

Machine learning-driven design of bored piles: optimization of embedded depth and capacity assessment using spatial soil distribution

Wanhuk Seo, Tae Sup Yun

School of Civil and Environmental Engineering, Yonsei University, Seoul, Republic of Korea, taesup@yonsei.ac.kr

ABSTRACT: Current design methods for bored piles typically rely on empirical formulas and simplified models. These approaches often fail to capture the complex spatial characteristics of soil conditions, leading to potential inaccuracies in bearing capacity prediction and suboptimal determination of embedded depth. This study presents a novel machine learning framework to address these limitations by utilizing spatially distributed soil parameters. A comprehensive dataset combining geotechnical investigation reports and dynamic load test results from 11 construction sites was used to construct a continuous 3D subsurface model through spatial interpolation. Subsurface characteristics such as soil classification, standard penetration test (SPT) N-values, and pile properties were integrated into transformer-based neural network models to establish relationships between ground conditions and pile performance. The proposed framework performs two key tasks: predicting the bearing capacity (end bearing and skin friction) of bored piles and optimizing their embedded depth for target design capacities. The bearing capacity model achieved R^2 values of 0.813 for end bearing and 0.728 for skin friction on test data, while the embedded depth model showed an R^2 of 0.771. This approach significantly enhances conventional pile design by providing a more accurate and efficient tool for foundation engineering, contributing to more economical and reliable construction practices.

KEYWORDS: Bored piles, Machine learning, Spatial soil parameters, Bearing capacity, Embedded depth, Transformer architecture.

1 INTRODUCTION

Bored pile foundations are extensively used in residential construction projects due to their versatility and load-carrying capacity. The accurate prediction of pile bearing capacity and determination of optimal embedded depth are crucial for safe and economical foundation design. Traditional design approaches rely heavily on empirical correlations and simplified analytical methods that often fail to capture the complex spatial variability of geotechnical properties and their influence on pile behavior (Meyrhof 1976; Masouleh & Fakharian 2008).

Current practice typically involves conducting limited dynamic load tests (pile driving analysis, PDA) on selected test piles. This approach presents several critical limitations: (1) Bearing capacity evaluation is performed only on a small fraction (typically less than 10%) of total piles, leaving the majority without direct capacity verification. (2) Empirical formulas tend to overestimate required embedded depths by applying excessive safety factors. In practice, PDA results frequently show that the measured bearing capacity significantly exceeds the target design capacity, indicating unnecessary overdesign in pile length. (3) The spatial uncertainty in soil properties derived from limited borehole investigations is not adequately addressed in the design process (Phoon & Kulhawy 1999).

Recent advances in machine learning, particularly deep neural networks, have shown promising results in various geotechnical engineering applications including pile foundation (Kumar et al. 2024; Pham et al. 2020; Shahin 2016). However, most existing studies have focused on simplified representations of soil conditions without considering the sequential nature of soil stratification and spatial distribution of properties. The transformer architecture, originally developed for natural language processing (Vaswani et al. 2017), has demonstrated exceptional capability in processing sequential data and capturing long-range dependencies, making it particularly suitable for analyzing depth-wise soil property variations.

This study aims to develop an integrated machine learning framework that leverages transformer architecture to process depth-wise sequences of soil properties for accurate prediction of bored pile bearing capacity and optimal embedded depth. The proposed approach explicitly considers the spatial

distribution of soil parameters through 3D interpolation and captures complex soil-pile interaction mechanisms through advanced neural network architectures. By addressing the limitations of current empirical methods, this study contributes to more reliable and economical pile foundation design practices.

2 MATERIALS AND METHODS

2.1 Data collection and preprocessing

2.1.1 Dataset characteristics

A comprehensive dataset was compiled from 432 dynamic load tests performed on bored piles across 11 residential construction sites in South Korea. These sites represent typical geological conditions found throughout the country, characterized by upper reclamation layers or silty sand (SM) layers transitioning to weathered rock and ultimately to granite bedrock. While the data is from specific regional conditions, the proposed methodology is generalizable to various geological settings. The dataset included:

