

RESEARCH ARTICLE

A novel machine learning-based approach to thermal integrity profiling of concrete pile foundations

Javier Sánchez Fernández^{1,2} , Agustín Ruiz López^{2,3} and David M.G. Taborda²

¹Science and Solutions for a Changing Planet DTP, Imperial College London, London, UK

²Department of Civil and Environmental Engineering, Imperial College London, London, UK

³Seequent—The Bentley Subsurface Company, Delft, The Netherlands

Corresponding author: Javier Sánchez Fernández; Email: js8918@ic.ac.uk

Received: 28 January 2025; **Revised:** 16 April 2025; **Accepted:** 04 June 2025

Keywords: concrete piles; defect detection; finite element; machine learning; thermal integrity profiling

Abstract

Thermal integrity profiling (TIP) is a nondestructive testing technique that takes advantage of the concrete heat of hydration (HoH) to detect inclusions during the casting process. This method is becoming more popular due to its ease of application, as it can be used to predict defects in most concrete foundation structures requiring only the monitoring of temperatures. Despite its advantages, challenges remain with regard to data interpretation and analysis, as temperature is only known at discrete points within a given cross-section. This study introduces a novel method for the interpretation of TIP readings using neural networks. Training data are obtained through numerical finite element simulation spanning an extensive range of soil, concrete, and geometrical parameters. The developed algorithm first classifies concrete piles, establishing the presence or absence of defects. This is followed by a regression algorithm that predicts the defect size and its location within the cross-section. In addition, the regression model provides reliable estimates for the reinforcement cage misalignment and concrete hydration parameters. To make these predictions, the proposed methodology only requires temperature data in the form standard in TIP, so it can be seamlessly incorporated within the TIP workflows. This work demonstrates the applicability and robustness of machine learning algorithms in enhancing nondestructive TIP testing of concrete foundations, thereby improving the safety and efficiency of civil engineering projects.

Impact Statement

This study provides a new approach to interpreting thermal integrity profiling (TIP) using numerical methods and machine learning techniques. By enabling the early identification and characterization of soil and bentonite defects that might otherwise go undetected, the proposed method improves the reliability of TIP as a non-destructive technique to evaluate the integrity of concrete piles. The implications of this research include the potential for more reliable and higher-quality pile construction, reducing the risk of defective pile failures, contributing to the broader adoption of nondestructive testing technologies, and improving industry standards for safety, efficiency, and environmental responsibility.

1. Introduction

Bored piles are one of the most common piling techniques used for structure foundations, sometimes constructed using bentonite support fluids, as first described by Veder (1953). It improves the construction

 This research article was awarded Open Data badge for transparent practices. See the Data Availability Statement for details.

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (<http://creativecommons.org/licenses/by-nc/4.0>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use.



of bored piles through the stabilization of the walls and reduces the likelihood of soil collapse. One of the most common construction defects resulting from this method consists of pockets of bentonite getting trapped throughout the shaft, preventing complete placement of concrete from the intended pile volume. It is important to evaluate the integrity of these foundations effectively, as they play a vital role in the safety and durability of the structures they support. As an example, voids occupying ~15% of the cross-sectional area in the central region of a pile can result in a reduction of its flexural resistance by 27–32% (Sarhan et al., 2002).

Some common nondestructive techniques used to evaluate the integrity of concrete foundations are based on the application of different physical principles, including the transmission of ultrasound waves, low-frequency acoustic waves, or electromagnetic waves. Popular options include cross-hole sonic logging (CSL), impulse response, and low-strain pile integrity testing (ASTM, 2016a; ASTM, 2016b; French and Whittaker, 2023; Zhai et al., 2023). These methods can offer detailed subsurface information for many types of foundations. CSL is particularly effective in detecting defects between access tubes, while low-strain integrity testing and impulse response testing, typically performed by applying an impulsive force at the pile head and analyzing the reflected waves, are better at evaluating major defects near the surface.

Each method, however, has its limitations. CSL, in addition to requiring the preinstallation of access tubes, is restricted to detecting defects that are located between access tubes, often missing smaller anomalies and struggling with defects outside the direct test path. There are also some concerns regarding its high false positive rate and poor integrity evaluation outside the rebar cage (Boeckmann and Loehr, 2019). Lozovsky et al. (2024) recommend combining CSL with other nondestructive methods and conclude that the main way to overcome this method's limitations is by adding more access tubes. Low-strain testing and impulse response, while valuable for general assessments, struggle to predict any defects in the lower portions of the pile and require careful interpretation of complex signals from varying soil conditions (Finno et al., 2002). Furthermore, ASTM (2016a) states that this method may not be able to identify all imperfections, but rather major defects along the effective length. When aiming to detect necking or weak concrete, any defects smaller than 0.5 m are very challenging to detect (Liu et al., 2019). Conclusions from Briaud et al. (2002) and Lim and Cao (2013) indicate that none of the testing techniques mentioned above are fully dependable at predicting defects or pile lengths, with further usage or a combination of different methods being needed to improve their reliability.

Other techniques, like ground penetrating radar, can be used in conjunction with CSL for a more precise evaluation of defects. This is most useful in other concrete foundation types, such as retaining walls. The minimum defect size detectable with CSL is typically 5–10% of the pile's cross-sectional area, while it is expected that defects occupying between 9 and 20% of the pile's cross-sectional area be detected with certainty (Li et al., 2005).

In recent years, alternative methodologies have been developed to improve the accuracy and reliability of integrity tests. Thermal integrity profiling (TIP) is a concrete pile testing technique that measures the temperature distribution caused by concrete hydration along the shaft depth (Mullins and Kranc, 2004; Mullins, 2010). Since cement hydration is an exothermic reaction, the heat generated dominates the internal temperature distribution of the concrete during hydration. This is a crucial factor for its early-stage performance, strength development, and long-term durability (Neville, 1986). This process affects curing, mechanical properties, and can lead to potential cracking due to thermal gradients if not adequately controlled. A reliable method to estimate the heat of hydration (HoH) is, therefore, essential, not only to predict the curing behavior and mechanical outcomes but also to improve the interpretation of TIP results by allowing for more precise identification of temperature anomalies that may indicate soil or bentonite defects (Liu et al., 2022).

The core concept behind TIP is that defects in piles reduce the total HoH and can, therefore, be detected from carefully analyzing temperature against depth data, measured at different times. In effect, having deviations from the expected temperature profile, such as cooler areas, may indicate an absence of concrete mass and thus suggest the presence of defects. By accurately identifying these anomalies, remedial action can be planned to ensure that the pile satisfies the performance requirements. A careful

analysis of the temperature field measured within a given section can also establish lateral cage misalignments arising from the cage installation process (Johnson, 2016; Sun et al., 2021). The effectiveness and accuracy of TIP have been extensively validated (Mullins et al., 2007; Winters, 2014; Boeckmann and Loehr, 2019).

The standard procedure for TIP consists of attaching thermocouples to the reinforcement cage, with a longitudinal spacing of ~300 mm. These are then connected to a temperature logging device, and the temperature versus depth graphs are analyzed post-hydration. A minimum of four testing locations within the pile cross-section should be provided. Testing procedures and specifications are outlined in ASTM D7949 (ASTM, 2014).

Recent technological advancements have significantly improved TIP, particularly regarding the measurement equipment. One notable development is the use of fiber optics for temperature readings, replacing thermocouples, through a technique called distributed temperature sensing (DTS). Using DTS instead of traditional thermocouples provides measurements that are more precise, higher spatial and temporal resolution, and thus more reliable detection of defects along the pile depth. Studies have demonstrated that DTS can better detect anomalies, such as soil and bentonite inclusions or other structural deficiencies, by providing a more detailed thermal profile (Xiao et al., 2017; Liu et al., 2020; Zhong et al., 2018; Wang et al., 2023). Despite its advantages, it is worth mentioning that DTS still requires adequate equipment calibration to ensure accurate readings, and that it does not provide informative data about areas far from any sensor.

The detection of defects through the interpretation of TIP data is far from straightforward. Temperature readings are affected by several factors, such as the soil type, concrete mix, pile diameter, and misalignments of the reinforcement cage. The size and location of the defects are also very important, and small defects located near the center of the pile might not cause significant temperature variations close to the pile edge, where the sensors are installed, and so can be difficult to detect. In practice, the interpretation of the readings often simply involves a comparison of temperatures at a given section with the average temperature at that section or along the pile. Alternatively, the temperature measurements can be compared with finite element (FE) predictions of the temperature evolution during hydration. While FE models adopting well-calibrated HoH models (Schindler and Folliard, 2005; De Freitas et al., 2013; Liu et al., 2022) have been shown to be accurate, this interpretation approach is still subjective and does not yield estimates of the size and location of the defects.

This article addresses the shortcomings described above by proposing a new data-driven approach that can determine whether a pile is defective and the extent of a potential defect from a standard set of TIP measurements. By leveraging the ability of machine learning (ML) to identify complex patterns between input and output data, the developed methodology allowed the relationship between the temperature fields and the various influencing factors outlined above to be established. Two artificial neural networks (ANNs) were developed using FE analysis results of the HoH of concrete piles as training data. Given a set of temperature measurements, the first ANN classifies piles as defective or not. This is followed by a regression ANN that predicts the size and location of the pile defect, the thermal properties of the concrete, and the misalignment of the reinforcement cage. The robustness and predictability of the developed ML-based approach are assessed thoroughly, including the use of noisy data, hence ensuring its reliability and generalization ability for real-world applications. This research amplifies significantly the impact and scope of TIP, bridging the gap between data collection and quantitative prediction and providing considerable insights into the state of newly constructed concrete foundations.

2. Proposed approach

As outlined above, the working principle behind TIP is to identify defects based on HoH temperature discrepancies within concrete pile foundations. Existing research has not yet produced a comprehensive model capable of predicting defect inclusions in concrete piles, accounting for a wide range of concrete and soil properties, as well as reinforcement cage misalignments. The proposed methodology aims to overcome these challenges by training an ML model with the results from two-dimensional (2D) simulations of the HoH generated by concrete piles. The outputs of these simulations are processed and used to train two

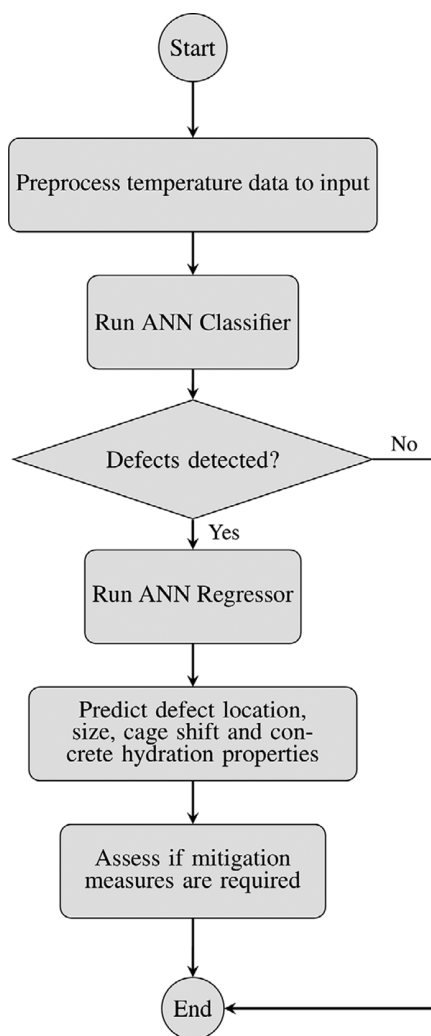


Figure 1. Flowchart of the defect detection model.

ANNs: a classifier and a regressor. Figure 1 shows a flowchart of the ANN prediction process. Once these models are available, the proposed methodology is set to fulfill the following tasks:

1. Classify piles as defective or nondefective.

For samples classified as defective:

2. Determine the defect size and its location.
3. Predict reinforcement cage misalignments.
4. Back-calculate concrete HoH parameters used in the numerical simulation, thereby enabling further predictions of other concrete structures built using the same concrete mix.

2.1. Numerical modeling of HoH

Several theoretical models have been proposed to calculate HoH by considering cement's chemical reaction kinetics, but they require a large number of parameters (Parrot and Killoh, 1984; Kishi and

Maekawa, 1996; Swaddiwudhipong et al., 2002). Schindler and Folliard (2005) proposed a model with fewer parameters, where the HoH of a specific mixture can be characterized from the activation energy, the degree of hydration development at the reference temperature, and the total HoH. It makes use of the “equivalent age” concept, described as a way to normalize the concrete curing progress over time, given the varying temperature conditions (Schindler and Folliard, 2005). A similar model is proposed by Liu et al. (2022), which has been calibrated and validated against field measurements of hydration temperature in piles. Numerical modeling of HoH has become a valuable tool in predicting the thermal behavior of concrete during the early stages of curing. Some researchers have computationally modeled the exothermic reaction using empirical equations (Xie and Qian, 2023; Tan and Tang, 2024).

