8th GRACM International Congress on Computational Mechanics
Volos, 12 July – 15 July 2015

SEMI SUPERVISED IDENTIFICATION OF NUMERICALLY SIMULATED PILE DEFECTS USING GRAPH LABEL PROPAGATION

Eftychios Protopapadakis[1], Marco Schauer[2], Anastasios Doulamis[3], Georgios Stavroulakis[1], Jens Böhrnsen[2], Sabine Langer[2]

[1]School of Production Engineering and Management
Technical University of Crete,
Kounoupidiana, Chania, 73100, Greece
e-mail: eprotopapadakis@isc.tuc.gr, gestavr@dpem.tuc.gr

[2] Institut für Konstruktionstechnik
Technische Universität Braunschweig
Braunschweig , 38106, Germany
e-mail: {m.schauer, j-u.boehrnsen, s.langer}@tu-braunschweig.de

[3] School of Rural and Surveying Engineering
National Technical University of Athens
Iroon Polytexneiou 9, Zografou, 15780, Greece
e-mail: adoulam@cs.ntua.gr

Keywords: Pile integrity test, Scaled boundary finite element method, Semi supervised learning.

Abstract. *A graph-based detector is utilized for the identification of structural defects in piles. The approach utilize semi supervised learning assumptions, which allow the exploitation of the unlabeled data to produce accurate defect identifications. A coupled finite element and scaled boundary finite element method approach is used to simulate the pile and its surrounding soil. The proposed defect detection system provides information about the type and the placement of the defect(s), given the surface's oscillation patterns.*

1 INTRODUCTION

Structures foundation in the form of concrete piles is a commonly adopted approach in many cases. These piles are usually built by using precast and cast-in-situ techniques. Sometimes, "necks" or "bulbs" may be created in the process of drilling. These defects may affect the bearing capacity of the piles. Hence the structural evaluation and monitoring of new and existing piles are becoming increasingly important. In this paper we deal with the detection of such defects using a graph based approach. Piles' surfaces' oscillations, produced through numerical simulation, serve as row data for the detection mechanism.

Surface oscillation is non-destructive testing (NDT) approach [1], adopted for many practical reasons. The test is based on wave propagation theory; the impact generates a compression wave that travels down the pile at a constant wave speed. Changes in cross sectional area (e.g. reduction in diameter) produce wave reflections. Engineers would desire the use of an intelligent software tool, able to automatically analyze these complex waveforms generated as a result of a pile integrity test (PIT) testing and produce classification outputs, regarding piles' condition states.

Towards that direction, an innovative work for the NDT of piles is employed using a mixture of state of the art soft computing techniques, appropriate feature extraction and data generation procedures. Regarding the classification process, a graph based label propagation approach is adopted over a graph, which is constructed under semi supervised learning (SSL) assumptions [2]. The innovation of the current methodology is the fully automatic post processing technique, which results in high classification performance, easy implementation and noise tolerance, using a limited training sample.

The results have been obtained on experimental data originating from numerical experiments. These data, as described below, simulate as much as possible real-life phenomena of "neck" or "bulb" type structural defects. Application of novel intelligent classification algorithms for defects' prediction should be first experientially validated and tested under laboratory conditions to guarantee the successful performance of the classifier and then to be validated on real-data, which requires huge financial effort while it is also risky in such infrastructures.

1.1 Related work

Surface reflection techniques are a common approach in the foundation assessment field. The work of [3] investigates the relative performance of the sonic echo (SE), impulse response (IR), and parallel seismic (PS) tests using a field constructed pile foundation incorporating simulated defects. The work of [4] presents the state of the art approaches of NDT on building structures. Such studies suggest the NDT approach appropriateness.

Numerical simulation of NDT cases [5] is an alternative approach, which allow the investigation of various defect types, providing accurate results very close to real life situations. The work of [6] focuses on drilled shaft defects identification. Other approaches focuses on the testing conditions, such as the effects of the source on wave propagation [7]. Generally, the quality of numerically generated waveforms is close to the actual ones. Thus, accurate interpretation of such waveforms could be extremely beneficial.