- Borehole investigation data: Location coordinates (x, y), ground elevation, groundwater level, standard penetration test (SPT) N-values and soil classification according to the unified soil classification system (USCS) at vertical intervals of 1.0–1.5 m
- Pile specifications: Location coordinates (x, y), geometric properties (diameter, thickness, cross-sectional area), material properties (elastic modulus, compression wave velocity, unit weight)
- Pile embedded depths: Ranging from about 5 to 30 m, with most piles penetrating through soil layers into weathered rock
- Pile dynamic load test (PDA) results: Time-series force and velocity measurements, impact energy, driving efficiency, and case pile wave analysis program (CAPWAP)-analyzed end bearing and skin friction capacities

2.1.2 Selection of restrike data and data splitting strategy

From the total dataset of 432 tests, only restrike test results (220 samples) were selected for model development. The decision to exclude end of initial driving (EOID) data was based on the

time-dependent nature of pile capacity. Restrike tests, performed after a waiting period following initial installation, allow for full mobilization of both end bearing and skin friction through soil setup effects. This provides more reliable and representative values of long-term pile capacity, which is essential for training accurate prediction models.

The dataset was divided into training, validation, and test sets with a ratio of 8:1:1. This splitting was performed randomly while ensuring that each subset maintained a representative distribution of pile depths, soil conditions, and capacity ranges. The relatively large training set was necessary to capture the complex nonlinear relationships between spatially variable soil properties and pile behavior.

2.2 Spatial interpolation and 3D domain construction

2.2.1 Interpolation methodology

A continuous 3D representation of subsurface conditions was created using spatial interpolation techniques. Both ordinary kriging and inverse distance weighting (IDW) methods were implemented and compared. The ordinary kriging method, with its statistical foundation and ability to provide uncertainty estimates, was ultimately selected for final domain generation. The interpolation process established a regular 3D grid with 2 m spacing in horizontal directions and 1 m spacing in the vertical direction. Data from different boreholes were aligned based on ground elevation to ensure consistent depth references across the site.

2.2.2 Feature extraction at pile locations

For each pile location, depth-wise profiles were extracted from the interpolated 3D domains. This extraction included:

- N-value series sampled at 1 m intervals from ground surface to pile tip
- USCS classification series representing soil type at each meter
- Identification of major stratigraphic boundaries for enhanced feature representation

Figure 1 illustrates the 3D domain construction process and the extraction of soil profiles at specific pile locations. The visualization shows how discrete borehole data are transformed into continuous spatial representations.

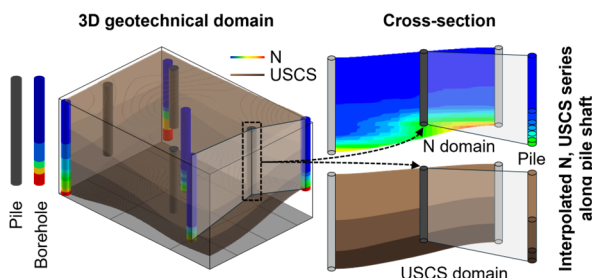


Figure 1. 3D soil domain construction and spatial data extraction process.

2.3 Deep learning model architecture

The proposed framework consists of two main components: a bearing capacity prediction model and an embedded depth prediction model, both using transformer-based architectures to process sequential soil data.

2.3.1 Bearing capacity prediction model

The bearing capacity model employs a dual-output architecture to predict end bearing and skin friction separately, as illustrated in Figure 2. This separation allows each component to be

optimized for its specific physical mechanisms. The model architecture consists of three main components:

- Series data processing: N-value series along the pile shaft are normalized to [0,1] range using min-max scaling. USCS series are first converted from categorical soil types to numerical indices, then transformed into learnable embedding vectors to avoid imposing artificial ordinal relationships between different soil types. A transformer encoder with multi-head attention and positional encoding processes these sequences, with sequence padding and masking applied to handle variable pile lengths.
- Point data processing: Pile properties including diameter, thickness, cross-sectional area, elastic modulus, wave velocity, and unit weight are processed along with embedded depth and tip conditions (N-value and USCS at pile tip) through a multi-layer perceptron (MLP) architecture consisting of 3 hidden layers with layer normalization and Gaussian error linear unit (GELU) activation.
- Feature fusion: Outputs from series and point processing branches are concatenated and passed through residual blocks with skip connections for enhanced gradient flow, with separate output heads generating predictions for end bearing and skin friction.