Heat transfer in soils is predominantly governed by conduction, and thus it will be the only mode considered in this study. The transient heat conduction equation is described in Equation (1) as follows:

$$\rho c_p \frac{\partial T}{\partial t} - k \nabla^2 T = P_{hoh} \quad (1)$$

where $\rho [kg \cdot m^{-3}]$, $c_p [J \cdot kg^{-1} \cdot K^{-1}]$, and $k [W \cdot m^{-1} \cdot K^{-1}]$ are the density, volumetric heat capacity, and thermal conductivity of the material, respectively, and $P_{hoh} [W \cdot m^{-3}]$ is the HoH power per unit volume, which serves as the only internal heat source. The HoH of concrete curing is calculated herein based on the formulation given in Liu et al. (2022). This model establishes a relation between the HoH power per unit volume and the accumulated HoH since the start of hydration. The model accounts for key factors, such as the composition of cement, temperature, and reaction kinetics, but with fewer parameters than other models previously mentioned, simplifying its use in practical applications. The key aspects of the model formulation are summarized in the following, with further details available in Liu et al. (2022).

An S-shaped curve describes the evolution of the reference-accumulated heat per unit mass of cement ($Q_{ref} [J \cdot kg_{cement}^{-1}]$) with time (van Breugel, 1997):

$$Q_{ref}(t) = Q_{max} e^{-\left(\frac{t}{\tau}\right)^\beta} \quad (2)$$

where $Q_{max} [J \cdot kg_{cement}^{-1}]$ is the maximum accumulated heat per unit mass of cement, and $\beta [-]$ and $\tau [s]$ are model parameters to fit observed hydration data.

The heat accumulation law (Equation (2)) is subsequently differentiated with respect to time to establish a power per unit volume, with a further correction being added to introduce the effect of temperature via the activation energy $E_a [J \cdot mol^{-1}]$, resulting in:

$$P(T, t) = \chi \rho_{conc} \frac{dQ_{ref}}{dt} e^{-\frac{E_a}{RT} \left[\frac{1}{T} - \frac{1}{T_{ref}} \right]} \left(\frac{Q_{max} - Q_{acc}(t)}{Q_{max} - Q_{ref}(t)} \right)^{n^*(t)} \quad (3)$$

where $P(T, t) [W \cdot m^{-3}]$ is the HoH power as a function of time and temperature, $R [J \cdot mol^{-1} \cdot K^{-1}]$ is the universal gas constant, and $n^*(t)$ is a factor that encapsulates the effect of reaction kinetics. The accumulated HoH, $Q_{acc}(t)$, is given by:

$$Q_{acc}(t) = \frac{1}{\chi \rho_{conc}} \int_0^t P(T, t) dt \quad (4)$$

where $\chi [kg_{cement} \cdot kg^{-1}]$ is the mass of cement per unit mass of concrete and $\rho_{conc} [kg \cdot m^{-3}]$ is the concrete density. Solving these equations gives an accurate way to predict the HoH of concrete piles.

Given its relatively simple calibration and reliance on a limited number of parameters, the model is well suited for field applications requiring rapid predictions, such as temperature control during curing and, most importantly, TIP of deep foundations (Liu et al., 2022).

The model was implemented with the heat transfer module in COMSOL Multiphysics, which was used to conduct transient 2D-plan thermal analyses of the HoH process. The 2D approximation assumes that heat fluxes in the out-of-plane direction are negligible due to the large pile depth compared to its other dimensions. This assumption holds because the applied heat and material properties remain constant

along the pile depth (for specific soil layers), with heat transfer occurring solely within the plane containing the cross-section. The only exception for this is the end effect, estimated to be of 1-diameter depth (Sun et al., 2021). A parametric, time-dependent study is run across different geometric configurations.

The geometry consisted of a concrete pile centered within a circular soil domain of 15 m radius. The boundary condition on the outer edge of the soil domain was set to a constant temperature throughout the simulation. The size of the domain was chosen to be sufficiently large to prevent any boundary effects from affecting the heat distribution on the pile while limiting the computational effort required. The initial temperature across the entire domain was assumed to match that of the poured concrete and was included in the model as an additional variable.

The HoH was modeled only within the concrete pile domain, with heat transfer occurring through conduction in both the concrete and the surrounding soil. In samples where a defect is included, no HoH occurs in this domain. The mesh was generated using second-order triangular elements for solving the heat transfer problem, with a higher resolution applied to the concrete pile and its vicinity. The meshing process was conducted using the COMSOL in-built mesher, and a sample mesh is shown in Figure 2. Given that the model is generated parametrically, several cases considered in this study were inspected to ensure that adequate spatial discretizations were always obtained, particularly when the edge of the defect was located close to that of the pile. Each sample was modeled with $\sim 1,000$ – $1,600$ elements, depending on the radius adopted for the pile.

2.2. Numerical data

A dataset comprising 100,000 simulations was produced by adopting the numerical model described above. It included variations in a wide range of geometric and material parameters, listed in Table 1. These

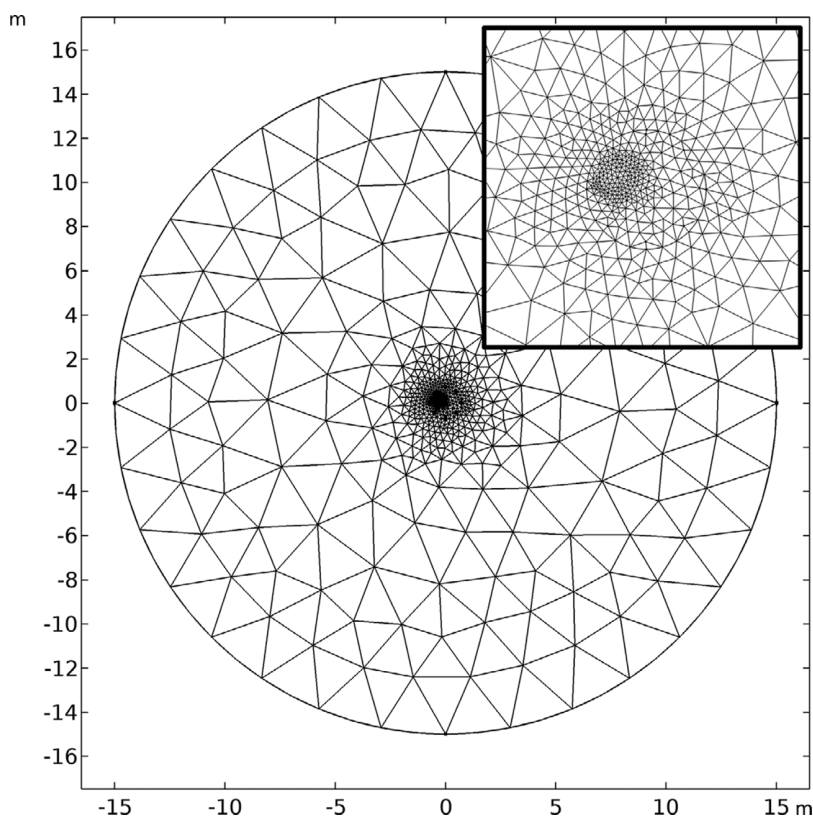


Figure 2. Sample mesh geometry.

Table 1. Numerical model parameters

Parameter	Lower bound	Upper bound	Unit	Description
θ_{defect}	0.0	2π	[rad]	Angular position of the defect
$r_{d,\text{def}}$	0.0	1.0	[–]	Ratio of radial distance of the defect centroid to the pile center
r_{def}	0.1414	0.8	[–]	Ratio of defect radius to pile radius
λ_{soil}	0.5	5.0	[W/(m K)]	Thermal conductivity of soil (Grirate et al., 2016; Zhang et al., 2017; Wu et al., 2025)
λ_{conc}	0.5	4.0	[W/(m K)]	Thermal conductivity of the concrete (Asadi et al., 2018; Lu et al., 2021; Di Girolamo et al., 2022)
λ_{def}	0.5	5.0	[W/(m K)]	Thermal conductivity of the defect
$C_{p,\text{soil}}$	1,000	3,000	[kJ/(m ³ K)]	Volumetric heat capacity of soil (Kodesová et al., 2013)
$C_{p,\text{conc}}$	1,500	2,500	[kJ/(m ³ K)]	Volumetric heat capacity of concrete (Tatro, 2006)
$C_{p,\text{def}}$	1,000	3,000	[kJ/(m ³ K)]	Volumetric heat capacity of the defect
r_{pile}	0.3	1.2	[m]	Pile radius
θ_{cage}	0.0	2π	[rad]	Angular position of the cage shift
r_{cage}	0.0	0.065	[m]	Radial distance of the cage shift
β	0.35	1.05	[–]	HoH model parameter
τ	35,000	105,000	[s]	HoH model parameter
Q_{max}	150	450	[kJ/kg]	Maximum heat generation capacity of the concrete mix
$\chi\rho$	300	500	[kg/m ³]	Mass of cement in concrete
T_{pour}	10	25	[°C]	Initial (concrete pour) model temperature

include the defect geometric properties, thermal conductivities, and volumetric heat capacities of the materials, pile, and cage displacement properties, and hydration model parameters. In this study, material properties are assumed to be homogeneous across the domain. The input values for each of the simulations were established using the near-random Latin hypercube sampling approach. This statistical method, first introduced by McKay et al. (1979), is used to sample points from a multidimensional space, ensuring that each parameter’s range is well represented across the entire dataset. It divides each parameter’s distribution into equal intervals and selects sample points such that all intervals are covered without repetition, improving sampling efficiency in large models. This ensured that all geometric and material parameter combinations were representative of the real-world variability of data in a computationally efficient way, thereby enabling the model to predict defects among a wide range of soil, concrete, and defect types. The lower and upper bounds for each of the parameters being varied are provided in Table 1. The parameter range defined by these bounds was deliberately made very large, hence ensuring that they covered any potential parameter combination to be encountered in practice. Even though some combinations are unlikely to occur in real-life conditions, this broader range is beneficial for the generalization of the model, thus preventing overfitting to a narrow subset of conditions.

Soil or bentonite defects were modeled as circles in the cross-sectional plane. Complex defect shapes were not evaluated throughout this study, as the impact of a defect on the HoH generation is more dependent on its area (or volume when considering three-dimensional [3D] problems) rather than its shape. Only defects with an equivalent area >2% of the pile cross-sectional area were considered, as the temperature variations generated by smaller defects would be too small to allow identification, particularly when these are located away from the probes, that is, toward the center of the pile. Furthermore, it could be argued that these are not defects at all. In turn, the largest modeled radius of the defect is 80% of that of the pile (i.e., equivalent to 64% of the total pile area). Due to the randomized nature of the defect parameter generation, the defect-to-pile radius ratios follow a uniform distribution. When these ratios are

converted to defect radii and the associated defect areas are calculated, 53% of defects have an equivalent area of <20% of the pile's cross-sectional area, thus aligning with the most commonly observed defect size in practice (Sarhan et al., 2002; Sun et al., 2024). The material properties of the defect are generated independently from the surrounding soil or concrete, accounting for other alternatives, such as loose soil or bentonite.

In order for the dataset to be most relevant to real-world scenarios, where the presence of a defect within the pile would not be known a priori, samples with and without defects were included. This approach enables the ANN models to generalize beyond defect-only cases, producing predictions that are more robust.

Temperature values were measured across eight probe locations spaced radially at equal angles, as shown in Figure 3. The temperature probes are located near the edge of the concrete pile, leaving a 75 mm cover. As the outer part of the shaft is the most relevant for its bending capacity, identifying defects in these regions is particularly important (Mullins, 2010). Temperature values from each of these probe locations were stored every 0.05 days for the first 5 days after pouring, leading to 101 temperature values per probe, per sample. A node was placed at each of the probe locations such that the temperature values were established directly from the FE calculation, rather than relying on interpolation.

A total of 586 parameter combinations, including a defect, could not be computed because of meshing complications; they all corresponded to cases where the edge of the defect was very close to one of the probes. The affected samples represent ~0.6% of the total. Latin hypercube sampling ensures that samples are distributed across the entire range of the input variables, making it robust to small changes (Forrester et al., 2008). This suggests that even if a few samples are removed, the remaining dataset can still be representative of the design space due to the stratified nature of the sampling technique, and therefore, the parameters can still be assumed to be randomly distributed.

When pile rebar cages are inserted into bored piles, they can displace from the design position. To account for this effect, two parameters, θ_{cage} and r_{cage} are generated to simulate the potential cage shift as a function of the angular direction and radial distance from the origin. These define the angle of cage shift and its magnitude, ranging between $0-2\pi$ and $0-6.5$ cm, respectively.

