCR	Datapoints
DS1	2—10	6—150	15—30	1	100,000
DS2	2—4	6—60	30	5—20	4,300
DS3	2—10	6—150	1—45	1—50	5500

(Table 1) to reflect realistic project scenarios specific certain pile geometries are considered. This dataset is designed to evaluate the foundation model's adaptability to new site conditions characterized by highly over-consolidated soil.

- **DS3:** extends the loading eccentricity range to 1–45 m and the OCR range to 1–50. This dataset is designed to enhance the foundation model's ability to handle a wider variety of loading and soil conditions, leveraging transfer learning for improved adaptability and performance.

The quality and quantity of training data directly influences model performance, making data management a critical component of any machine learning pipeline. Typically, the data is divided into three subsets: training, validation, and test sets. The training set, which constitutes the majority of the data, is used to optimize the model's parameters. A smaller subset, the validation set, is employed to monitor and evaluate the model's performance during training, helping to fine-tune hyperparameters and prevent overfitting. Finally, the test set, consisting of entirely unseen data, is reserved for assessing the model's overall performance and generalizability.

To explore the impact of dataset size on model performance, we employed two distinct data-splitting strategies based on the dataset size. For large datasets containing more than 500 data points, we allocated 90 % of the data for training, 5 % for validation, and 5 % for testing. This approach ensured that the model had ample data for training while maintaining a balanced distribution for validation and testing. Conversely, for smaller datasets with fewer than 500 data points, we utilized a fixed dataset approach. In this scenario, the entire small dataset was dedicated to training, while an additional 500 data points were separately reserved as the validation set. This strategy was designed to expand the validation set, addressing concerns about the model's ability to generalize effectively when training data is limited.

2.2. Model architecture

Deep Learning (DL) is a subset of machine learning that uses neural network architectures with multiple layers, also known as Deep Neural Networks (DNNs), to extract patterns and process information from input data, ultimately producing meaningful outputs. DL encompasses various architectures, including Fully-Connected Neural Networks (FCNNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs), each suited for specific types of data and tasks. These architectures are inspired by biological neural processes and aim to replicate how the brain processes information (Agatonovic-Kustrin & Beresford, 2000). Fully Connected Neural Networks (FCNNs), one of the simplest forms of DL, generally consist of an input layer, multiple hidden layers, and an output layer. The process begins with an input vector, which is transformed using a linear function and then passed through a non-linear activation function. This process is repeated across neurons in the hidden layers, enabling the network to learn complex patterns. Among the popular activation functions, the Rectified Linear Unit (ReLU) is particularly effective for generalization and adaptation to varying data (Nair & Hinton, 2010; Feng & Lu, 2019).

Convolutional Neural Networks (CNNs), a specialized type of DNN, are designed to process structured data, such as images, by reducing complexity and parameter counts through techniques like receptive

fields, weight sharing, and pooling operations (Goodfellow et al., 2016; Ren et al., 2017). While traditional “2D CNNs” are tailored for 2D inputs like images, a variant known as “1D CNNs” is optimized for 1D data. These 1D CNNs have demonstrated exceptional effectiveness in processing one-dimensional vector data, such as the Cone Penetration Test (CPT) data utilized in this study (Kiranyaz et al., 2021). A typical CNN architecture includes three primary layers: input, convolutional, and pooling layers (Fig. 1). In convolutional layers, input data is processed using kernels, followed by non-linear activation (e.g., ReLU) and a bias term to generate feature maps. Pooling layers, such as average pooling and max pooling, further reduce dimensionality by summarizing feature maps. For this study, MaxPooling was employed to downsample features.

Fig. 2 outlines the hybrid neural network architecture used in this research (Taherkhani et al., 2023a; Taherkhani et al., 2023b). The model utilizes convolutional layers specifically to extract depth-wise patterns from Cone Penetration Test (CPT) data, while fully connected layers are employed to capture the impact of pile parameters and loading conditions on the lateral capacity. The architecture consists of four core components: CPT_Conv, CPT_FC, PILE_FC, and Predictor_FC. The CPT_Conv component, comprising nine blocks, uses 1D convolutional and MaxPooling layers to capture patterns in CPT data. CPT_FC, consisting of four fully connected layers, further extracts and refines these trends for enhanced understanding. The PILE_FC component, also with four fully connected layers, models the effects of pile parameters and loading conditions on monopile behavior. Lastly, the Predictor_FC

component, with two fully connected layers, combines outputs from previous components to capture soil-pile interactions. Implemented using PyTorch version 1.11.0 and Python 3.9.10, the model was trained on a computer equipped with an AMD Ryzen 5 5600X processor, 16 GB RAM, and an NVIDIA GeForce RTX 3070 TI GPU. This computational setup ensured efficient training and robust performance.

2.3. Transfer learning

Transfer learning is a machine learning technique that facilitates the reuse of knowledge acquired from a large dataset to tackle a distinct but related task involving a smaller dataset (Mitelman & Uralinis, 2023). This process begins by training a general foundation model on a large dataset to optimize its parameters (weights and biases), which are subsequently fine-tuned with a smaller, task-specific dataset. This approach eliminates the need to train the target model from scratch see Fig. 3. Transfer learning is typically done in two steps:

Foundation model training: A general foundation model is trained on a large and diverse dataset (source domain) to capture generic patterns and features. The resulting parameters (e.g., weights and biases) are preserved for subsequent use in a target model.

Fine-tuning the target model: The preserved parameters from the foundation model are transferred to the target model and fine-tuned using a smaller, task-specific dataset. This step involves selectively