There are many methods for the analysis of such waveforms; artificial neural networks (ANNs) is a common one. Relating work on inverse analysis and defect identification problems solved by optimization and ANNs can be found in [8–10]. Relative work can be, also, found in [11, 12]. These approaches exploit relatively simple ANN topologies, using a few selected inputs. Therefore, the users must have quite extended experience in order to choose the measurements to use, while the effectiveness of the neural network cannot be guaranteed or optimized.

1.2. Our contribution

In this paper, we exploit the unlabeled data in order to improve detection accuracy. In order to do so we employ a graph based approach and a noise modeling approach for the mapping of the waveforms onto a new manifold. Given a set of waveforms, an experts' help is required only for a small subset (i.e. $\leq 40\%$ of the available samples). The proposed approach reduces labeling effort, which is both costly and time consuming. Such an approach encourages the data sampling, since larger the data base better the classification accuracy.

Data availability was not an issue, since we used numerical simulation in order to create various pile types with defects. The rest of the paper is organized as follows: Section 2 briefly describes the proposed methodology, regarding data generation and defects detection. Section 3 refer to the experimental setup and provides the results.

2 PROPOSED METHODOLOGY

The proposed approach is suitable for low strain integrity tests, carried out in time domain. In time domain reflectometry, the wave is generated by a hand held hammer blow impact and the response as a function of time is picked up by multiple accelerometers, placed on piles' head and around it, on a circle base, close to the location of hammer blow. Monitoring and analysis of these reflections form the basis of PIT [13].

In this paper similar tests with the ones performed in the laboratories are modelled by employing a coupled finite element method (FEM) together with scaled boundary finite element method (SBFEM) approach [14]. Numerical simulation is used for the data generation. Graph label propagation [15] is, then, used for the defect identification. Data generation involves the generated waveforms (time domain), while graph detectors provide results regarding the integrity testing, by exploiting the information of *all available data* provided, following a specific feature extraction.

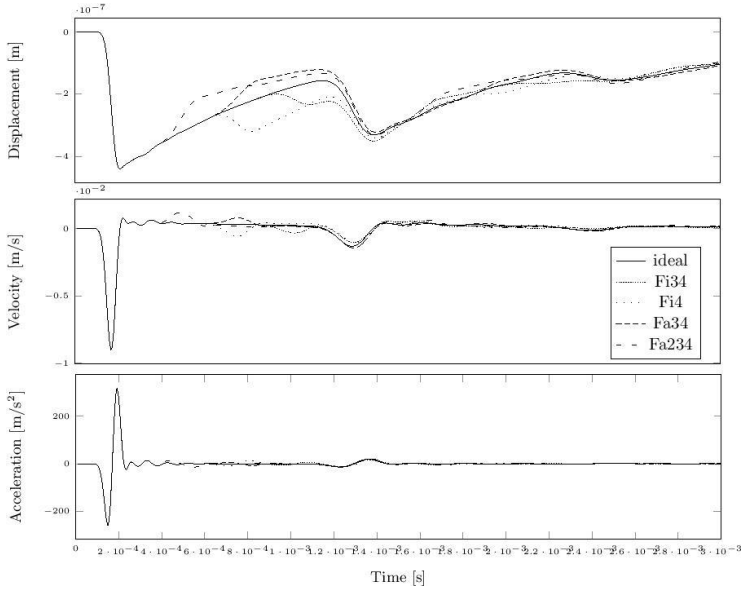

Figure 1.Time dependent plots for displacement, velocity and acceleration for pile configuration: ideal, Fi34, Fi4, Fa34 and Fa234. More information about piles' defects can be found in Sec. 3.1.

2.1 Piles Numerical Simulation

In order to simulate the wave propagation through the piles a coupled FEM and SBFEM approach is used. This approach satisfies Sommerfeld's radiation condition and allows simulating an infinite half space. This ensures that the applied impulse will not be reflected at the artificial boundary which is introduced by the boundary of the numerical discretization. The coupled approach proposed here requires only the discretization of a small domain compared to a purely FEM-based approach.