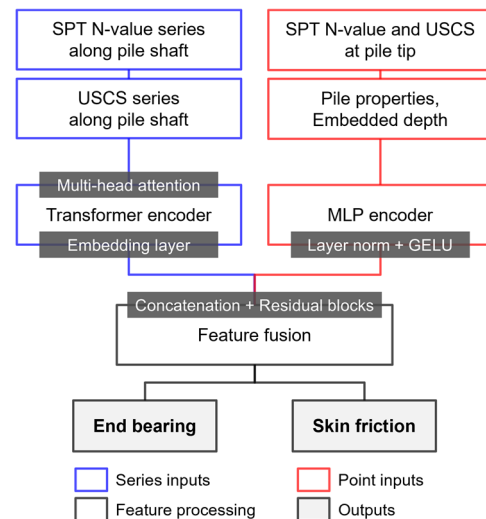


Figure 2. Neural network architecture for bearing capacity prediction (end bearing and skin friction).

2.3.2 Embedded depth prediction model

The embedded depth model predicts the required pile length to achieve a target bearing capacity. While sharing a similar architecture to the bearing capacity model shown in Figure 2, key differences exist in the input configuration.

Instead of processing N-value and USCS series only along the pile shaft, this model receives the complete soil profile from ground surface to bedrock at the pile location. Additionally, rather than embedded depth as a point input, the model takes the target total bearing capacity (sum of end bearing and skin friction) as input. During training, true measured bearing capacities from PDA tests are used as inputs with corresponding embedded depths as ground truth outputs. This enables the model to learn the complex relationship between required capacity and necessary pile penetration depth.

2.3.3 Model optimization

Neural architecture search (NAS) combined with Bayesian optimization was employed to determine optimal

configurations across an extensive search space. Architecture parameters explored included encoder dimensions, attention heads, encoder layers, and hidden layer sizes. Training hyperparameters such as optimizer, learning rate, batch size, dropout rate, and weight decay were simultaneously optimized.

The optimization strategy employed AdamW optimizer with gradient clipping, one-cycle learning rate scheduling with maximum learning rate at 30% of training, and early stopping with patience of 20 epochs based on validation loss. Mean squared error (MSE) was used as the loss function for all prediction tasks. The final optimized configuration consisted of a transformer encoder with 256 dimensions, 16 attention heads, and 3 layers, dropout rate of 0.13, batch size of 64, and learning rate of 8.36×10^{-5} with one-cycle scheduling. The optimized model configurations were then evaluated on the test dataset to assess their generalization performance, as detailed in the following section.

3 RESULTS AND DISCUSSION

3.1 Model performance evaluation

3.1.1 Quantitative metrics

The developed models were evaluated using multiple metrics including coefficient of determination (R^2), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE). Table 1 summarizes the performance metrics for all three prediction tasks on the test dataset.

Table 1. Performance metrics for pile capacity and embedded depth prediction (Units: MAE and RMSE in [kN] for bearing capacity components, [m] for embedded depth)

Metric	End bearing		Skin friction		Depth	
	Train	Test	Train	Test	Train	Test
R^2	0.897	0.813	0.882	0.728	0.975	0.771
MAE	227.1	304.2	227.7	425.3	0.557	1.974
RMSE	311.5	385.7	332.5	531.5	0.831	3.104
MAPE [%]	8.610	10.97	15.44	26.30	5.623	19.15

The models demonstrate strong generalization capability with minimal overfitting, as evidenced by the relatively small gaps between training and test performance.

The end bearing model achieved the highest accuracy ($R^2 = 0.813$ on test set), which can be attributed to the relatively localized nature of tip resistance primarily dependent on soil conditions near the pile tip. The MAE of 304.2 kN and MAPE of 10.97% demonstrate that predictions typically fall within acceptable engineering tolerances. In contrast, skin friction showed lower accuracy ($R^2 = 0.728$) with higher error metrics (MAE = 425.3 kN, MAPE = 26.30%), reflecting the challenge of capturing cumulative soil-pile interactions along the entire shaft length through highly variable soil strata.