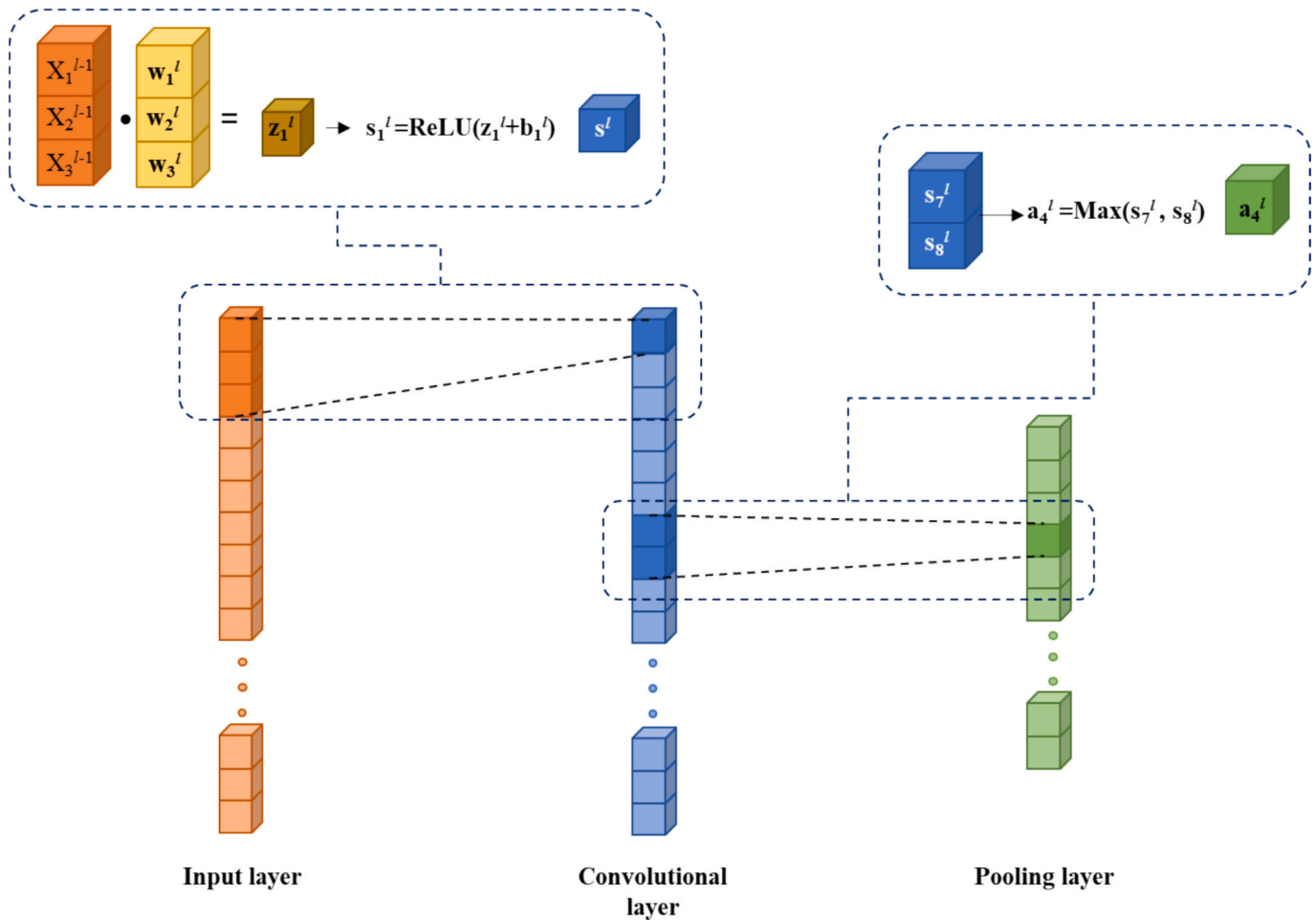


Fig. 1. Schematic representation of a one-dimensional Convolutional Neural Network (CNN). In layer $l-1$, the vector X^{l-1} is involved in a dot product computation with the kernel vector w with a size 3, resulting in z . This value is then passed through a nonlinear activation function, combined with a bias term b , to produce the activation output s . Subsequent max pooling is applied to the activation outputs, where a moving window selects the maximum value within each window.

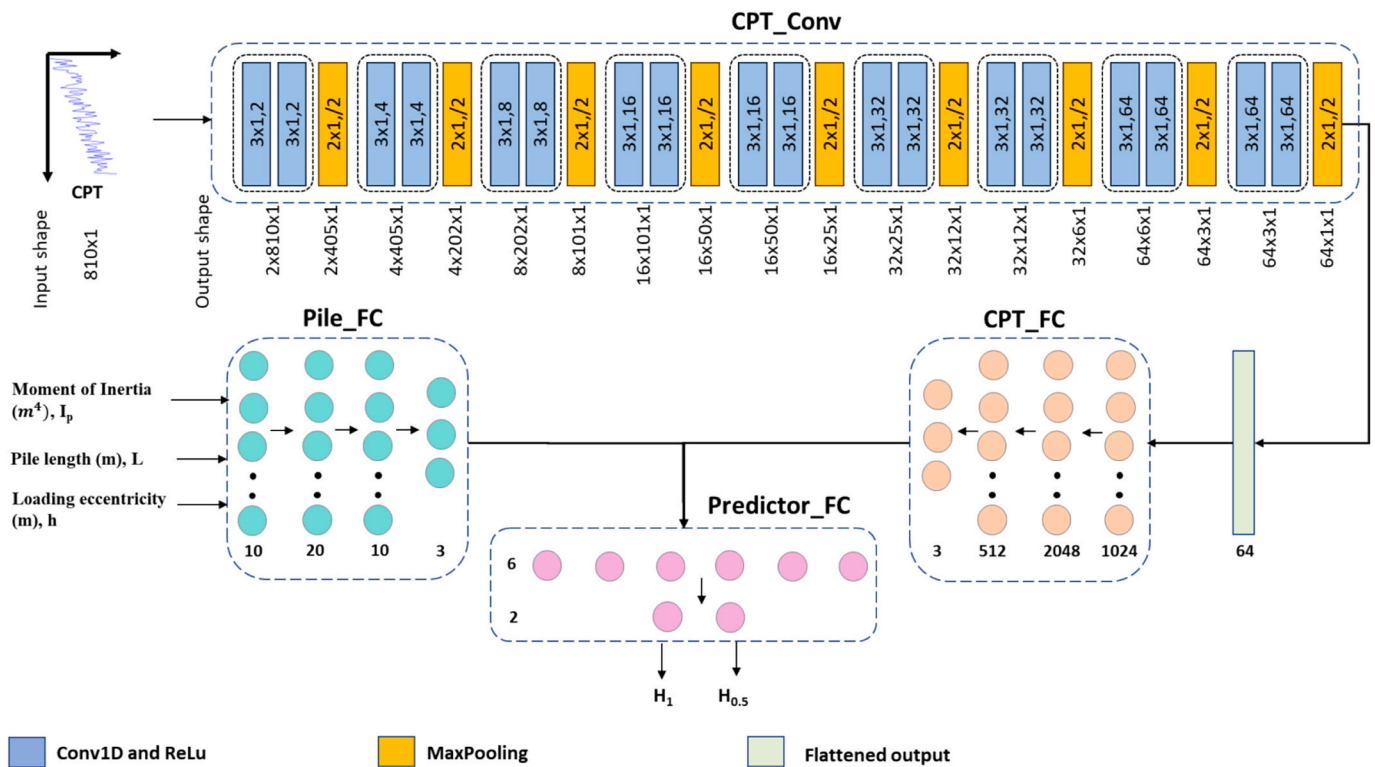


Fig. 2. Our proposed model architecture. The CPT_Conv component consists of nine blocks utilizing 1-D convolutional layers and MaxPooling layers to capture CPT data trends. CPT_FC comprises four fully connected layers, processing the flattened CPT_Conv output. Simultaneously, Pile_FC integrates pile parameters and loading conditions through four fully connected layers. The outputs of CPT_FC and Pile_FC serve as inputs to Predictor_FC, composed of two fully connected layers, yielding lateral capacities $H_{0.5}$ and H_1 for the monopile. The numbers inside the blue boxes represent the filter and the output channels, while in the yellow boxes represent the filter size and the stride. The model uses a kernel size of 3 with a padding of 1 for all convolutional layers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

updating a subset of the model's parameters based on the new dataset.

The application of transfer learning in geotechnical engineering has shown promising results in various studies. For instance, Fu et al., (2022) utilized transfer learning to predict cutterhead torque in tunnel boring machines using historical datasets. Man et al., (2022) employed transfer learning on a Residual Network (ResNet) to identify water leakage and cracks, demonstrating improved model accuracy. Wang et al., (2022) proposed a transfer learning-based model for landslide susceptibility assessment, achieving a 30 % increase in performance on the target domain test set. Similarly, Chen et al., (2023) developed a ResNet model with transfer learning to classify seven types of rock, showing a significant accuracy improvement from 88.1 % without transfer learning to 99.1 % with transfer learning.