The defect area and the centroid's Cartesian coordinates were calculated analytically. Defects could overlap the soil domain, potentially leading to noncircular shapes. In cases when these defects partly extend beyond the edge of the concrete pile, they were trimmed such that only the portion originally within the concrete section was considered as defect, with the material properties being assigned

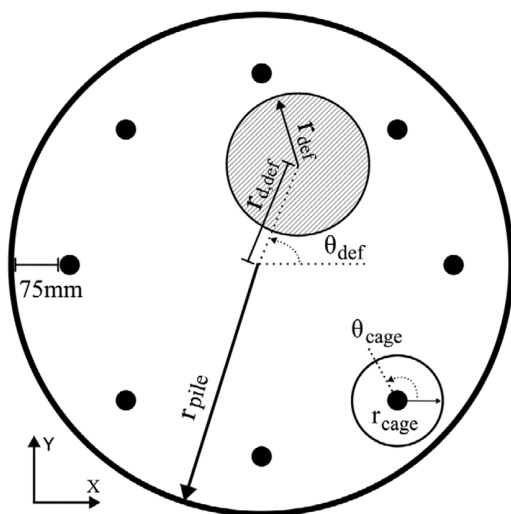


Figure 3. Pile and defect geometry.

accordingly, whereas the remaining portion of the circle (i.e., outside the concrete section) was assumed to be soil. The area and centroid coordinates were then calculated using Monte Carlo integration, a statistical method commonly used to estimate the area of complex shapes, although analytical solutions exist for this specific geometric problem. This involves the random sampling of points (N_{tot}) within the region bounded by the concrete pile and defect, and counting how many falls within the intersecting section. The proportion of points (N_{in}) that fall within the intersection of the two circles (pile and defect) was used to estimate the area of the defect (Equation (5)), while the average coordinate of the points inside the intersection provided an estimate of the section centroid, as shown in Equations (6) and (7). To produce accurate area and centroid estimations, four million points were randomly generated for each sample.

$$A \approx \frac{N_{in}}{N_{tot}} \cdot A_{bound} \quad (5)$$

$$x_c = \frac{1}{A} \int_A x dA \approx \frac{1}{N} \sum_{i=1}^N x_i \quad (6)$$

$$y_c = \frac{1}{A} \int_A y dA \approx \frac{1}{N} \sum_{i=1}^N y_i \quad (7)$$

2.3. ML model

In order to develop the proposed methodology for defect detection in bored piles using TIP, two ANN models were created: the first to establish whether a defect is present, and the second to estimate the defect size and location, cage shift, and concrete properties of the sample pile when a defect is identified. Figure 4

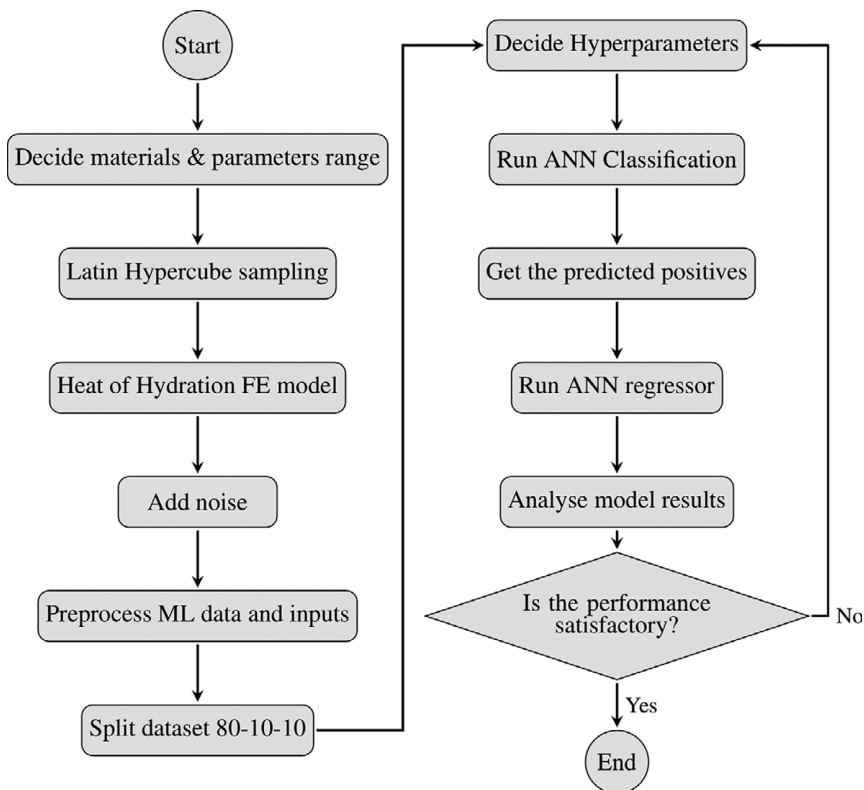


Figure 4. Flowchart showing the ANN creation process.

summarizes the process to generate data and train the ANN. The ANN models were built using the deep learning package *Keras* (Chollet, 2021) and the procedures employed to do so are detailed in the following.

2.3.1. Data processing and ML inputs

The FE output for all samples and probes was preprocessed before being passed to the ML training process. To decide on the model inputs, a statistical analysis of temperature values was conducted for each probe, including the maximum temperature recorded at any time, the average temperature over the 5-day simulation period, the maximum temperature differences between all probes, and the time instants at which these maximum temperatures occurred. The concrete ratio of the mix, initial temperature, and pile radius were also used for model training. These were included as inputs for both the classification and regression modes. In addition to those inputs, the regression ANN model also incorporated temperature data recorded at intervals of every 0.25 days between $t = 0.25$ and $t = 2$ days, as well as at $t = 3, 4$, and 5 days for each probe. It is worth noting that by using exclusively temperature values as inputs to the ANN models, the proposed approach does not require any field or laboratory testing to assess hydration parameters, which are, in fact, returned by the regression ANN. The dataset used throughout this publication is available on Sánchez Fernández (2025).

To maximize the performance of the models, it is important to consider the different input parameters that have the greatest impact on the resulting output. Spearman's correlation coefficient (r_s , see Equation (8)) was used to study these relationships. This coefficient measures the relationship between two continuous variables, ranging from -1 to $+1$. Unlike Pearson's correlation coefficient, which assumes a linear relationship and normally distributed data, Spearman's correlation coefficient measures the monotonic relationship between two variables, regardless of whether they are linear or not, by comparing their rank orders instead of their actual values. Spearman's correlation is more robust to outliers than Pearson's, as it evaluates the relative ranking of the data points rather than their magnitude (Moscarelli, 2023).

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (8)$$

In Equation (8), d_i is the difference in the ranks of corresponding data points, and n is the number of data points. Figure 5 shows the Spearman's correlation coefficients for the regression model inputs and outputs, while Figure 6 presents the Spearman's correlation coefficient for the input variables of the classifier algorithm. Each individual cell in the heatmap corresponds to a probe value within the overall cluster. The results indicate that most variables exhibit weak correlations, except for the average and maximum temperatures, which are naturally highly correlated as they arise from the same time series. Despite their correlation, these inputs are essential for capturing the characteristics of the temperature field within the hydrating pile problem, which is fundamental to the model's performance at predicting the presence, size, and location of defects.

In addition to the possibility of reinforcement cage misalignment, which affects the position at which temperatures are measured, the accuracy of temperature readings for TIP is a critical consideration. Measurement inaccuracies can significantly affect the thermal profiles and their subsequent assessment and interpretation. Introducing noise serves as a powerful generalization mechanism that improves the reliability of ML outputs by simulating the variability typically observed in real-world data (Vogt et al., 2010). By adding controlled perturbations to the raw deterministic data, the model learns patterns more robustly, leading to reduced overfitting. In TIP measurements, most of the noise comes from the sensor accuracy during testing.

Raman-based distributed fiber optic sensors (DFOS) are the preferred choice when using nondestructive integrity evaluation techniques such as DTS (Hartog, 2017). Some distributed temperature-sensing models, such as the Sentinel DTS, can achieve temperature measurement accuracies as fine as 0.01°C , with resolutions of 1 m for up to 45 km (Sensornet, 2021). Other sensors, summarized in Table 2, also

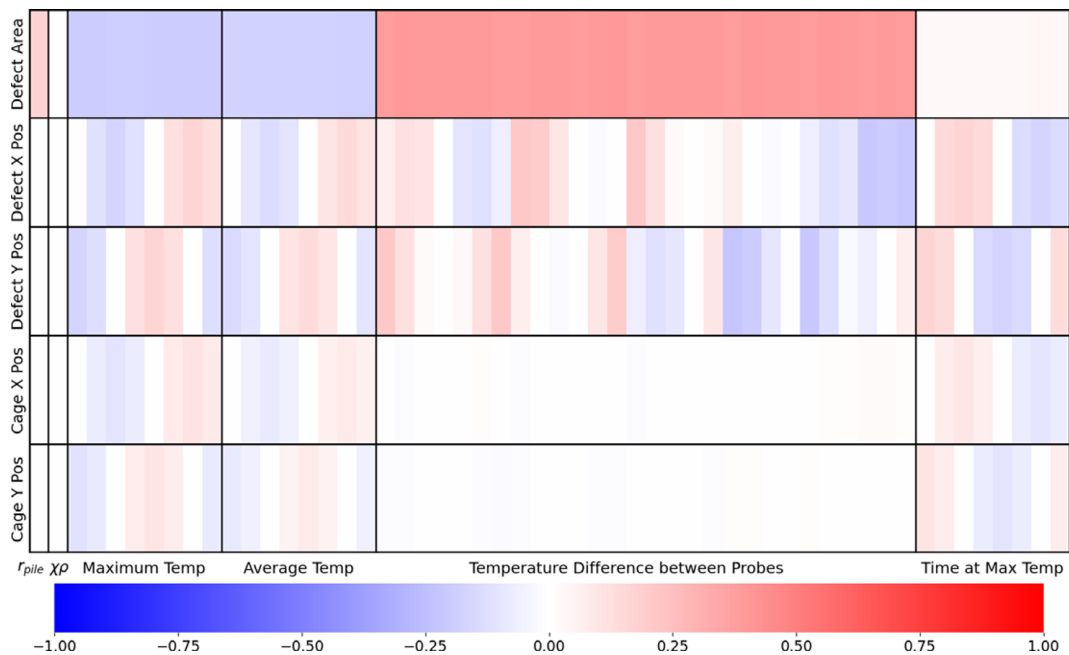


Figure 5. Spearman's correlation coefficient between the model inputs and outputs.

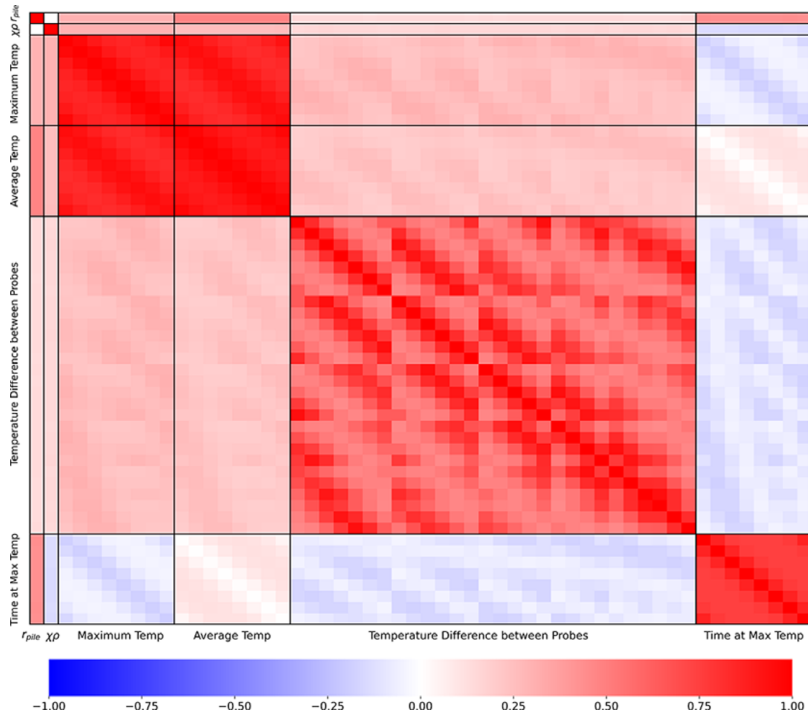


Figure 6. Spearman's correlation coefficient for the model inputs.

provide very high accuracies and resolutions. In addition to sensor accuracy, in practice, spatial resolution (SR) plays a large role in DFOS temperature readings (Bao and Chen, 2012). Since Raman-based DFOS returns a weighted average temperature across its SR, discrepancies between actual and measured temperatures can arise (Pradhan and Sahu, 2015). However, as the model in this study relies on 2D numerical data, SR effects cannot be explicitly incorporated. Therefore, to account for these effects, a normally distributed noise with a standard deviation of 0.1°C will be added to all temperature readings, resulting in 99.7% of the noise values being within $\pm 0.3^\circ\text{C}$. This level of accuracy, in line with that of the sensors listed in Table 2, aids in ensuring the generalization of the ML model while maintaining realistic conditions for TIP. Introducing greater variation would potentially result in predictions that would be unreasonably inaccurate when compared to those in the field, as high-accuracy equipment is an inherent requirement for the method, both with respect to SR and sensor accuracy.