For embedded depth prediction, despite a relatively lower R^2 of 0.771, the practical accuracy remains excellent. The MAE of 1.974 m represents only 15.7% of the average pile length (approximately 12.5 m). This level of accuracy is particularly valuable for preliminary design, where traditional empirical methods often result in overconservative estimates with safety factors leading to 20–30% excess length.

3.1.2 Visual performance analysis

Figure 3 presents scatter plots comparing predicted versus true values for all three prediction tasks on both training and test datasets. The plots reveal several important characteristics:

- End bearing prediction (Figs. 3a and b): Shows excellent correlation with minimal scatter, particularly for values

between 2000–4000 kN. The tight clustering around the diagonal line indicates consistent model performance across different capacity ranges. Notably, the model maintains accuracy even for high-capacity piles near 5000 kN, suggesting robust performance within the trained domain.

- Skin friction prediction (Figs. 3c and d): Exhibits greater scatter than end bearing, with a tendency for slight underestimation at higher capacities (> 2000 kN). This pattern reflects the cumulative uncertainties in soil property interpolation along the pile shaft and the complex load transfer mechanisms in heterogeneous soil profiles.
- Embedded depth prediction (Figs. 3e and f): Demonstrates consistent accuracy across the full range of depths. One outlier is observed in the test set, which may be attributed to unusual soil conditions not well-represented in the training data.

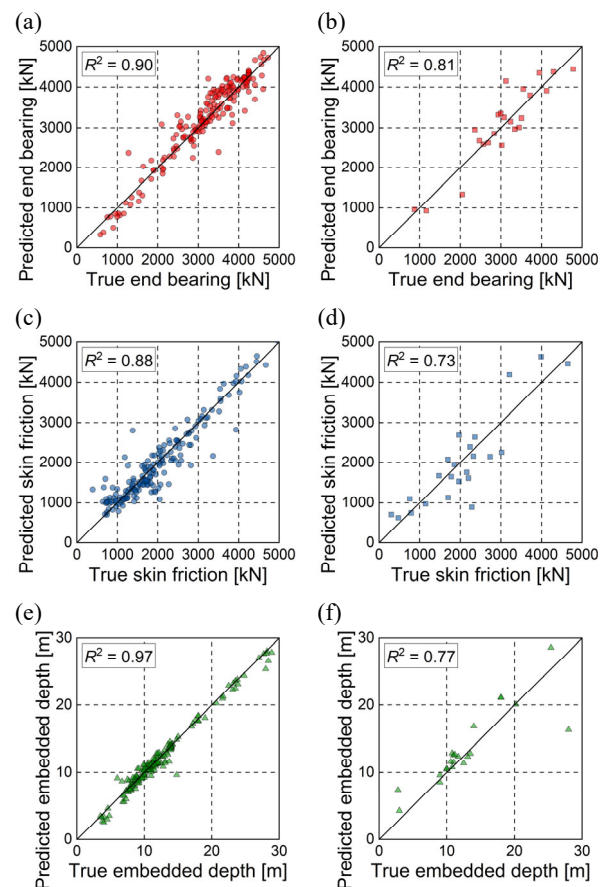


Figure 3. Comparison of predicted versus actual values for (a, b) end bearing, (c, d) skin friction, and (e, f) embedded depth on training and test datasets, respectively.

3.2 Physical interpretation of model behavior

To understand the superior performance of the proposed models, it is essential to examine how the transformer architecture processes the sequential soil data. The transformer architecture's ability to process sequential soil data provides significant advantages over traditional empirical methods. The model can identify long-range dependencies between soil properties at different depths. Multi-head attention enables simultaneous focus on different aspects of the soil profile, such as weak layers affecting skin friction and competent layers controlling end bearing. The model implicitly learns typical stratigraphic patterns and their influence on pile behavior,

improving its relevance to local geotechnical conditions. The model processes soil profiles resampled at 1 m intervals, providing consistent resolution for the transformer architecture. Masking is applied to ensure that only the soil properties within the actual pile length contribute to the predictions, preventing inaccurate correlations from deeper soil layers that the pile does not interact with.