To better understand the datasets used in this study, it is important to clarify the roles of the source and target domains. DS1 serves as the source domain and is used to pre-train a general foundation model. This model is subsequently fine-tuned using either DS2 or DS3 to adapt to new, site-specific conditions serving as the target domains.

DS2 represents over-consolidated soil conditions and is employed to evaluate the model's adaptability under a significantly shifted data distribution. In contrast, DS3 is an extended version of DS1 that incorporates a broader range of CPT profiles and loading scenarios. This enables the training of a more generalizable foundation model, which demonstrates strong transferability and high predictive accuracy when adapted to new conditions using limited data.

From a data distribution perspective, the most notable variation across datasets lies in the CPT profiles. Other input features—such as pile geometry, loading conditions, and soil layering—are either identical or fall within overlapping parameter ranges. Fig. 4.a illustrates the

differences in CPT profiles between DS1 and DS2, while Fig. 4.b highlights the broader yet partially overlapping feature space of DS3 relative to DS1. These shared input distributions help establish domain similarity, which provides a strong theoretical basis for effective knowledge transfer. This similarity minimizes the risk of negative transfer and supports the application of transfer learning (TL) within this context.

To enhance the performance of the new model, we experimented with different freezing and unfreezing strategies. Freezing refers to preventing the weights of certain layers from updating during fine-tuning, while unfreezing allows layer weights to be updated, typically using the parameters of the foundation model as initialization. In this study, we discovered a specific freezing and unfreezing layout, as detailed later in the paper under the section titled 'Selection of Transfer Learning Strategies.' Our findings emphasize the effectiveness of transfer learning with pre-trained foundation models in mitigating data scarcity challenges and demonstrate the critical role of carefully selecting freezing and unfreezing strategies to optimize the performance of deep learning models.

3. Results and discussion

To evaluate the effectiveness of transfer learning, a foundation model was first trained on 90,000 data points from DS1. The training process was conducted over 600 epochs using a batch size of 128, a learning rate of 0.001, and Adam as the optimizer. This extensive dataset enabled the foundation model to achieve high accuracy, with Mean Absolute Percentage Error (MAPE) values of 3.41 % and 3.75 % for predicting $H_{0.5}$ and H_1 , respectively. The proposed hybrid deep learning architecture was designed to capture the complex, nonlinear relationships in the dataset—particularly those arising from the interaction between CPT measurements and pile geometry. To assess whether such a complex

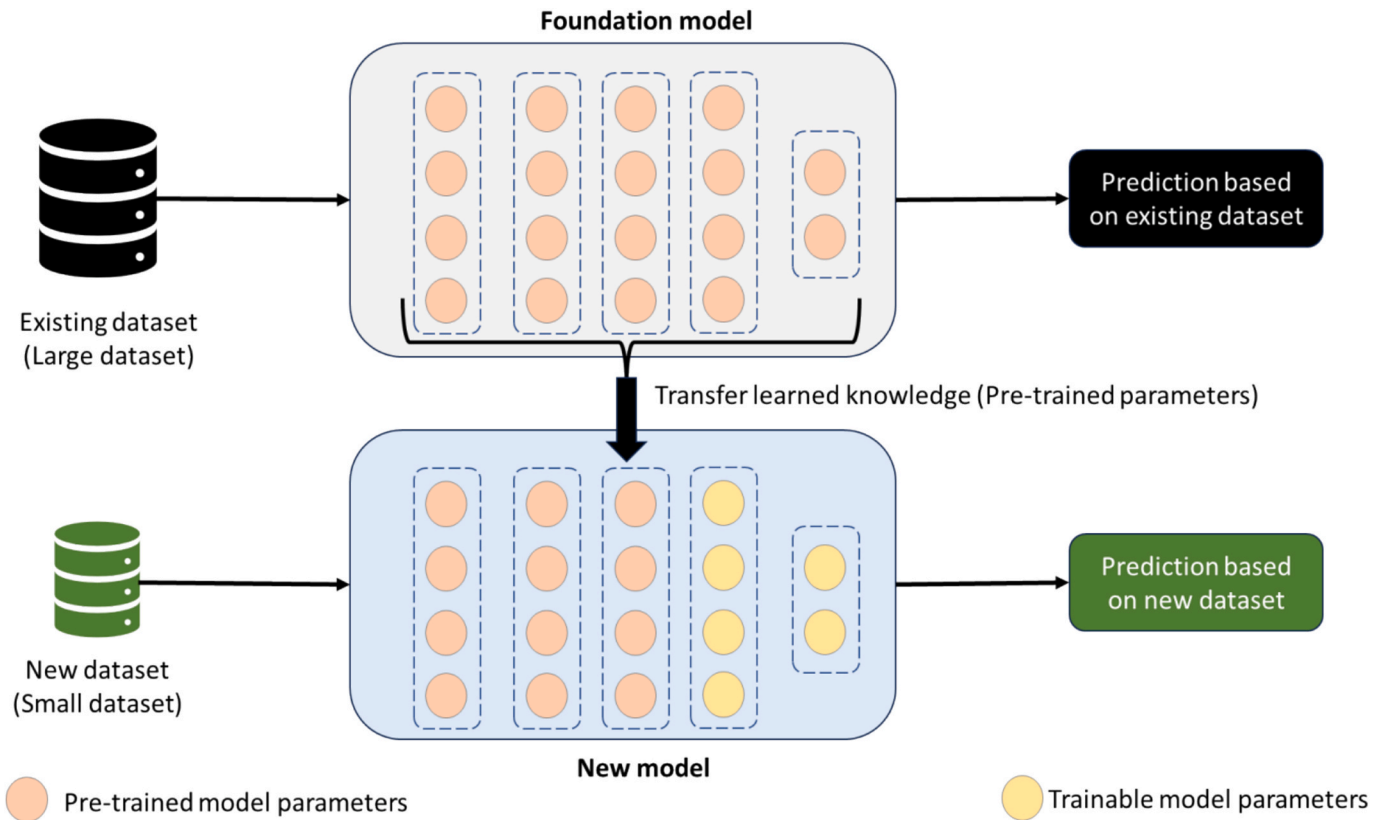


Fig. 3. Overview of the transfer learning framework. The training starts with training a foundation model using a large amount of data. Once the training finishes, a new model can be trained using the new data while preserving the pre-trained parameters from the foundation model for some of the shallow layers.