2.3.2. Classification algorithm

An ANN was developed to perform the binary classification of piles into defective and nondefective, using the “KerasClassifier” algorithm. The dataset was split into training, validation, and testing sets, using an 80-10-10 split to ensure sufficient data is provided for both model training and validation. In terms of the number of samples, this corresponds to 79531-9941-9942. Before training, all input features were standardised using z-score normalization. The architecture of the ANN was established following a grid search approach (Raschka, 2017). The best architecture was identified based on its performance from a 10-fold cross-validation. It consisted of three densely connected hidden layers with neuron counts of 3,000, 1,500, and 300, respectively. Rectified linear unit (ReLU) activation functions were used in all hidden layers, allowing the model to learn nonlinear patterns in the data. A sigmoid activation function was applied to the output layer to perform binary classification (defect or no defect). The loss function was set to binary cross-entropy. The model was trained using the Adam optimizer with an initial learning rate of 0.001. A learning rate reduction strategy was also implemented, where the learning rate was reduced by

Table 2. Summary of different temperature sensor accuracies as reported in the literature

Publication	Sensor type	Model	Temperature accuracy (°C)	Spatial resolution (m)	Sensing range (km)
Chen et al. (2015)	DFOS (Raman)	–	± 0.2	0.93	2.5
Soga and Luo (2018)	DFOS (Raman)	Halo-DTS	± 0.45 (2σ)	2	4
		LIOS-EN- SURE-OTS700	± 1.0 (2σ)	1–5	30–70
		Sentinel DTS-XR-SM	± 0.05 (2σ)	1	20
		Foresight-BDTS	± 0.3 (2σ)	1	10
Moser et al. (2024)	Vibrating wire	Geokon–4700	± 0.5	N/A	N/A
	DFOS (Raman)	MMF SEN 2 XRM	± 0.02 –0.12	1–2	15–45
Fairbairn and Azenha (2019)	Thermocouple	Type K	± 1.5 –2	–	–
	Thermocouple	Type T	± 0.5 –1	–	–
	Resistance	PT100	± 0.2	–	–
	temperature detector				
Du et al. (2019)	DFOS	–	± 0.08	–	–
Van de Giesen et al. (2012)	DFOS (Raman)	–	± 0.1	1–2	5

a factor of 0.1 if no improvement was observed in validation loss for five consecutive epochs. The model was trained using a batch size of 800 samples and a maximum of 150 epochs. To mitigate overfitting, a dropout rate of 0.4 was applied to each hidden layer. In addition, early stopping was implemented, stopping the training process if no improvement in validation loss was detected after 10 epochs. This prevented overfitting by ensuring the model did not continue training once it had sufficiently learned from the data.

The resulting defect-labeled samples, both true and false positives, are then passed on to the regressor to evaluate each sample further.

2.3.3. Regression algorithm

The regression algorithm presents a similar structure in complexity compared to the classifier. The algorithm chosen for this task was “*KerasRegressor*.” This regressor was trained to predict the size and location (x and y coordinates) of defects within the pile cross-section, potential cage shifts, and back-analyze three concrete HoH parameters: β , τ , and Q_{max} .

The configuration of the model included three densely connected hidden layers with 1,500, 750, and 250 neurons, respectively, each using ReLU activations. Batch normalization was applied after each hidden layer to improve training stability, along with a dropout rate of 0.4. The regressor was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 400 samples over 150 epochs. Reduced learning rate and early stopping were also included. The model’s loss function was the mean squared error (MSE) function, as shown in Equation (9). The model was trained on the same dataset as the classifier and subsequently validated and tested on the true and false positive samples labeled by the classifier. This produced an environment that not only determines the presence of inclusions, but also characterizes their size and location.

$$MSE = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad (9)$$

3. Results and discussion

3.1. Classification ANN

A total of 9,942 samples were used to evaluate the performance of the classification ANN model. The confusion matrix for the test set, shown in Figure 7, demonstrates the excellent classification ability of the model: about 94% of the samples were classified correctly. The model only misidentified a limited number of pile samples with defects, mislabeling 571 of these as false negatives (i.e., the classifier indicates the absence of a defect when there was indeed one). The number of false positives is lower in comparison, mislabeling 77 samples (i.e., the classifier returns the existence of a defect when one was not present). The receiver operating characteristic (ROC) curve, shown in Figure 8, further illustrates the model’s effectiveness by depicting its true positive rate against the false positive rate at various threshold levels. The area under the curve (AUC) is 0.97, indicating a high level of classification accuracy (a value of 0.5 would indicate that the classification was as good as a random guess and 1.0 would suggest perfect classification).

True Label	No Defect	4949	77
	Defect	571	4345
		No Defect	Defect
		Predicted Label	

Figure 7. Confusion matrix for the test set.

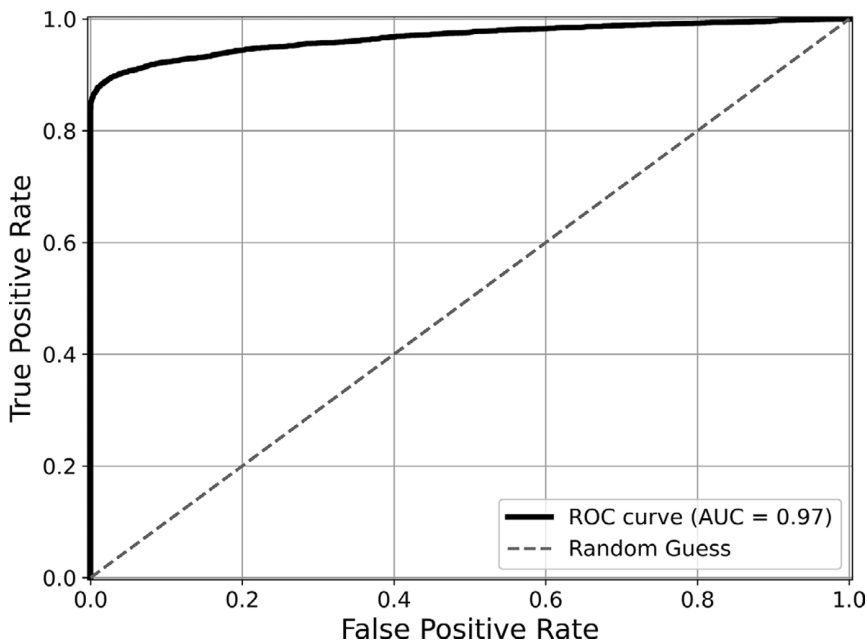


Figure 8. ROC curve for the test set.

Figure 9 presents the accuracy curve for the training and validation datasets. These curves serve as useful indicators of potential overfitting. Consistent with the above, the model shows very good predictive performance. The obtained training and validation accuracies are 0.938 and 0.935, respectively, suggesting that the model is effectively capturing patterns in the data without overfitting. This allows the model to generalize well to new, unseen data.

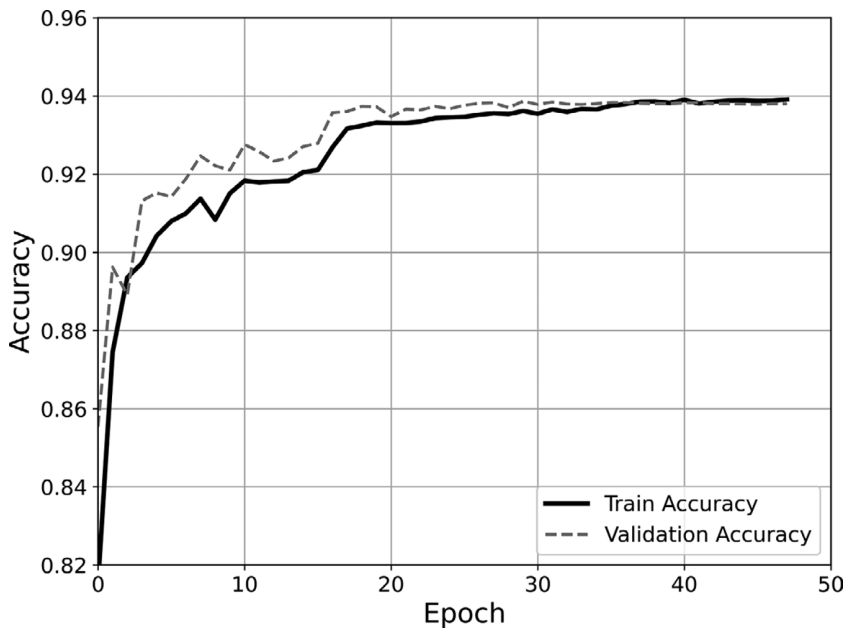


Figure 9. Train and validation accuracy for each epoch.

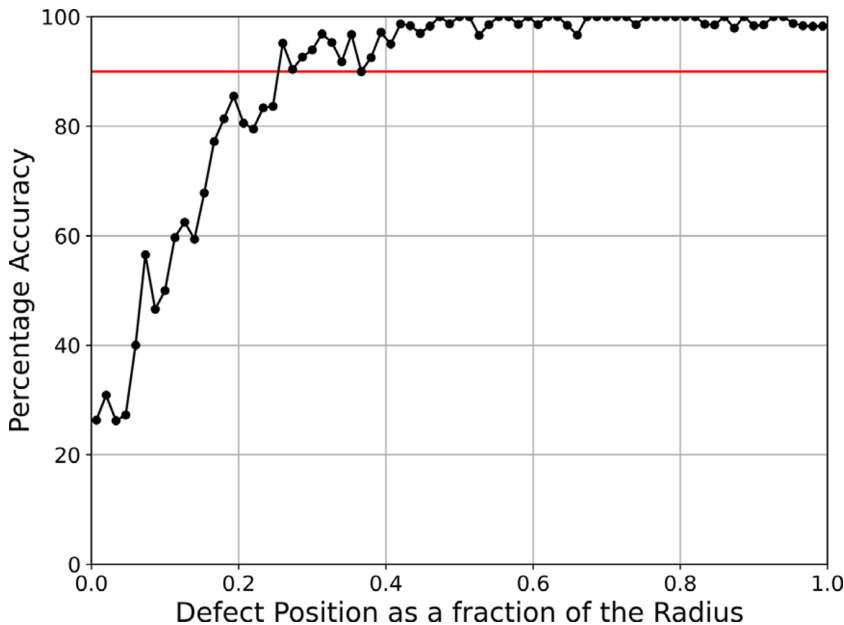


Figure 10. Accuracy versus defect radial position for defect-labeled samples.

An attempt was made to lower the classifier's decision threshold to reduce the number of false negatives. However, the model exhibited a high degree of confidence in its false negative predictions. Adjusting the decision threshold, therefore, provides no meaningful improvement in model performance, as the false positive rate increases by similar amounts as the false negative decreases.

Figure 10 shows the model's accuracy at predicting defect samples based on their radial distance from the center of the pile; for reference, samples have been grouped in bins of equal width, and the 90% accuracy mark is indicated by the solid red line. As expected, given the location of the probes, the model struggles to identify some of the defects located within the central 20% of the pile's radius. In contrast, for defects situated in the outer 80% of the pile's radius (or 96% of the equivalent pile area), the model's accuracy is nearly perfect. This trend aligns with the findings of Boeckmann and Loehr (2019), who observed that TIP techniques tend to underperform when attempting to detect inclusions within the interior of the reinforcement cage, that is, at locations distant from measurement optic cables.

When evaluating the model's performance across defects categorized by relative and nominal area, as shown in Figures 11 and 12, defect size is shown to play an important role. In effect, the model demonstrates higher accuracy for larger defects, consistent with the expectation that larger anomalies generate more distinguishable disturbances in the temperature field, making it easier for the model to classify correctly.

Smaller defects are likely to produce only subtler disturbances in the temperature distribution, which, added to the additional noise included artificially, lead to a higher rate of misclassification. For a target accuracy of 0.9, the existence of defects with areas $>0.17 \text{ m}^2$ (0.23 m radius) will be correctly predicted. Even the smallest defects, with an area of $\sim 0.05 \text{ m}^2$ (0.12 m radius), are detected with over 70% accuracy.

Ashlock and Fotouhi (2014) conclude that a lower bound of at least 8% in relative area is required for TIP to predict inclusions accurately, as these produce barely detectable temperature differences. In this respect, TIP performance is proven to be similar to that of other common pile integrity testing methods such as CSL. Li et al. (2005) report that the minimum detectable defect size ranges between 2 and 7% of the pile cross-sectional area, with defects with radii smaller than 0.1 m being undetectable. According to that study, achieving a target accuracy of 0.9 requires defect sizes ranging from 7.5 to 15% of the pile area, which is similar to the results of the ML model presented herein. Amir and Amir (2009) performed a CSL

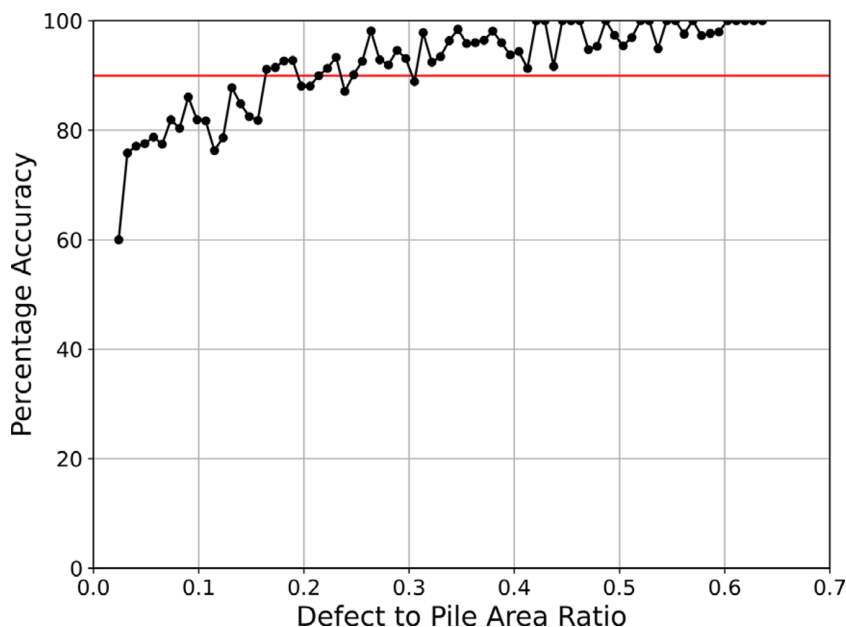


Figure 11. Accuracy versus defect area–pile ratio for defect-labeled samples.