Unlike traditional empirical formulas resulting in conservative designs, the proposed machine learning models adapt to local soil variability through 3D interpolation and process complete soil profiles rather than simplified parameters like average N-values. By predicting end bearing and skin friction separately, the model offers enhanced insight into load distribution, enabling more efficient and targeted optimization of pile design parameters. This approach supports safer and more economical foundation design, particularly in stratified ground conditions where empirical methods may lack precision.

3.3 Practical implications for pile design

3.3.1 Design optimization potential

The developed embedded depth model enables iterative design optimization by accurately predicting the required penetration depth for achieving target bearing capacity. The process begins with using the model to estimate required depth for the design capacity, followed by applying bearing capacity models to verify the design. Based on predicted capacity components, depths can be adjusted and refined to balance material costs against capacity requirements, enabling economic optimization.

3.3.2 Limitations and future work

Several limitations should be acknowledged. The models were trained on typical Korean soil conditions and may require retraining for significantly different geological settings. The current framework is optimized for bored piles, and adaptation to other piles would require consideration of installation effects. Additionally, most training data involved similar pile specifications, potentially limiting model applicability to piles with significantly different specifications.

Future research directions include expansion to diverse geological conditions and pile types, integration of uncertainty quantification methods, development of real-time optimization algorithms for construction phase adjustments, and investigation of transfer learning approaches for data-scarce regions.

4 CONCLUSIONS

This study successfully demonstrates the application of transformer-based machine learning architectures for optimizing bored pile design through accurate prediction of bearing capacity and embedded depth. The transformer-based approach effectively processes depth-wise sequences of soil properties, capturing complex stratigraphic variations and their influence on pile behavior. The multi-head attention mechanism enables identification of relevant soil layers and their interactions, providing significant advantages over traditional empirical methods.

The developed models achieved high prediction accuracy with R^2 values of 0.813 for end bearing, 0.728 for skin friction, and 0.771 for embedded depth on test datasets, demonstrating robust generalization capability. By integrating spatially interpolated 3D soil domains with pile properties, the proposed framework provides a comprehensive approach to pile design that addresses the limitations of simplified soil representations in traditional methods. The separate prediction of bearing capacity components enables iterative design refinement,

potentially reducing pile lengths while maintaining required safety levels.

The practical implications of this research are substantial. The framework enables bearing capacity evaluation for all piles rather than just selected test piles, and optimizes embedded depths based on actual soil conditions rather than conservative empirical formulas. This contributes to more economical and reliable foundation design, with potential cost savings through reduced pile lengths and improved capacity utilization. With inference times under one second and standard computational requirements, the models can be readily integrated into existing design workflows.

Future work should focus on expanding the dataset to include diverse geological conditions beyond typical Korean soil profiles, and developing uncertainty quantification methods to provide confidence intervals for predictions. As the construction industry increasingly embraces digital transformation, such data-driven approaches will become essential for optimizing foundation systems while maintaining safety and reliability. The success of this framework demonstrates the potential for machine learning to transform geotechnical design practices.

5 ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Nos. RS-2021-NR060085, RS-2023-NR076991).

6 REFERENCES

- Kumar, M., Kumar, D. R., Khatti, J., Samui, P., and Grover, K. S. 2024. Prediction of bearing capacity of pile foundation using deep learning approaches. *Frontiers of Structural and Civil Engineering* 18(6), 870-886.
- Masouleh, S. F., and Fakharian, K. 2008. Application of a continuum numerical model for pile driving analysis and comparison with a real case. *Computers and Geotechnics* 35(3), 406-418.
- Meyerhof, G. G. 1976. Bearing capacity and settlement of pile foundations. *Journal of the Geotechnical Engineering Division* 102(3), 197-228.
- Pham, T. A., Ly, H.-B., Tran, V. Q., Giap, L. Van, Vu, H.-L. T., and Duong, H.-A. T. 2020. Prediction of pile axial bearing capacity using artificial neural network and random forest. *Applied Sciences* 10(5), 1871.
- Phoon, K. K., and Kulhawy, F. H. 1999. Characterization of geotechnical variability. *Canadian Geotechnical Journal* 36(4), 612-624.
- Shahin, M. A. 2016. State-of-the-art review of some artificial intelligence applications in pile foundations. *Geoscience Frontiers* 7(1), 33-44.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. 2017. Attention is all you need. *Advances in Neural Information Processing Systems* 30.