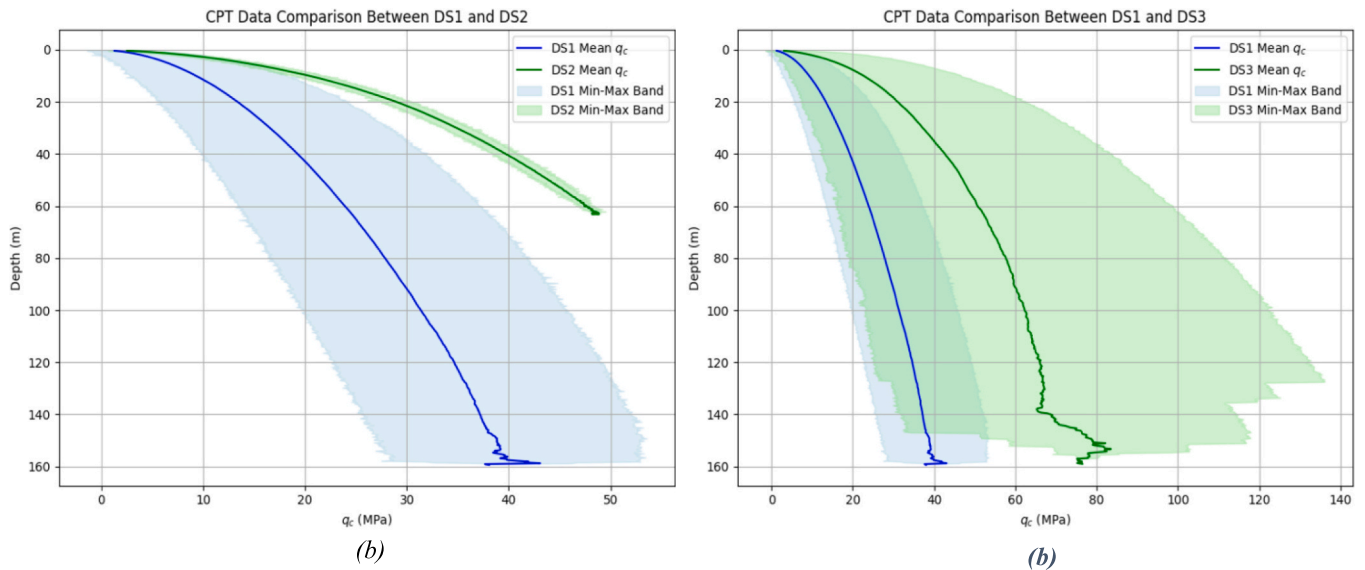


Fig. 4. Comparison of CPT profiles between the source and target domains: (a) DS1 vs. DS2, (b) DS1 vs. DS3.

architecture is indeed necessary, we conducted a comparative experiment by evaluating three alternative ANN models with varying model complexity. Specifically, ANN1 comprises 3 hidden FC layers for processing CPT data, 3 FC layers for pile geometry, and 2 FC layers for final prediction. ANN2 and ANN3 are similar to ANN1 but with progressively reduced number of layers. As shown in Table 2, models with simpler architectures exhibit notably poorer predictive power when compared with the proposed model, supporting the need for deeper and more

Table 2

A comparison between different models in terms of complexity and prediction accuracy.

Model	Number of layers	MAPE for $H_{0.5}$ (%)	MAPE for H_1 (%)
Hybrid model	18 (CNN) + 10 (FC)	3.41	3.75
ANN1	8FC	14.08	16.27
ANN2	6FC	22.16	24.97
ANN3	4FC	48.05	46.78

complex architecture in this application.

Since synthetic data were used for training and evaluating the model, concerns arise about their ability to represent the complexity and variability encountered in real-world conditions. Specifically, synthetic data are unable to fully represent critical aspects such as soil heterogeneity, inherent variability and measurement noise in CPT data, and variability in pile capacities resulting from installation processes. To assess the model's robustness against these limitations and input uncertainties, Gaussian noise was introduced to the CPT data in the test dataset in DS1. The noise was applied at ten different levels, corresponding to standard deviation values ranging from 1 MPa to 10 MPa, to simulate uncertainty and reflect real-world soil variability. The pre-trained foundation model was then tested using these datasets with added noise. As shown in Fig. 5, the model is robust in predicting pile capacities $H_{0.5}$ and H_1 for a wide range of noise levels. While the introduction of Gaussian noise does not entirely replicate the complexity of real-world conditions, it effectively highlights the model's practical applicability and generalizability, particularly in scenarios where comprehensive real datasets are unavailable.

In the transfer learning stage, the pre-trained foundation model was fine-tuned using the new dataset DS2. The Fine-tuning involved 700 epochs with a smaller batch size of 5, a reduced learning rate of 0.0001, and the Adam optimizer. The new model's performance was evaluated using different subsets of DS2 as training data, containing 5, 25, 100, and 250 data points. The errors for these models are compared in Table 3.

These results highlight the effectiveness of transfer learning in leveraging knowledge from the larger dataset DS1 to improve performance on the smaller dataset DS2. Despite using only a fraction of the new dataset, the fine-tuned model demonstrated strong generalization capabilities and achieved relatively low MAPE values. With as low as 5 new data points, the model can yield predictions with less than 10 % relative error. This underscores the potential of transfer learning to perform accurate predictions in new data domains with limited data availability.

Table 3

The Mean Absolute Percentage Error (MAPE) of the new model trained using transfer learning with data sets of different sample sizes.

Data size for transfer learning	MAPE for $H_{0.5}$ (%)	MAPE for H_1 (%)
5	8.91	8.92
25	5.12	5.51
100	3.64	3.91
250	2.88	3.09

3.1. Transfer learning strategies

The layer-freezing configuration in deep learning plays a critical role in the effectiveness of transfer learning and the performance of the fine-tuned model. A common approach to determine the optimal configuration of frozen and fine-tuning layers involves gradually unfreezing layers starting from the output side of the pre-trained foundation model and evaluating performance for each configuration (Howard & Ruder, 2018). To identify the optimal freezing layout for this specific task, we conducted experiments on eight different freezing configurations using 5 randomly selected subsets of DS2, each containing 25 data points for training and 500 data points for validation. The training process involved 600 epochs, a learning rate of 0.001, and a batch size of 5.

Fig. 6 illustrates the eight layer-freezing configurations explored in this study. In Layout 1, only the last layer in Predictor_FC (lower box) was unfrozen, while all other model layers remained frozen. In Layout 2, both layers in Predictor_FC were unfrozen, with the rest of the model frozen. These two configurations yielded poor performance compared to the remaining layouts, likely due to the limited number of trainable parameters, which restricted the model's ability to effectively learn from the new data. In Layouts 3 through 6, both layers in Predictor_FC remained unfrozen, and additional layers in Pile_FC (lower-left box) were incrementally unfrozen with each layout. Model performance improved progressively as additional layers in Pile_FC, which capture the impact of pile geometry, were unfrozen. Notably, performance reached a peak when all layers in Predictor_FC and Pile_FC were unfrozen, along with one layer in CPT_FC (lower-right box), as seen in Layout 7. However, in Layout 8, as we unfreeze more layers in CPT_FC, the model's performance began to degrade. This indicates that