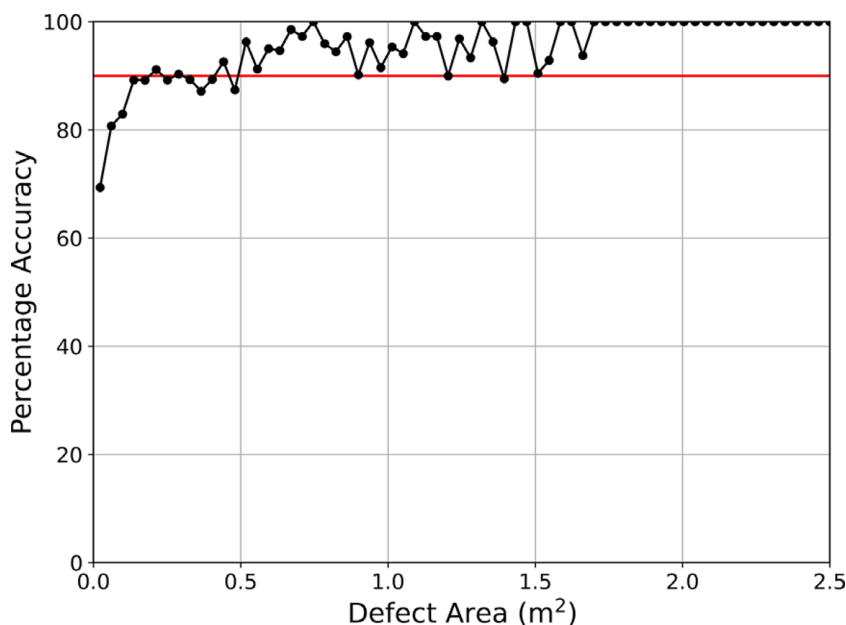


Figure 12. Accuracy versus defect area for defect-labeled samples.

test on a pile with three artificial defects and concluded that flaws are only detectable between access tubes if they are >10% of the pile cross-sectional area.

Clearly, despite the results suggesting that the classification model has a lower prediction accuracy for defects located toward the pile center and of reduced size, a very high accuracy can be achieved for all types of soil, concrete, and defect characteristics.

The previous results show that most cases where the model misclassifies a defective pile share one condition: the defects are located at, or close to, the pile center. By excluding all samples located within the

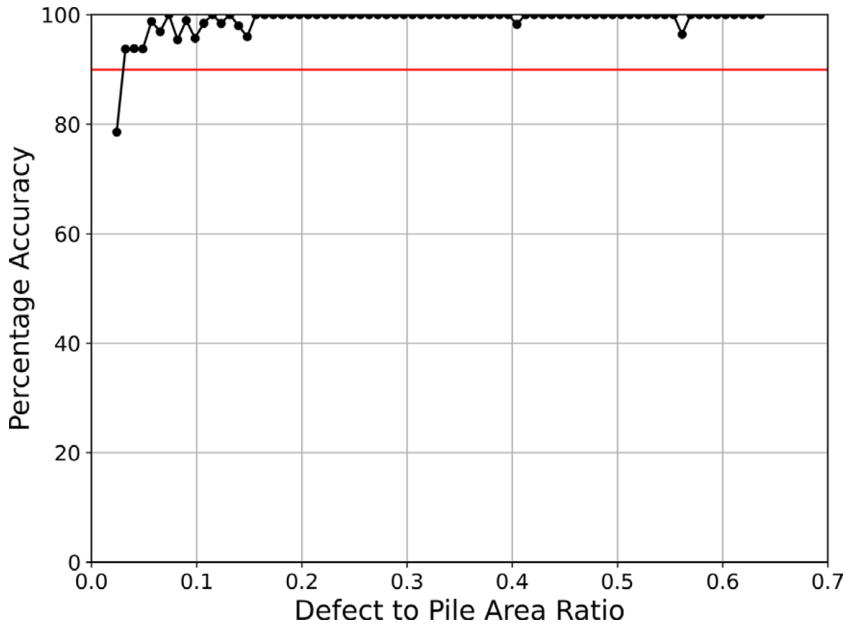


Figure 13. Accuracy versus defect area–pile ratio for samples in the outer 80% of the radial distance.

central section defined by 20% of the pile radius, near-perfect accuracy is achieved, as depicted in Figure 13. This further confirms that the primary limitation of TIP in defect classification is the proximity to the nearest probe rather than defect size. TIP's overall accuracy would significantly increase should a central optic cable be installed in the pile being monitored. To further corroborate this statement, an additional study was carried out where a ninth fiber optic cable is included at the pile center, whose results are shown in Section 3.3.

3.2. Regression ANN

The regression model was tested on 4,422 positive-labeled samples, 77 of which were false positives returned by the classifier. The performance metrics used to evaluate the results are the root MSE (RMSE), the mean absolute error (MAE), and the coefficient of determination (R^2), defined in Equations (10)–(12), respectively. In these expressions, x_i represents the true value, $\hat{x}_i$ is the predicted value, $SS_{residual}$ is the sum of squared residuals, and SS_{total} is the total sum of squares. Clearly, given their definitions, better predictions are characterized by values of RMSE and MAE closer to 0 and R^2 values closer to 1. The errors and performance results for all predicted output variables are provided in Table 3.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2} \quad (10)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - \hat{x}_i| \quad (11)$$

$$R^2 = 1 - \frac{SS_{residual}}{SS_{total}} \quad (12)$$

Figure 14 shows the real versus predicted values for the output variables. Predictions are particularly accurate for the defect position and area, arguably the most important parameters to identify. The cage

Table 3. Regressor performance metrics for the test dataset

Variable	RMSE	MAE	R^2
β [—]	0.043	0.033	0.955
τ [s]	10,099	7,336	0.747
$Q_{\max}$ [kJ · kg ⁻¹]	46.07	35.98	0.719
Cage misalignment (x) [m]	0.010	0.007	0.858
Cage misalignment (y) [m]	0.010	0.007	0.858
Defect position (x) [m]	0.057	0.036	0.963
Defect position (y) [m]	0.057	0.037	0.964
Defect area [m ²]	0.128	0.079	0.923

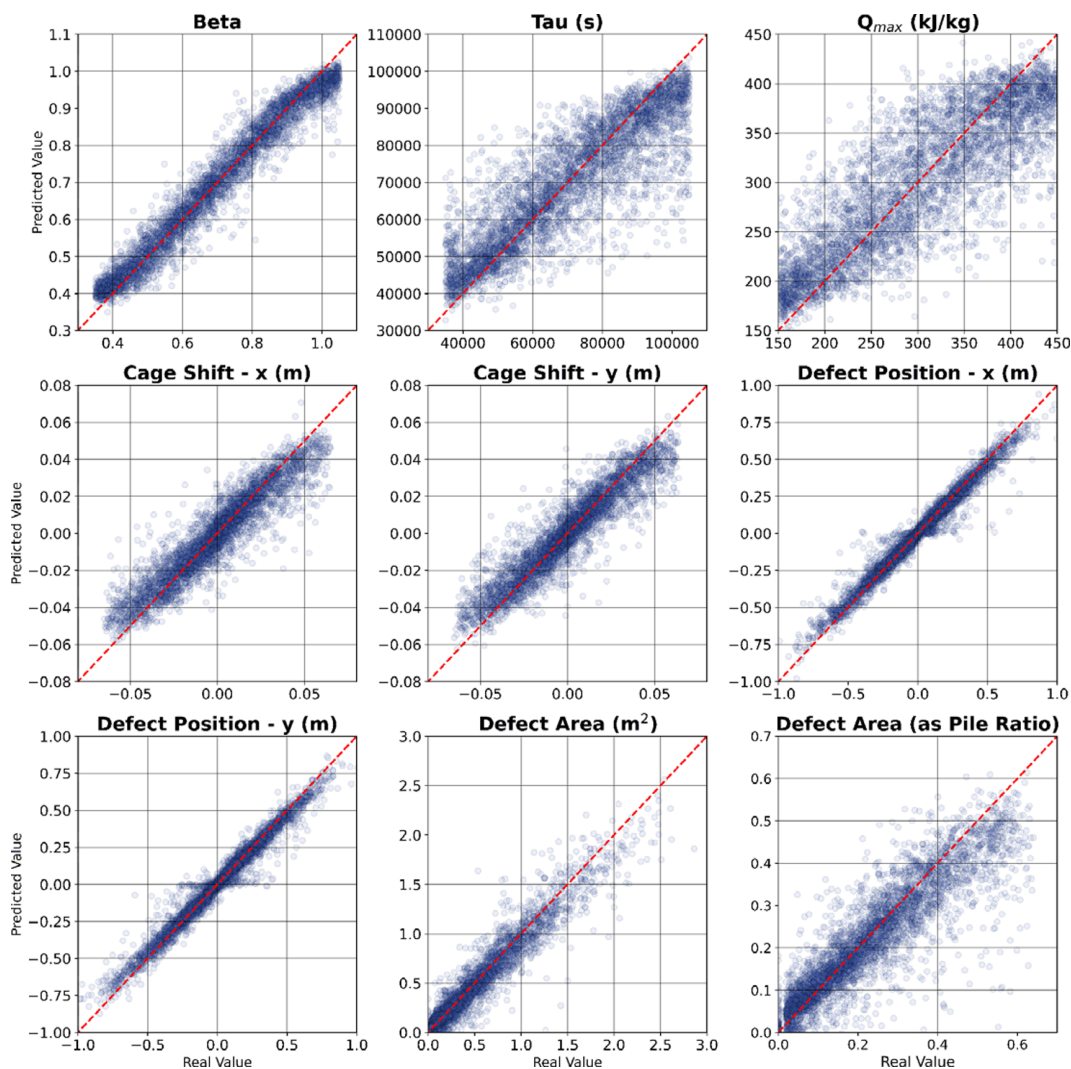


Figure 14. Correlation between real and predicted variables for 4,422 samples.

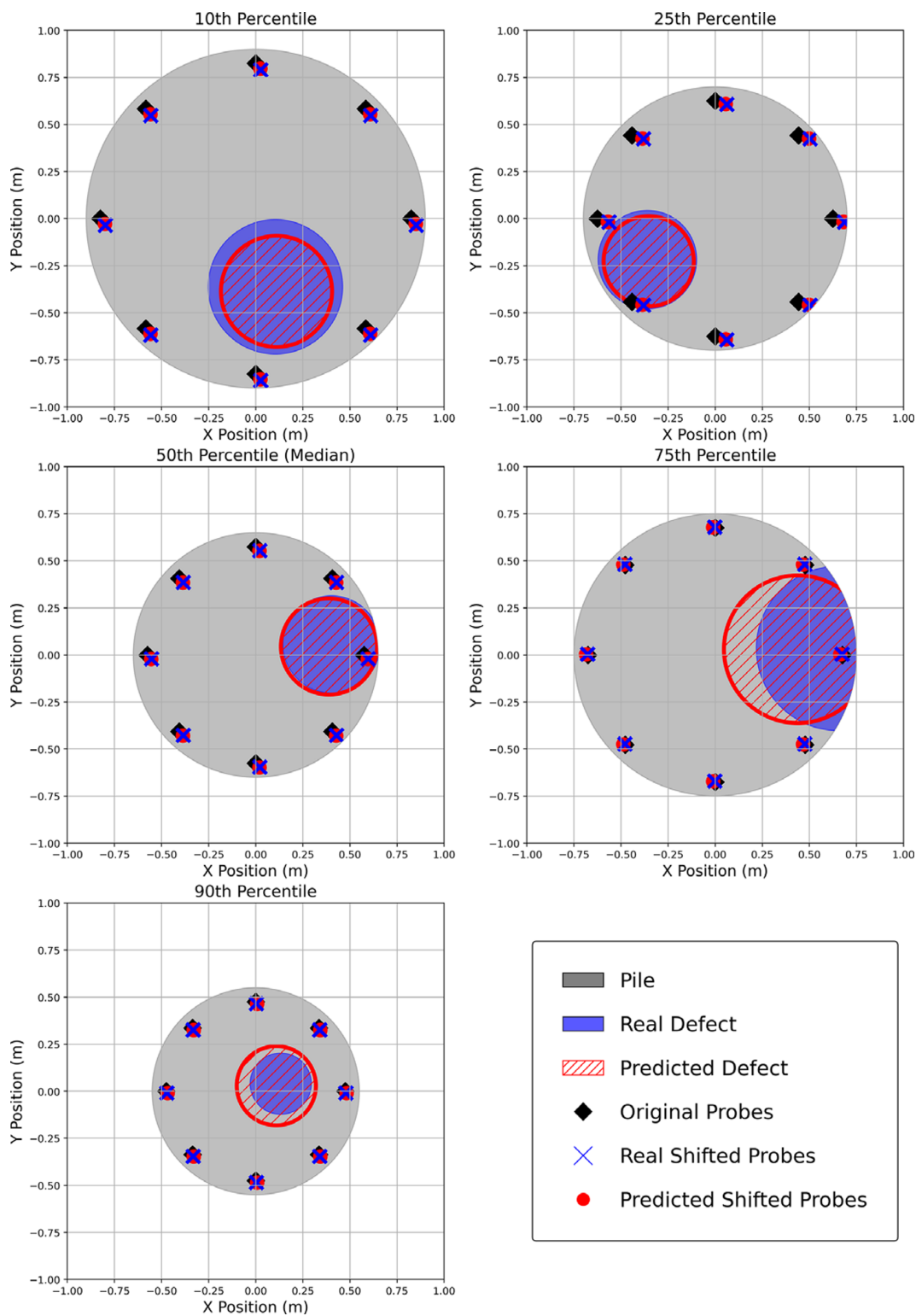


Figure 15. Graphical representation of predicted and real defects and cage displacements for five representative samples.