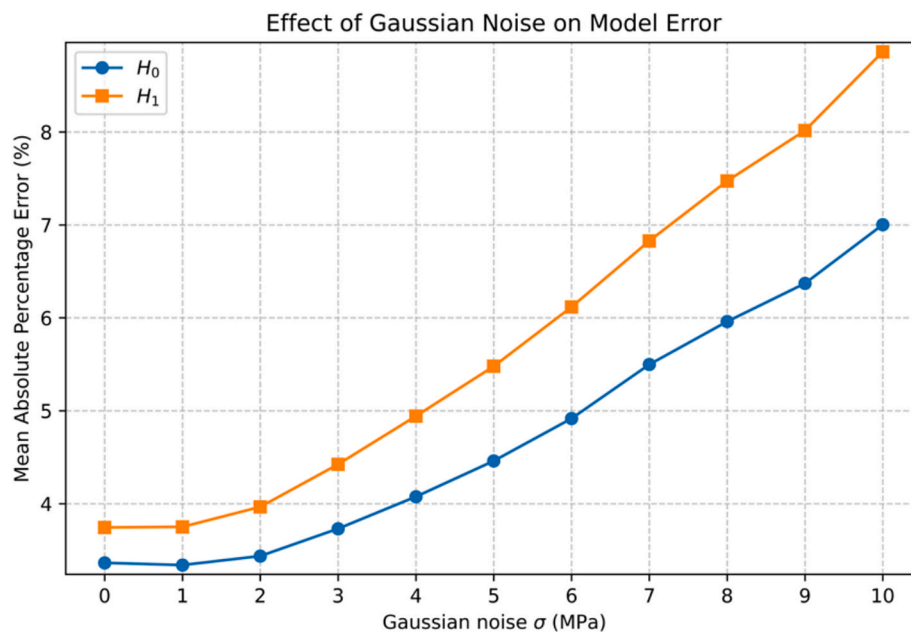


Fig. 5. Foundation model performance under varying levels of Gaussian noise. "0" represents predictions using the original test dataset without any noise, while the remaining values indicate predictions made with normally-distributed noise added to the CPT cone resistance profile at standard deviation levels ranging from 1 MPa to 10 MPa.

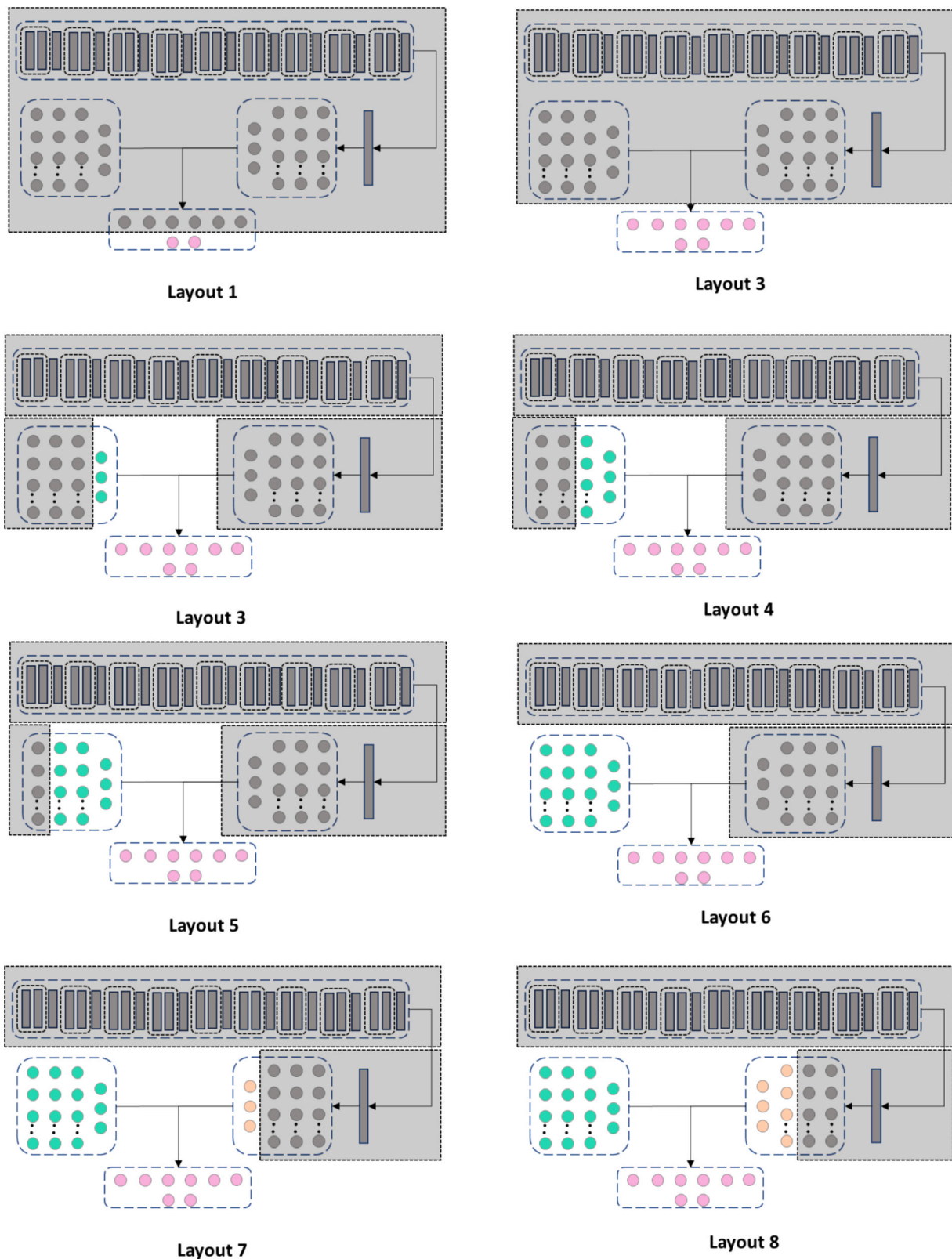


Fig. 6. Overview of freezing layouts 1 to 8 used in this paper. The gray-shaded areas represent layers that are frozen, meaning that the pre-trained parameters from the foundation model are retained and not updated during training. In contrast, the colored layers are unfrozen and allowed to update their parameters during the fine-tuning process.

unfreezing too many layers may lead to overfitting or reduce the model's ability to generalize effectively to new data.

A likely explanation for this behavior lies in the distributional differences between the DS1 (source domain) and the DS2 (target domain). Compared to DS1, DS2 introduces a substantially higher OCR range (from 1 to 5–20), while also narrowing pile geometry parameters (diameter reduced from 2–10 m to 2–4 m; length from 6–150 m to 6–60 m; eccentricity fixed at 30 m). These changes result in a significant distribution shift. When only the final layers are unfrozen, the model struggles to adapt to the new data distribution. As more layers in the Pile_FC branch are opened, performance improves due to better adaptation to the specific pile geometries. Layout 7 further unfreezes layers in the CPT_FC branch, enabling the model to adjust earlier feature extraction layers in response to altered CPT patterns resulting from new OCR values. This configuration achieves an optimal balance between retaining pretrained knowledge and adapting to the new domain. In contrast, Layout 8 opens too many layers, increasing the number of trainable parameters and leading to overfitting given the limited size of DS2—thus explaining the observed performance drop.

It is important to note that the optimal freezing layout may vary depending on dataset size. For smaller datasets, such as the 25 data points used in this study, limiting the number of unfrozen layers helps prevent overfitting while allowing sufficient adaptability. In contrast, with larger datasets, unfreezing additional layers, such as more layers in CPT_FC, could enhance performance by leveraging the increased amount of training data and reducing the risk of overfitting. Future work could explore the relationship between dataset size and optimal layer-freezing strategies to further generalize these findings.

Table 4 provides a summary of the average performance metrics for new models trained across the five datasets. These results highlight the critical role of a well-designed layer-freezing strategy in transfer learning.

3.2. The impact of dataset size on the transfer learning effectiveness

To conduct a comprehensive analysis, we systematically generated four distinct sample sizes (5, 25, 100, and 250) from the DS2 dataset. For each sample size, we created ten different training sets, each randomly sampled from the DS2 dataset. The purpose is to mitigate the impact of random sampling variability, particularly for smaller datasets, which could otherwise introduce noise and lead to misleading conclusions. Each training set was independently used to train the model. Validation errors were calculated and summarized in terms of the minimum, maximum, and mean values of the Mean Absolute Percentage Error (MAPE) obtained from the 10 training sets for each sample size (Fig. 7). The training process involved 700 epochs, a batch size of 5, a learning rate of 0.0001, and the Adam optimizer.

To establish a baseline for comparison, we also trained a deep learning model using the same architecture entirely from scratch, without using transfer learning. The performance of these freshly-trained models without transfer learning is shown in Fig. 7 for comparison. To ensure consistent and fair comparison, a fixed testing dataset comprising 500 data points was used to evaluate all models.

Table 4
Performance Evaluation of Different Freezing Layouts.

Layout	Mean absolute percentage error (MAPE)	
	H _{0.5} (%)	H ₁ (%)
1	16.44	14.63
2	15.01	13.66
3	13.93	12.74
4	12.03	12.30
5	11.35	11.28
6	11.15	10.94
7	11.14	10.97
8	11.29	11.18

When trained from scratch, model performance was heavily influenced by dataset size. Models trained on smaller datasets exhibited high MAPE values due to insufficient training data and random weight initialization. For example, in predicting H_{0.5}, the mean MAPE decreased significantly from 99.96 % with 5 data points to 16.62 % with 250 data points. This demonstrates the model's inability to generalize effectively when trained with minimal data.

In contrast, transfer learning significantly reduced dependency on dataset size, resulting in consistent and reliable performance across all sample sizes. By leveraging pre-trained weights, models achieved improved generalization and reduced MAPE values. For instance, in predicting H_{0.5}, the mean MAPE ranged from 5.17 % (5 data points) to 3.03 % (250 data points). Additionally, the variability in performance, as measured by the interquartile range, was significantly reduced. With 100 data points, the interquartile range for training from scratch was [24.22 % – 34.37 %], while transfer learning narrowed this range to [3.26 % – 4.25 %]. This finding highlights the improved stability and reliability provided by the transfer learning approach.

To further validate these results, a paired two-sample *t*-test was conducted to compare models trained using transfer learning against those trained from scratch across four different dataset sizes. For the smallest dataset, the test on the H_{0.5} output yielded a *t*-statistic of 11.81 and a *p*-value of 5.67×10^{-16} , confirming a statistically significant improvement. Similar results were observed for dataset sizes of 25, 100, and 250, where transfer learning models consistently outperformed their non-transfer learning counterparts. The corresponding *t*-statistics (48.99, 23.22, and 155.11) and *p*-values (1.21×10^{-9} , 1.21×10^{-9} , and 4.88×10^{-17}) further support the robustness of these improvements. Overall, transfer learning provides a substantial and statistically significant performance boost, particularly for smaller datasets.

These results demonstrate the substantial benefits of transfer learning in data-limited scenarios. By leveraging pre-trained models, transfer learning not only improves predictive accuracy but also reduces the dependence on large datasets. This makes it a valuable approach for geotechnical and other domains where data availability is constrained.

4. General model and case study

As transfer learning has proved to be an effective scheme, we leverage the technique to obtain an even more general foundation model, further extending the model applicability. The layer-freezing configuration employed in this case study is Layout 7. Comparing to our original foundation model trained using DS1, the new foundation model was trained using a new dataset DS3 that extends the load eccentricity range from 15–30 m to 1–45 m, and the OCR range from 1 to 1–50 (Table 1). The DS3 dataset contains 5500 datapoints, where 5000 were for training and 500 for validation. We trained the model for 600 epochs, using a batch size of 2, a learning rate of 0.001, and Adam as the optimizer.

To assess the performance of the new foundation model, we utilized full-scale pile load test data obtained from PISA project, in which a 2-m-diameter and 10.5-m-long pile was laterally loaded with an eccentricity of 10 m in multi-layered sandy soil (Byrne et al., 2019). The CPT data was obtained from (Byrne et al., 2019). Fig. 8 compares the load-pile rotation curves obtained from the load test, 3D Finite Element (FE) analysis performed by Hu et al. (2022a, 2022b), and the new foundation DL model trained in this study. Note that the DL model predicts pile capacities H_{0.5} and H₁ corresponding to 0.5° and 1° mudline pile rotation. The load-rotation curve is then obtained using the equation proposed by Hu et al., (2022b) as follows:

$$H = \frac{\theta}{k + \eta\theta} \quad \text{Where} \quad \begin{cases} \eta = \frac{2}{H_{1^\circ}} - \frac{1}{H_{0.5^\circ}} \\ k = \frac{1}{H_{0.5^\circ}} - \frac{1}{H_{1^\circ}} \end{cases}$$

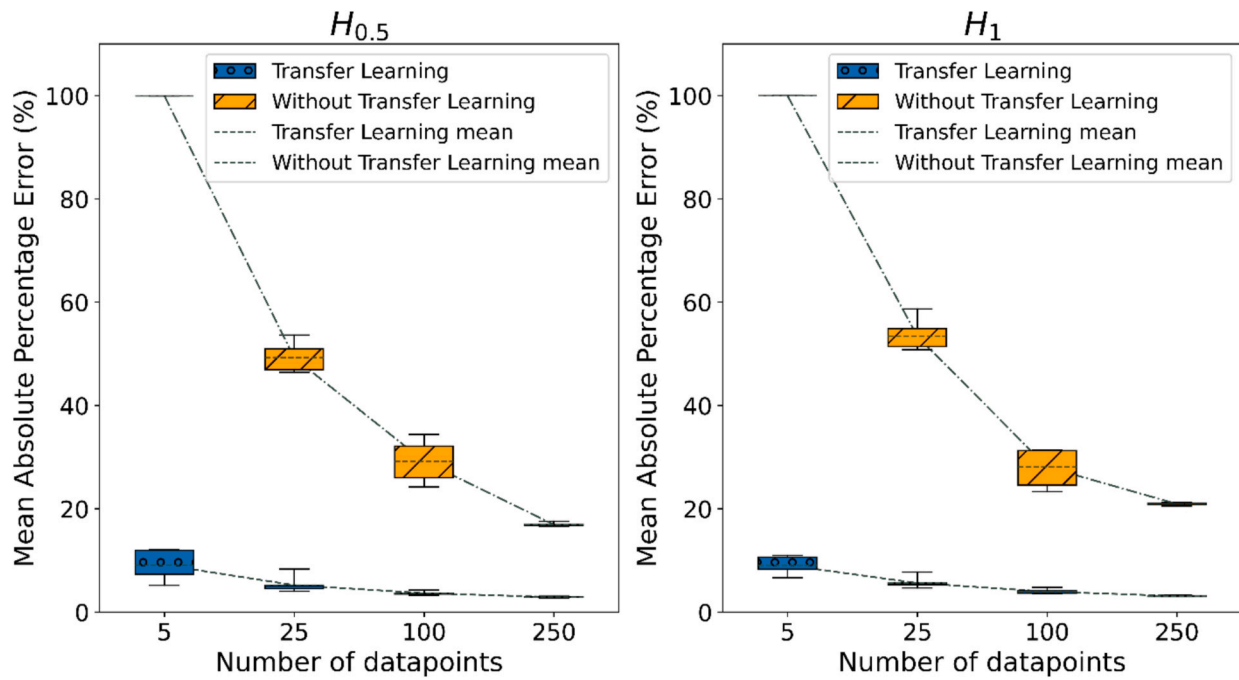


Fig. 7. Comparing training without Transfer learning and training with for $H_{0.5}$ and H_1 under varying training dataset sizes. The upper plots represent the MAPE values of training without transfer learning. The lower plots represent the MAPE values of training with transfer learning. The length of the box and the whiskers indicates the variability of the data. The longer whiskers or boxes, the more variability.