misalignment was identified within 7-mm precision, and the defect position within 4 mm, as given by the MAE. The back-calculation of the concrete HoH parameters presents lower accuracies, particularly the parameter controlling the heat evolution over time, τ , and the maximum heat generated per unit of mass, Q_{max} . However, while this information can be useful to predict the hydration response of other structures using the same concrete mix, these values are naturally less important as an outcome of a TIP analysis, where the focus is necessarily on the prediction of the characteristics of any existing defect.

Overall, the model captures general trends with very high accuracy, despite the artificial noise added to the data and the broad range considered for all parameters. Figure 15 shows the graphical comparison of the real and predicted defects, along with the rebar cage shift, for five samples representing the 10th, 25th, 50th, 75th, and 90th percentiles in percentage error performance. Samples in the lower percentiles predict defects smaller than the real value. The opposite is true for higher percentiles, with the 50th percentile representing an accurately predicted sample. To assess the impact of these prediction inaccuracies on the thermal profile of the samples, the temperature evolution was recalculated in the FE model using the new predicted values for predicted defect size, location, and reinforcement cage shift. These are shown in Table 4. Despite slight inaccuracies in defect characterization, temperature discrepancies are negligible, staying below 1°C for all samples and temperature statistical variables.

Figure 16 illustrates the real and predicted defect to pile area ratios, with each sample represented by a dot color-coded according to its distance from the pile center. Consistent with the findings in Section 3.1, the size of a significant number of central defects is substantially underestimated in size. This is because the temperature at the sensors does not clearly reflect the presence of the defects, and therefore, the ANN could not learn the patterns as well as for defects closer to the sensors. In some cases, these central defects are difficult to predict by the regressor, as the model likely interprets the overall lack of heat as resulting from factors like a reduced Q_{max} or elevated thermal conductivity, rather than from a significant absence of concrete mass. Consequently, the model predicts near-zero defect sizes for these samples, even though some of them are large. Conversely, a few smaller defects closer to the pile edge are slightly overestimated in size.

Figure 17 shows the model loss function (MSE), over the training process of 150 epochs. The similarity between the error obtained for the training and validation sets is indicative of the good generalization ability of the model to unseen data.

3.3. Additional central probe

To address the challenge of mislabeled samples concentrated near the center of the concrete pile, training both the classifier and regressor models by incorporating data from a central fiber optic probe was explored. Integrating a central probe within standard rebar cages can be challenging due to installation complexity,

Table 4. Comparison of hydration temperature values caused by real and predicted defects and cage displacements for five samples selected based on the associated error

Sample	Avg temp real (°C)	Avg temp pred. (°C)	Max temp real (°C)	Max temp pred. (°C)	Time at max temp real (d)	Time at max temp pred. (d)
10th percentile	19.18	18.21 (−5.1%)	29.16	28.73 (−1.5%)	1.12	1.16 (3.6%)
25th percentile	14.04	13.62 (−3.0%)	27.31	26.72 (−2.2%)	1.26	1.27 (0.8%)
50th percentile (near-zero error)	7.47	7.46 (−0.1%)	11.5	11.58 (0.7%)	2.28	2.22 (−2.6%)
75th percentile	6.53	6.77 (3.7%)	9.86	10.31 (4.6%)	2.26	2.4 (6.2%)
90th percentile	3.69	4.01 (8.7%)	5.69	6.32 (11.1%)	0.93	0.92 (−1.1%)

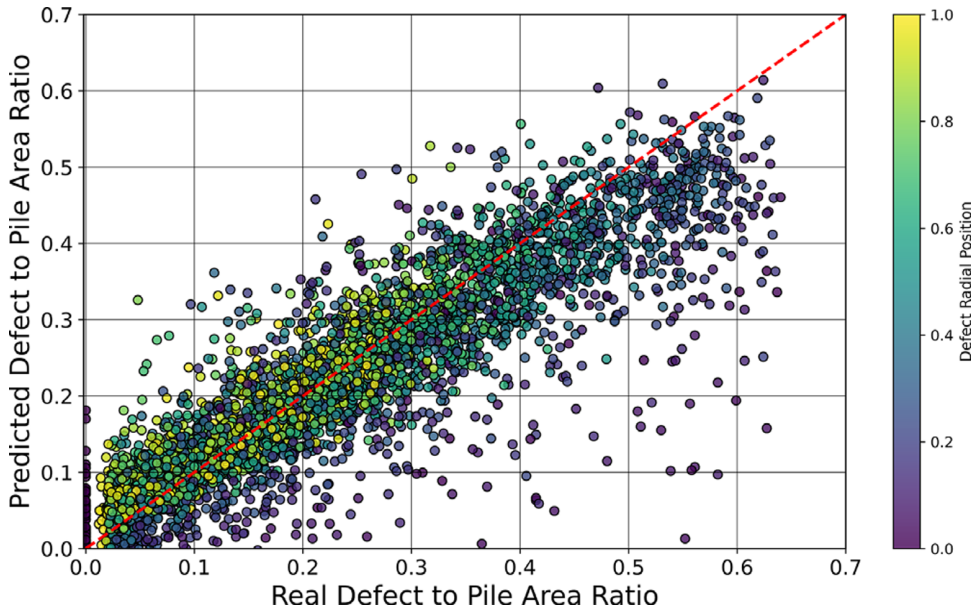


Figure 16. Correlation between real and predicted defect-to-pile area ratio, with samples colored based on radial distance to pile center.

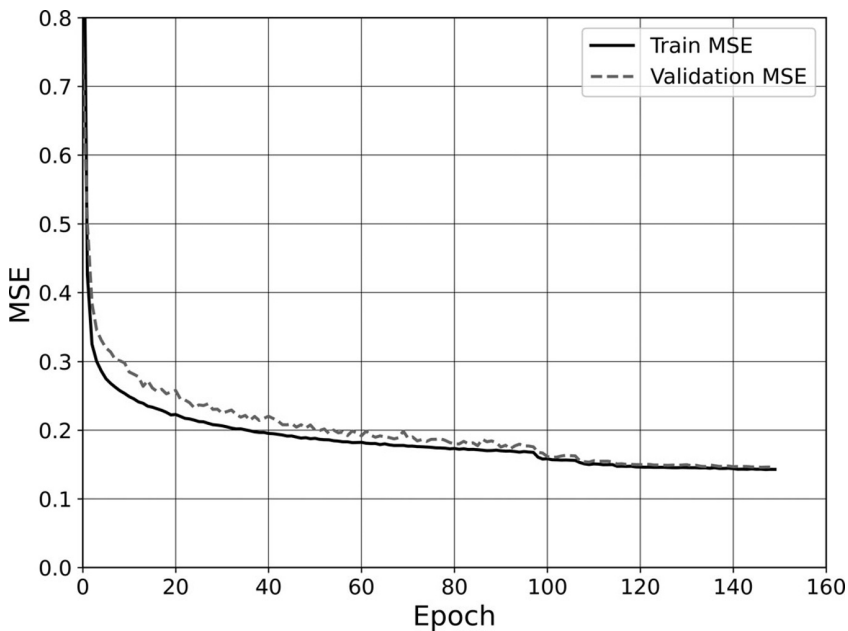


Figure 17. Evolution of training and validation loss with epochs.

making its feasibility dependent on project-specific considerations and contractor discretion. This section highlights the improvements in terms of predictive performance that can be expected by adding a central probe in TIP experiments, thus informing potential future adaptations to this methodology.

The classifier and regressor models were configured as described in Sections 2.3.2 and 2.3.3. With the inclusion of central fiber optic data, the test accuracy of the classification model increased, yielding an R^2 value of 0.974 (up from 0.935). Figures 18 and 19 show the confusion matrix and the training and validation loss curves during the training process, respectively. The false positive rate decreased by 38% compared to the original model, where only radial probes were used, while the false negative rate saw a reduction of 63%. The training and validation R^2 scores reached 0.980 and 0.978, respectively, indicating that overfitting is not expected to have taken place. Furthermore, the area under the ROC curve (AUC) increased from 0.97 to 0.99 by adding the central probe. This shows how adding the additional central probe can improve the model's ability to detect accurately the presence or absence of defects.

The results in terms of accuracy versus defect area are presented in Figure 20. The target accuracy is now reached for samples with defect areas larger than 0.07 m^2 , which is considerably smaller than the 0.17 m^2 required by the original model. Indeed, defects with an area exceeding 0.14 m^2 were detected with

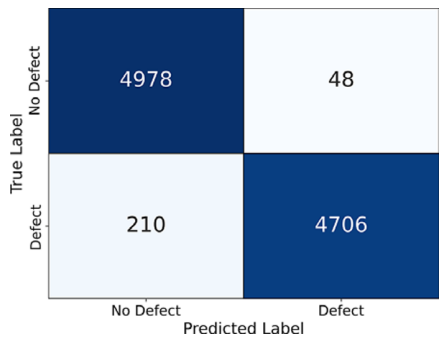


Figure 18. Confusion matrix for the nine-probe test.

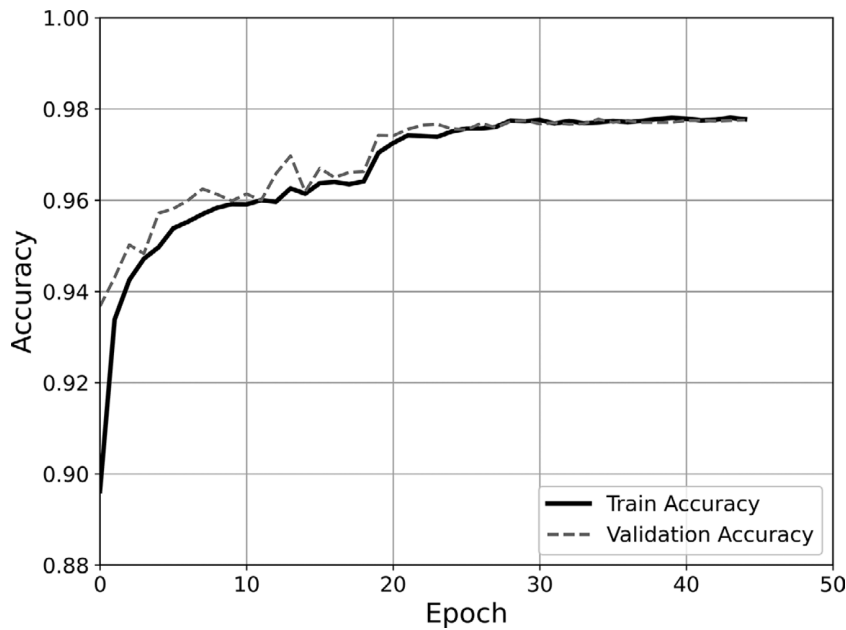


Figure 19. Train and validation accuracy for each epoch, for the nine-probe test.