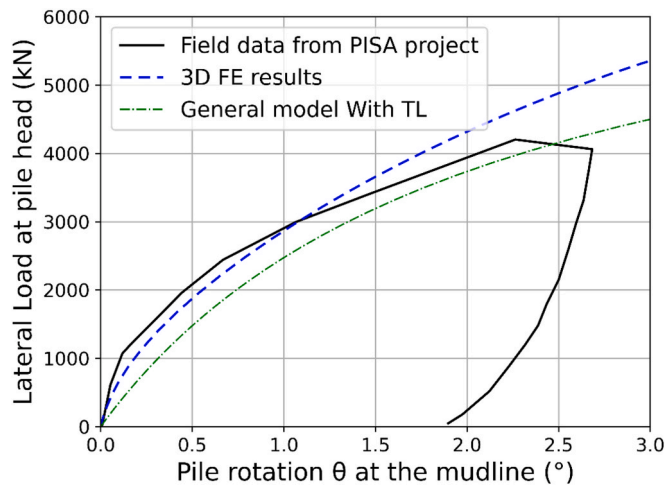


Fig. 8. Validation of transfer learning general model against PISA Project Data and 3D finite element results. The solid black line represents the field data from PISA project, the dashed blue line represents the 3D finite element results, and the dashed green line (lower line) represents the prediction made by the proposed general model with transfer learning. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The motivation behind using these empirical equations is to keep the model simple and efficient without losing important predictive accuracy. Predicting the full load–displacement curve directly would require a more complex model with higher output dimensions, which can be difficult to train, especially with limited data. Instead, we used well-known empirical equations that estimate the full curve based on just two key points, $H_{0.5}$ and H_1 . These equations have been widely tested and used in previous studies (Kulhaway & Chen, 1995; Hu et al., 2022a, 2022b; OuYang et al., 2025), making them a reliable and practical choice. This approach reduces computational effort while still capturing the main features of the monopile response. Notably, our DL foundation

model produced predictions that closely matched both the field data from pile load tests and the results of 3D finite element (FE) analyses. While the predictions from the transfer learning general model show reasonable agreement with both the field data and the 3D FE results, the 3D FE curve aligns more closely with the field measurements. This is expected, as the FE model was specifically developed for the PISA project site using soil constitutive model tailored to the soil condition at the PISA site. In contrast, our machine learning foundation model was trained on a broader dataset (DS3) that includes a wide range of pile geometries and soil profiles. Based on the findings in this study, we recommend fine-tuning the foundation model using a small number of site-specific data points, to further improve the prediction power.

5. Conclusion

This study presents a novel application of deep learning to predict the lateral capacity of monopiles, which are often used as offshore wind farm foundations. The proposed deep learning model employs a hybrid architecture that integrates convolutional and fully connected layers to extract patterns from Cone Penetration Test (CPT) data and model the relationships between pile parameters, loading conditions, and lateral capacity. The transfer learning technique is used to leverage knowledge gained from a larger dataset, enabling fine-tuning and adaptation of a general foundation model using smaller, task-specific datasets.

Key findings include the effectiveness of transfer learning in reducing the dependency on dataset size, achieving consistent and accurate performance across various data scenarios. The study also highlights the critical role of layer-freezing configurations in optimizing transfer learning, demonstrating that careful selection of frozen and fine-tuned layers significantly impacts model performance. Additionally, the analysis of dataset size confirms that transfer learning consistently outperforms models trained from scratch, even with minimal data availability.

The model's reliability was validated through a case study using real-world data from the PISA project. Predictions from the deep learning foundation model closely matched both field data and results from 3D finite element analyses, demonstrating the model's capability to

generalize to practical engineering problems. Furthermore, the development of a more generalized foundation model, trained on an extended dataset, expanded the applicability of the approach to a broad range of loading and soil conditions.

While the proposed transfer learning framework demonstrates strong predictive performance and data efficiency, it relies primarily on synthetic data generated from 3D finite element analyses. Although this data has been validated against field measurements, it may not fully capture the complexity and variability of real-world soil. Addressing this limitation through the incorporation of more extensive real-world datasets, end-to-end curve prediction models, and experimentation with different deep learning architectures represents valuable directions for future work.

In conclusion, this research demonstrates that transfer learning is a powerful tool for addressing data scarcity and enhancing predictive accuracy in geotechnical engineering. By enabling effective utilization of limited data, the approach has the potential to improve the efficiency and cost-effectiveness of monopile design for offshore wind energy projects, ultimately contributing to the advancement of renewable energy infrastructure.

CRedit authorship contribution statement

Mohammed Alduais: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation. **Qipei (Gavin) Mei:** Funding acquisition, Writing – review & editing, Methodology, Conceptualization. **Amir Hosein Taherkhani:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Fei Han:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Reese, Lymon C., Cox, William R., and Francis D. Koop. "Field Testing and Analysis of Laterally Loaded Piles on Stiff Clay." Paper presented at the Offshore Technology Conference, Houston, Texas, May 1975. <https://doi.org/10.4043/2312-MS>.
- Agatonovic-Kustrin, S., Beresford, R., 2000. Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharm. Biomed. Anal.* 22 (5), 717–727. [https://doi.org/10.1016/S0731-7085\(99\)00272-1](https://doi.org/10.1016/S0731-7085(99)00272-1).
- Arshad, M., O'Kelly, B.C., 2016. Analysis and design of monopile foundations for offshore wind-turbine structures. *Mar. Georesour. Geotechnol.* 34 (6), 503–525. <https://doi.org/10.1080/1064119X.2015.1033070>.
- Benbouras, M.A., Petrişor, A.-I., Zedira, H., Ghelani, L., Lefleif, L., 2021. Forecasting the bearing capacity of the driven piles using advanced machine-learning techniques. *Appl. Sci.* 11 (22), 10908. <https://doi.org/10.3390/app112210908>.
- Byrne, B. W., Burd, H. J., Zdravkovic, L., Abadie, C. N., Houlsby, G. T., Jardine, R. J., Martin, C. M., McAdam, R. A., Pacheco Andrade, M., Pedro, A. M. G., Potts, D. M., & Taborda, D. M. G. (2019). PISA Design Methods for Offshore Wind Turbine Monopiles. *Day 1 Mon*, May 06, 2019, D0115005R002. Doi: 10.4043/29373-MS.
- Chen, W., Su, L., Chen, X., Huang, Z., 2023. Rock image classification using deep residual neural network with transfer learning. *Front. Earth Sci.* 10, 1079447. <https://doi.org/10.3389/feart.2022.1079447>.
- Choi, Y.S., Basu, D., Salgado, R., Prezzi, M., 2014. Response of laterally loaded rectangular and circular piles in soils with properties varying with depth. *J. Geotech. Geoenviron. Eng.* 140 (4), 04013049. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0001067](https://doi.org/10.1061/(ASCE)GT.1943-5606.0001067).
- Díaz, H., Guedes Soares, C., 2020. Review of the current status, technology and future trends of offshore wind farms. *Ocean Eng.* 209, 107381. <https://doi.org/10.1016/j.oceaneng.2020.107381>.

- Doherty, P., & Gavin, K., 2012. Laterally loaded monopile design for offshore wind farms. *Proceedings of the Institution of Civil Engineers - Energy*, 165(1), 7–17. Doi: 10.1680/ener.11.00003.
- Fu, T., Zhang, T., Song, X., 2022. A novel hybrid transfer learning framework for dynamic cutterhead torque prediction of the tunnel boring machine. *Energies* 15 (8), 2907. <https://doi.org/10.3390/en15082907>.
- Goh, A.T.C., Zhang, W., Zhang, Y., Xiao, Y., Xiang, Y., 2018. Determination of earth pressure balance tunnel-related maximum surface settlement: A multivariate adaptive regression splines approach. *Bull. Eng. Geol. Environ.* 77 (2), 489–500. <https://doi.org/10.1007/s10064-016-0937-8>.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep learning*. MIT Press. Retrieved from <http://www.deeplearningbook.org>.
- Han, F., Salgado, R., Prezzi, M., 2015. Nonlinear analyses of laterally loaded piles – A semi-analytical approach. *Comput. Geotech.* 70, 116–129. <https://doi.org/10.1016/j.compgeo.2015.07.009>.
- Han, F., Prezzi, M., Salgado, R., 2017. Energy-based solutions for nondisplacement piles subjected to lateral loads. *Int. J. Geomech.* 17 (11), 04017104. [https://doi.org/10.1061/\(ASCE\)GM.1943-5622.0001012](https://doi.org/10.1061/(ASCE)GM.1943-5622.0001012).
- Henderson, A.R., Morgan, C., Smith, B., Sørensen, H.C., Barthelme, R.J., Boesmans, B., 2003. Offshore wind energy in europe—a review of the state-of-the-art. *Wind Energy* 6 (1), 35–52. <https://doi.org/10.1002/we.82>.
- Howard, J., Ruder, S., 2018. Universal Language Model Fine-tuning for Text Classification (arXiv:1801.06146). arXiv. <http://arxiv.org/abs/1801.06146>.
- Hu, Q., Han, F., Prezzi, M., Salgado, R., Zhao, M., 2022. Finite-element analysis of the lateral load response of monopiles in layered sand. *J. Geotech. Geoenviron. Eng.* 148 (4), 04022001. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0002745](https://doi.org/10.1061/(ASCE)GT.1943-5606.0002745).
- Hu, Q., Han, F., Prezzi, M., Salgado, R., Zhao, M., 2022. Lateral load response of large-diameter monopiles in sand. *Géotechnique* 72 (12), 1035–1050. <https://doi.org/10.1680/jgeot.20.00002>.
- Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., Inman, D.J., 2021. 1D convolutional neural networks and applications: A survey. *Mech. Syst. Sig. Process.* 151, 107398. <https://doi.org/10.1016/j.ymssp.2020.107398>.
- Kulhawy, F.H., Chen, Y.J., 1995. December). A thirty-year perspective of Broms' lateral loading models, as applied to drilled shafts. In: *In Proceedings, Bengt B. Broms Symposium in Geotechnical Engineering*, pp. 13–15.
- Lozano-Minguez, E., Kolios, A.J., Brennan, F.P., 2011. Multi-criteria assessment of offshore wind turbine support structures. *Renew. Energy* 36 (11), 2831–2837. <https://doi.org/10.1016/j.renene.2011.04.020>.
- Man, K., Liu, R., Liu, X., Song, Z., Liu, Z., Cao, Z., Wu, L., 2022. Water leakage and crack identification in tunnels based on transfer-learning and convolutional neural networks. *Water* 14 (9), 1462. <https://doi.org/10.3390/w14091462>.
- Mitelman, A., Urlainis, A., 2023. Investigation of transfer learning for tunnel support design. *Mathematics* 11 (7), 1623. <https://doi.org/10.3390/math11071623>.
- Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)* (pp. 807–814).
- Oh, K.-Y., Nam, W., Ryu, M.S., Kim, J.-Y., Epreanu, B.I., 2018. A review of foundations of offshore wind energy converters: Current status and future perspectives. *Renew. Sustain. Energy Rev.* 88, 16–36. <https://doi.org/10.1016/j.rser.2018.02.005>.
- OuYang, H., Gong, Z., Zhu, M., Liu, H., Liu, Z., Dai, G., Gong, W., 2025. Numerical analysis of lateral bearing capacity of monopiles in clay, part 1: Single uniform and isotropic soil layer with strength profile $s_u = s_{u1} \times z$. *Comput. Geotech.* 184, 107294. <https://doi.org/10.1016/j.compgeo.2025.107294>.
- 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 (12), e0243030. <https://doi.org/10.1371/journal.pone.0243030>.
- Ren, X., Guo, H., Li, S., Wang, S., Li, J., 2017. A Novel Image Classification Method with CNN-XGBoost Model. In: *Kraetzer, C., Shi, Y.-Q., Dittmann, J., Kim, H.J. (Eds.), Digital Forensics and Watermarking, Vol. 10431. Springer International Publishing*, pp. 378–390. https://doi.org/10.1007/978-3-319-64185-0_28.
- Suryasentana, S.K., Lehane, B.M., 2016. Updated CPT-based $p - y$ formulation for laterally loaded piles in cohesionless soil under static loading. *Géotechnique* 66 (6), 445–453. <https://doi.org/10.1680/jgeot.14.P.156>.
- Taherkhani, A., Mei, Q., Han, F., 2023. Capacity prediction and design optimization for laterally loaded monopiles in sandy soil using hybrid neural network and sequential quadratic programming. *Comput. Geotech.* 163, 105745. <https://doi.org/10.1016/j.compgeo.2023.105745>.
- Taherkhani, A.H., Mei, Q. (Gavin), Han, F., 2023b. A Deep Learning Model to Predict the Lateral Capacity of Monopiles, in: *Geo-Congress 2023. Presented at the Geo-Congress 2023, American Society of Civil Engineers, Los Angeles, California*, pp. 220–227. <https://doi.org/10.1061/9780784484685.023>.
- Vázquez, K., Rodríguez, R.R., Esteban, M.D., 2022. Inventory proposal for monopiles in offshore wind farms. *Ocean Eng.* 247, 110741. <https://doi.org/10.1016/j.oceaneng.2022.110741>.
- Wang, H., Wang, L., Zhang, L., 2022. Transfer learning improves landslide susceptibility assessment. *Gondw. Res.* S1342937X22002131. <https://doi.org/10.1016/j.gr.2022.07.008>.
- Xu, L.-Y., Cai, F., Wang, G.-X., Ugai, K., 2013. Nonlinear analysis of laterally loaded single piles in sand using modified strain wedge model. *Comput. Geotech.* 51, 60–71. <https://doi.org/10.1016/j.compgeo.2013.01.003>.
- Zhang, W., Li, H., Li, Y., Liu, H., Chen, Y., Ding, X., 2021. Application of deep learning algorithms in geotechnical engineering: A short critical review. *Artif. Intell. Rev.* 54 (8), 5633–5673. <https://doi.org/10.1007/s10462-021-09967-1>.