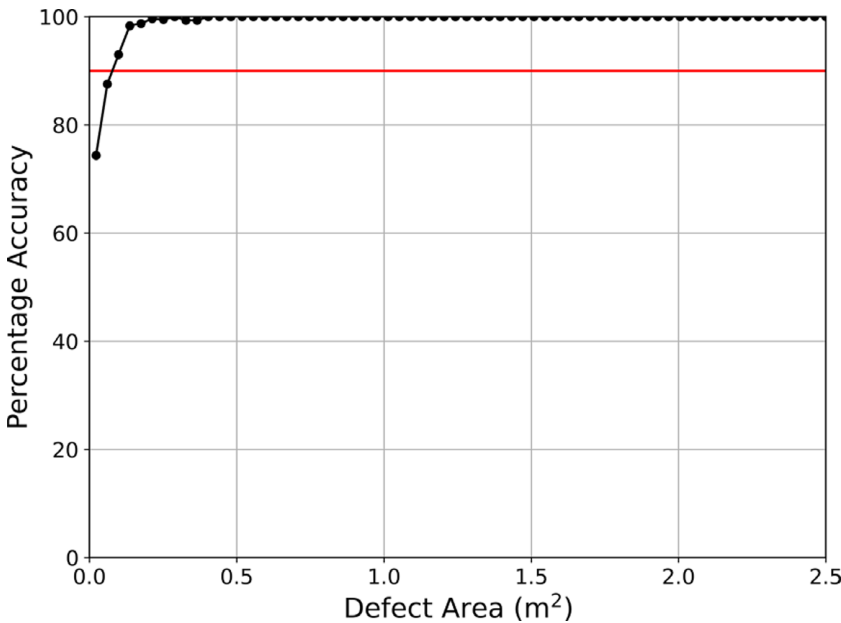


Figure 20. Accuracy versus defect area for the nine-probe results.

Table 5. Performance metrics and improvement (%) for regression model with nine probes

Variable	RMSE	MAE	R^2
β [—]	0.037 (−14.1%)	0.028 (−15.7%)	0.967 (+1.24%)
τ [s]	9,223 (−8.67%)	6,359 (−13.3%)	0.789 (+5.63%)
$Q_{\max}$ [kJ/kg]	40.95 (−11.1%)	31.73 (−11.8%)	0.779 (+8.16%)
Cage misalignment (x) [m]	0.008 (−23.0%)	0.005 (−29.0%)	0.916 (+6.76%)
Cage misalignment (y) [m]	0.008 (−24.2%)	0.005 (−28.0%)	0.919 (+7.1%)
Defect position (x) [m]	0.044 (−23.2%)	0.026 (−29.1%)	0.977 (+1.41%)
Defect position (y) [m]	0.042 (−26.1%)	0.026 (−30.0%)	0.979 (+1.53%)
Defect area [m ²]	0.085 (−34.1%)	0.052 (−34.4%)	0.965 (+4.63%)

near-perfect accuracy. When considering defect to pile area ratios, the minimum defect sizes are found to be 3 and 6% for target accuracies of 0.8 and 0.9, respectively. The results agree with the conclusions presented by Sun et al. (2024), who identified the minimum detectable defect area to be 4–5% of the pile cross-sectional area when using eight radial and one central fiber optic cables.

The regression results are tested on 4,754 samples. Table 5 shows the performance metrics (MAE, RMSE, and R^2) for the updated regressor, together with percentage improvements with respect to the results discussed in Section 3.2. It can be seen that all performance metrics improve when an additional central probe is included. Among all the output variables, the defect area and position show the greatest improvement in predictive accuracy, with 34 and 29% improvement is the MAE, respectively. Figure 21 shows the real versus predicted variable plots for all output variables. The improvement is obvious with respect to Figure 14, as all variables plot in a narrower band around the 1:1 dashed red line.

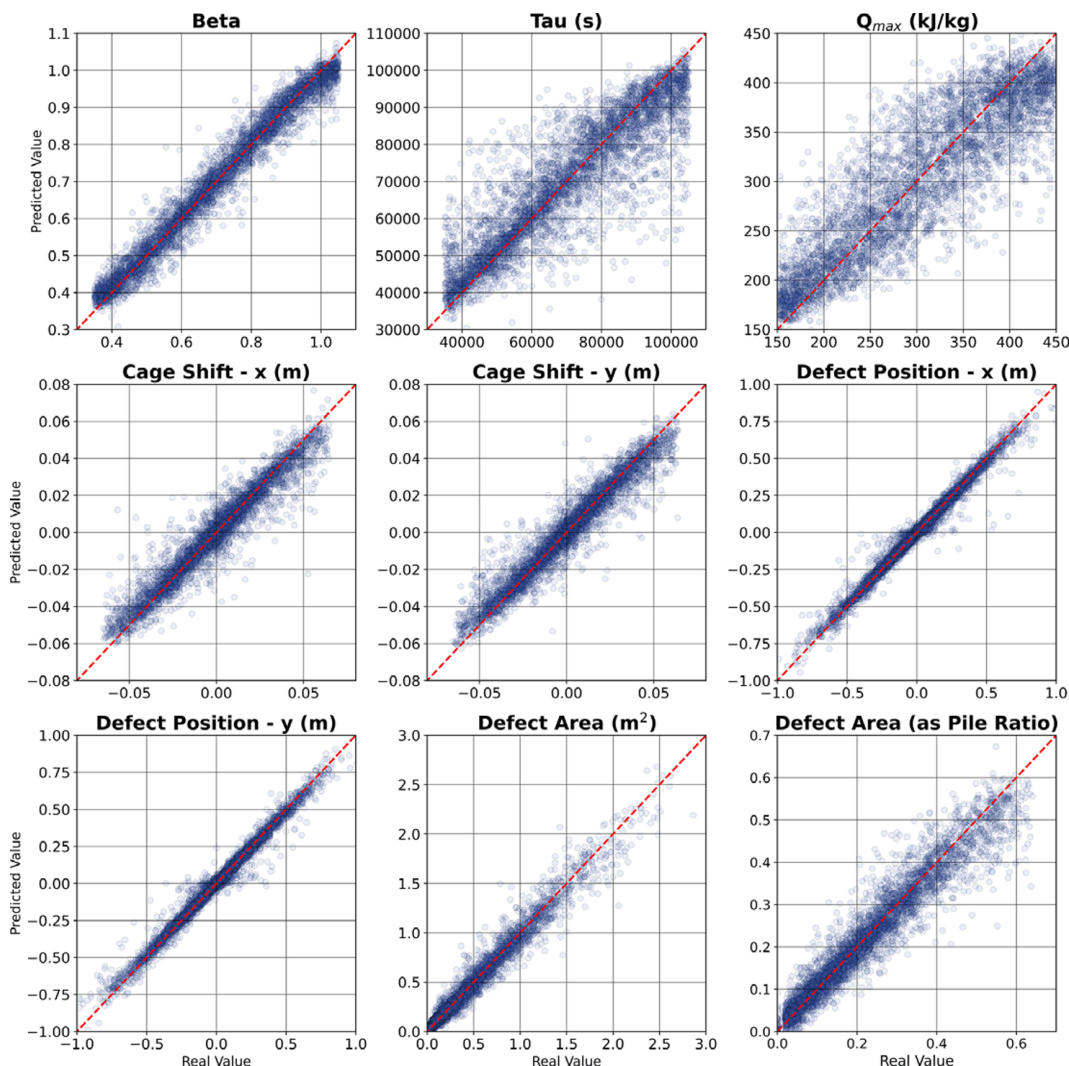


Figure 21. Correlation between real and predicted variables for the nine-probe model.

4. Conclusions

This study presents a comprehensive ML classification and regression model developed as a proof of concept to detect soil or bentonite inclusions in drilled foundations based on TIP data. The model is trained based on a conservative scenario, where a large range of variation in material properties is considered. The model's robustness was improved by adding random noise to all temperature readings, thereby mitigating the inherent determinism of numerical FE data. The following conclusions can be drawn:

- The classifier can label correctly 94% of the processed samples. Thermal variations in the central section of the pile are challenging to capture due to the probes' positioning near the edge of the pile, limiting detection capability in these regions. The overall accuracy of the model is very good ($R^2 = 0.935$).
- The false positive rate is significantly lower than the false negative rate. False negatives are primarily associated with defects of varied sizes located within the central 25% of the pile radius. These defects remain undetected due to the considerable distance to the nearest optical fiber probes.
- Through regression, very high accuracies can be achieved when predicting defect size, location, and cage misalignment. Hydration parameters can also be back-calculated with reasonable precision.

- Including a central probe can significantly improve the ability to predict defects located close to the pile center, particularly through the reduction of false negatives in the classification stage. The inclusion of an additional probe leads to improvements across all labels, with reductions in MAE for defect location, size, and cage misalignment ranging between 28 and 35%. Therefore, it is recommended to include an additional probe if conditions allow.

This study focused on demonstrating the potential of ML techniques in TIP using data generated through 2D modeling. For greater applicability to complex structures, the model can be expanded to simulate the pile behavior in 3D, potentially covering additional defect types, such as necking, uneven radii, or internal steps. Currently, each sample with defects includes only a single circular defect. Incorporating multiple defects and/or varying their shape in future simulations may provide a more accurate representation of real-world construction scenarios.

List of abbreviations

ANN	artificial neural network
AUC	area under the curve
FE	finite element
HoH	heat of hydration
MAE	mean absolute error
ML	machine learning
RMSE	root mean square error
TIP	thermal integrity profiling

Data availability statement. The data that support the findings of this study are available in Zenodo at <http://doi.org/10.5281/zenodo.14674381>.

Author contribution. Javier Sánchez Fernández: Data curation (lead), formal analysis (equal), investigation (lead), methodology (lead), writing—original draft (lead), writing—review and editing (supporting).
Agustin Ruiz López: Conceptualization (equal), methodology (Supporting), supervision (equal), writing—review and editing (equal).
David M.G. Taborda: Conceptualization (equal), methodology (supporting), project administration (lead), supervision (equal), writing—review and editing (equal).

Funding statement. This work was supported by the Natural Environment Research Council [NE/S007415/1] and by the Engineering and Physical Sciences Research Council (EPSRC) through the project *intigeo* [EP/R511547/1].

Competing interests. The authors declare none.

References

- Amir JM and Amir EI (2009, 10 March) Capabilities and limitations of cross hole ultrasonic testing of piles. In Iskander, M., Laefer, D.F. and Hussein, M.H. (eds.) *Contemporary Topics in Deep Foundations*. Reston, VA: American Society of Civil Engineers, pp. 536–543. [http://doi.org/10.1061/41021\(335\)67](http://doi.org/10.1061/41021(335)67).
- Asadi I, Shafigh P, Abu Hassan ZFB and Mahyuddin NB (2018) Thermal conductivity of concrete – A review. *Journal of Building Engineering* 20(July), 81–93. <http://doi.org/10.1016/j.jobe.2018.07.002>.
- Ashlock JC and Fotouhi MK (2014) Thermal integrity profiling and crosshole sonic logging of drilled shafts with artificial defects. In *Geo-Congress 2014: Geo-Characterization and Modeling for Sustainability*, pp. 1795–1805. <http://doi.org/10.1061/9780784413272.176>.
- ASTM (2016a) D5882: Standard Test Method for Low Strain Impact Integrity Testing of Deep Foundations. Annual Book of ASTM Standards. American Society for Testing and Materials. <http://doi.org/10.1520/D5882-16>.
- ASTM (2016b) D6760: Standard Test Method for Integrity Testing of Concrete Deep Foundations by Ultrasonic Crosshole Testing. American Society for Testing and Materials. <http://doi.org/10.1520/D6760-16.1>.
- ASTM (2014) D7949: Test Methods for Thermal Integrity Profiling of Concrete Deep Foundations. ASTM International. <http://doi.org/10.1520/D7949>.
- Bao X and Chen L (2012) Recent Progress in distributed Fiber optic sensors. *Sensors* 12(7), 8601–8639. <http://doi.org/10.3390/s120708601>.
- Boeckmann AZ and Loehr JE (2019) Evaluation of thermal integrity profiling and Crosshole sonic logging for drilled shafts with concrete defects. *Transportation Research Record: Journal of the Transportation Research Board*. 2673(8), 86–98. <http://doi.org/10.1177/0361198119842112>.

- van Breugel K** (1997) *Simulation of hydration and formation of structure in hardening of cement-based materials*. 2nd ed. PhD Thesis. Delft: TU Delft, Department of Civil Engineering and Geosciences.
- Briaud J-L, Ballouz M and Nasr G** (2002) Defect and length predictions by NDT methods for nine bored piles. In: O'Neill, M. & Townsend, F.C. (eds.) *Deep Foundations 2002*. Reston, VA: American Society of Civil Engineers, pp. 173–192. [https://doi.org/10.1061/40601\(256\)13](https://doi.org/10.1061/40601(256)13).
- Chen K, Zhou X, Peng W and Yu Q** (2015) OFDR based distributed temperature sensor using the three-channel simultaneous radio-frequency lock-in technique. *Photonic Sensors* 5(3), 217–223. <http://doi.org/10.1007/s13320-015-0246-z>.
- Chollet F** (2021) *Deep Learning with Python*. New York: Manning Publications
- Du Y, Sun B, Li J and Zhang W** (2019) *Optical fiber sensing and structural health monitoring technology*. pp. 31–74. 1st edition. Singapore: Springer Singapore. doi:10.1007/978-981-13-2865-7
- Fairbairn EMR and Azenha M** (2019) *Thermal Cracking of Massive Concrete Structures. RILEM State-of-the-Art Reports*. Cham: Springer International Publishing. <http://doi.org/10.1007/978-3-319-76617-1>.
- Finno RJ, Chao H-C, Gassman SL and Zhou P** (2002, February 26) Non-destructive evaluation of drilled shafts at the Amherst NGES test section. In O'Neill, M. & Townsend, F.C. (eds.) *Deep Foundations 2002*. Reston, VA: American Society of Civil Engineers, pp. 883–897. [http://doi.org/10.1061/40601\(256\)62](http://doi.org/10.1061/40601(256)62).
- Forrester AIJ, Sôbester A and Keane AJ** (2008) *Engineering Design Via Surrogate Modelling*. Chichester, West Sussex: Wiley. <http://doi.org/10.1002/9780470770801>
- French S and Whittaker A** (2023) Pile integrity testing. In: Brown, M., Burland, J., Chapman, T., Higgins, K., Skinner, H. & Toll, D. (eds.) *ICE Manual of Geotechnical Engineering, Volume 2: Geotechnical Design, Construction and Verification*, pp. 1561–1597. <http://doi.org/10.1680/icemge.66830.1561>.
- Van de Giesen N, Steele-Dunne SC, Jansen J, Hoes O, Hausner MB, Tyler S and Selker J** (2012) Double-ended calibration of fiber-optic Raman spectra distributed temperature sensing data. *Sensors* 12(5), 5471–5485. <http://doi.org/10.3390/s120505471>.
- Di Girolamo L, Ausiello G, Russo G and Marone G** (2022) High thermal conductivity concrete for energy piles. *Acta Polytechnica CTU Proceedings* 33, 125–132. <http://doi.org/10.14311/APP.2022.33.0125>.
- Grirate H, Agalit H, Zari N, Elmchaouri A, Molina S and Couturier R** (2016) Experimental and numerical investigation of potential filler materials for thermal oil thermocline storage. *Solar Energy* 131, 260–274. <http://doi.org/10.1016/j.solener.2016.02.035>.
- Hartog AH** (2017) *An Introduction to Distributed Optical Fibre Sensors*. Boca Raton: CRC Press. <http://doi.org/10.1201/9781315119014>
- Johnson KR** (2016) Analyzing thermal integrity profiling data for drilled shaft evaluation. *DFI Journal* 10(1), 25–33. <http://doi.org/10.1080/19375247.2016.1169361>.
- Kishi T and Maekawa K** (1996) Complex hydration heat model for mixed cement using blast furnace slag and fly ash. *Proceedings of the Japan Society of Civil Engineers* 1996(550), 131–143. http://doi.org/10.2208/jsej.1996.550_131.
- Kodesová R, Vlasáková M, Fér M, Teplá D, Jakšík O, Neuberger P and Adamovsky R** (2013) Thermal properties of representative soils of the Czech Republic. *Soil and Water Research* 8(4), 141–150. <http://doi.org/10.17221/33/2013-swr>.
- Li DQ, Zhang LM and Tang WH** (2005) Reliability evaluation of cross-hole sonic logging for bored pile integrity. *Journal of Geotechnical and Geoenvironmental Engineering* 131(9), 1130–1138. [http://doi.org/10.1061/\(asce\)1090-0241\(2005\)131:9\(1130](http://doi.org/10.1061/(asce)1090-0241(2005)131:9(1130)
- Lim MK and Cao H** (2013) Combining multiple NDT methods to improve testing effectiveness. *Construction and Building Materials* 38, 1310–1315. <http://doi.org/10.1016/j.conbuildmat.2011.01.011>.
- Liu H, Wu W, Yang X, Jiang G, El Naggari MH, Mei G and Liang R** (2019) Detection sensitivity analysis of pipe pile defects during low-strain integrity testing. *Ocean Engineering* 194(June), 106627. <http://doi.org/10.1016/j.oceaneng.2019.106627>.
- Liu RYW, Fisher A, Taborda DMG and Bourne-Webb PJ** (2022) A practical heat of hydration model for concrete curing for geotechnical applications. *Geotechnical Research* 9(1), 23–31. <http://doi.org/10.1680/jgere.21.00054>.
- Liu Y, Ding H, Wang K, Xiao H and Li L** (2020) Thermal conduction characteristics of DTS when detecting the integrity of cast-in-place piles considering their environment. *Heat and Mass Transfer/Waerme- und Stoffuebertragung* 56(7), 2185–2202. <http://doi.org/10.1007/s00231-020-02848-3>.
- Lozovsky I, Sellountou-Rausche EA, Zhostkov R and Churkin A** (2024) Numerical study of flaw detectability in drilled shafts by cross-hole sonic logging. *Indian Geotechnical Journal* 54(6), 2303–2317. <http://doi.org/10.1007/s40098-023-00865-7>.
- Lu X, Tong F, Liu G and Guan J** (2021) Model for predicting the thermal conductivity of concrete. *International Journal of Thermophysics* 42(3), 1–20. <http://doi.org/10.1007/s10765-020-02786-6>.
- McKay MD, Beckman RJ and Conover WJ** (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 42(1), 55–61. <http://doi.org/10.1080/00401706.2000.10485979>.
- Moscarelli M** (2023) *Biostatistics with 'R': A Guide for Medical Doctors*. Cham: Springer International Publishing. <http://doi.org/10.1007/978-3-031-33073-5>
- Moser D, Martin-Candilejo A, Cueto-Felgueroso L and Santillán D** (2024) Use of fiber-optic sensors to monitor concrete dams: Recent breakthroughs and new opportunities. *Structure* 67, 106968. <http://doi.org/10.1016/j.jstruc.2024.106968>.
- Mullins AG and Kranc SC** (2004) Method for testing integrity of concrete shafts. US Patent. University of South Florida, USF Patents, 762. Available at: https://digitalcommons.usf.edu/usf_patents/762.
- Mullins G** (2010) Thermal integrity profiling of drilled shafts. *DFI Journal: The Journal of the Deep Foundations Institute* 4(2), 54–64. <http://doi.org/10.1179/dfi.2010.010>.

- Mullins G, Kranc S, Johnson K, Stokes M and Winters D (2007) Thermal Integrity Testing of Drilled Shafts. Available at http://www.fdot.gov/research/Completed_Proj/Summary_SMO/FDOT_BD544_20_rpt.pdf.
- Neville AM (1986) *Properties of Concrete*, 3rd edn. Harlow: Longman Scientific & Technical.
- Parrot L and Killoh D (1984) Prediction of cement hydration. *Proceedings of the British Ceramics Society 1984*, 41–53.
- Pradhan HS and Sahu PK (2015) Characterisation of Raman distributed temperature sensor using deconvolution algorithms. *IET Optoelectronics* 9(2), 101–107. <http://doi.org/10.1049/iet-opt.2014.0048>.
- Raschka S (2017) Chapter 6: Learning best practices for model evaluation and hyperparameter tuning. In *Python Machine Learning: Machine Learning and Deep Learning with Python, Scikit-Learn, and TensorFlow*, 2nd edn. Birmingham, UK: Packt Publishing.
- Sánchez Fernández J (2025) Thermal Integrity Profiling Modelled Data for ML Training. [Dataset]. Available at <http://doi.org/10.5281/zenodo.14674381>.
- Sarhan H, Tabsh SW, O'Neill M, Ata A and Ealy C (2002) Flexural behavior of drilled shafts with minor flaws. In: O'Neill, M. & Townsend, F.C. (eds.) *Deep Foundations 2002*. Reston, VA: American Society of Civil Engineers, pp. 1136–1150. [http://doi.org/10.1061/40601\(256\)81](http://doi.org/10.1061/40601(256)81).
- Schindler A and Folliard K (2005) Heat of hydration models for cementitious materials. *ACI Materials Journal* 102(1), 24–33. <http://doi.org/10.14359/14246>.
- Sensornet (2021) Sentinel DTS Sensors. Available at <https://www.sensornet.co.uk/sentinel-dts-sensors/> (accessed 28 October 2024).
- Soga K and Luo L (2018) Distributed fiber optics sensors for civil engineering infrastructure sensing. *Journal of Structural Integrity and Maintenance* 3(1), 1–21. <http://doi.org/10.1080/24705314.2018.1426138>.
- Sun Q, Elshafie M, Barker C, Fisher A, Schooling J and Rui Y (2021) Thermal integrity testing of cast in situ piles: An alternative interpretation approach. *Structural Health Monitoring* 20(5), 2493–2512. <http://doi.org/10.1177/1475921720960042>.
- Sun Q, Elshafie MZEB, Xu X and Schooling J (2024) Pile defect assessment using distributed temperature sensing: Fundamental questions examined. *Structural Health Monitoring* 23(3), 1701–1724. <http://doi.org/10.1177/14759217231189426>.
- Swaddiwudhipong S, Chen D and Zhang MH (2002) Simulation of the exothermic hydration process of Portland cement. *Advances in Cement Research* 14(2), 61–69. <http://doi.org/10.1680/ader.2002.14.2.61>.
- Tan Y and Tang K (2024) Evaluate the effect of coarse aggregates on cement hydration heat and concrete temperature modelling using isothermal calorimetry. *Heliyon* 10(19), e38322. <http://doi.org/10.1016/j.heliyon.2024.e38322>.
- Tatro SB (2006) Chapter 22: Thermal properties. In Pielert JH (ed.), *Significance of Tests and Properties of Concrete and Concrete-Making Materials*. West Conshohocken, PA: ASTM International, pp. 226–237. <http://doi.org/10.1520/STP37740S>
- Teixeira De Freitas JA, Cuong PT, Faria R and Azenha M (2013) Modelling of cement hydration in concrete structures with hybrid finite elements. *Finite Elements in Analysis and Design* 77, 16–30. <http://doi.org/10.1016/j.finel.2013.07.008>.
- Veder C (1953) Method for the construction of impermeable diaphragms at great depth by means of thixotropic muds. In *BT: Proceedings of the 3rd International Conference on Soil Mechanics and Foundation Engineering*, Zurich, Switzerland, pp. 91–94.
- Vogt C, Mottaghy D, Wolf A, Rath V, Pechinig R and Clauser C (2010) Reducing temperature uncertainties by stochastic geothermal reservoir modelling. *Geophysical Journal International* 181(1), 321–333. <http://doi.org/10.1111/j.1365-246X.2009.04498.x>.
- Wang C, Jin Z, Li J, Dong W, Chen R, Yang Y, Chen Y, Wang D and Pang B (2023) Experimental study on early shrinkage and later expansion of concrete under a simulated geothermal environment. *Journal of Building Engineering* 72(March), 106493. <http://doi.org/10.1016/j.jobbe.2023.106493>.
- Winters D (2014) Comparative study of thermal integrity profiling with other nondestructive integrity test methods for drilled shafts. In *Geo-Congress 2014 Technical Papers*. Reston, VA: American Society of Civil Engineers, pp. 1731–1742. <http://doi.org/10.1061/9780784413272.170>.
- Wu YH, Wu YF, Fan LW, Yu ZT and Khodadadi JM (2025) Thermal conductivity of soil: A review on the vast experimental data and predictive models. *International Journal of Thermal Sciences* 208(July 2024), 109486. <http://doi.org/10.1016/j.ijthermalsci.2024.109486>.
- Xiao HL, Liu YL, Fan M, Li LH and Lei WK (2017) Detecting soil inclusion inside piles in the laboratory using DTS method. *Journal of Testing and Evaluation* 45(2), 452–459. <http://doi.org/10.1520/JTE20150251>.
- Xie Y and Qian C (2023) A novel numerical method for predicting the hydration heat of concrete based on thermodynamic model and finite element analysis. *Materials and Design* 226, 111675. <http://doi.org/10.1016/j.matdes.2023.111675>.
- Zhai J, Wang Q, Xie X, Qin H, Zhu T, Jiang Y and Ding H (2023) A new method for 3D detection of defects in diaphragm walls during deep excavations using cross-hole sonic logging and ground-penetrating radar. *Journal of Performance of Constructed Facilities* 37(1), 1–11. [http://doi.org/10.1061/\(asce\)cf.1943-5509.0001776](http://doi.org/10.1061/(asce)cf.1943-5509.0001776).
- Zhang T, Cai G, Liu S and Puppala AJ (2017) Investigation on thermal characteristics and prediction models of soils. *International Journal of Heat and Mass Transfer* 106, 1074–1086. <http://doi.org/10.1016/j.ijheatmasstransfer.2016.10.084>.
- Zhong R, Guo R and Deng W (2018) Optical-Fiber-based smart concrete thermal integrity profiling: An example of concrete shaft. E. Miroshnichenko (ed.), *Advances in Materials Science and Engineering 2018* (1), 1–8. <http://doi.org/10.1155/2018/9290306>.

Cite this article: Sánchez Fernández J, Ruiz López A and Taborda DMG (2025). A novel machine learning-based approach to thermal integrity profiling of concrete pile foundations. *Data-Centric Engineering*, 6, e33. doi:10.1017/dce.2025.10013