FEM and SBFEM are used to model the near field and far field, respectively. The equation of motion at an arbitrary time step can be written as:

$$\begin{bmatrix} \boldsymbol{M}_{nn} & \boldsymbol{M}_{nf} \\ \boldsymbol{M}_{fn} & \boldsymbol{M}_{ff} \end{bmatrix} \ddot{\boldsymbol{u}} + \begin{bmatrix} \boldsymbol{K}_{nn} & \boldsymbol{K}_{nf} \\ \boldsymbol{K}_{fn} & \boldsymbol{K}_{ff} \end{bmatrix} \boldsymbol{u} = \begin{bmatrix} \boldsymbol{p}_{nn} \\ \boldsymbol{p}_{ff} \end{bmatrix} - \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{p}_{b} \end{bmatrix} \tag{1}$$

where the vector $\boldsymbol{u}$ represents the nodal displacement, $\ddot{\boldsymbol{u}}$ the nodal acceleration, and $\boldsymbol{p}$ denotes the applied nodal forces. $\boldsymbol{M}$ is the mass matrix and $\boldsymbol{K}$ stands for the stiffness matrix. Here, matrix blocks with subscript nn contain the nodes of the near field while blocks with subscript ff comprise the nodes of the far field. The coupling of near and far field nodes is reflected in those blocks subscribed with nf and fn. Vector $\boldsymbol{p}_b$ denotes the far field influence on the near field, so that the behavior of the infinite half space can be applied to the FEM sub-domain as a load.

The far field is represented by the forces of the far field $\boldsymbol{p}_b$ at the interface given by the convolution integral

$$\boldsymbol{p}_b(t) = \int_0^t \boldsymbol{M}^{\infty}(t-\tau)\, \ddot{\boldsymbol{u}}(\tau) d\tau \tag{2}$$

where $\boldsymbol{M}^{\infty}$ is the acceleration unit-impulse response matrix. Detailed information on how the acceleration unit-impulse response matrices are assembled are published in [14, 16]. An illustration of the generated waveforms at the piles' surface is shown in Figure *1*.

2.2 Feature Extraction

Once a load p is applied, at the top and center of the pile an oscillation occurs as a result of wave propagation through the piles' structure. For a predefined time duration T_t the oscillating patterns $\boldsymbol{O}_{p,i}$ are recorded for every node i. These patterns have the form of

$$\boldsymbol{O}_{p,i} = \begin{bmatrix} x_{d,\mathrm{i}} & y_{d,\mathrm{i}} & z_{d,\mathrm{i}} \\ x_{v,i} & y_{v,\mathrm{i}} & z_{v,\mathrm{i}} \\ x_{a,\mathrm{i}} & y_{a,\mathrm{i}} & z_{a,\mathrm{i}} \end{bmatrix} \tag{3}$$

where d, v and a stand for displacement, velocity and acceleration respectively. So all information regarding a piles' behavior is expressed by:

$$\boldsymbol{S}_{pile} = [\boldsymbol{O}_{p,1} \cdots \boldsymbol{O}_{p,m}] \tag{4}$$

where $\boldsymbol{S}_{pile}$ denotes the available information about the waveform in any of the m nodes for a total time T_t.

A waveform that describes the oscillating behavior (or recorded observation 340 for simplicity) for each node represents the base for our analysis. However, the recorded observations for each of the piles' nodes are too large to be processed and it is suspected to be non-informative after a time period. The simplification of the amount of resources, required to describe a large set of data accurately, is possible taking two main assumptions into account:

1. Ideal pile behavior is known. That can be achieved through CAD models and numerical simulation. Every one of the investigating nodes has its corresponding ideal waveform. As we will see that is of major importance during the feature extraction of the data.
2. There is a transient period with sufficient information, for every observed node. In other words, a short period of time includes most of the important signal variations, needed by the model in order to recognize the type of the defect.

The feature extraction is based on signal subtraction, a common technique applied in noise modelling [17]. Thus, the first step is the subtraction stage, so that:

$$\boldsymbol{S}_{pile}^{new} = \boldsymbol{S}_{pile}^{check} - \boldsymbol{S}_{pile}^{ideal} \tag{5}$$

where $\boldsymbol{S}_{pile}^{ideal}$ denotes the generated signal from a pile without defects. We also define:

1. The transient period time T_{trans}. During this period the wave propagates from the pile's top to the bottom and then to the top again. After T_{trans}, waveforms get a complicated form due to the waves deflections and reflections.
2. Feature space dimension, n_v. We map each oscillation pattern to a new space $\mathbb{R}^{n_v \times 1}$. Values of n_v less than 6 are unable to create descriptive feature vectors, while values greater than 40 may require additional training (labeled) data for a smooth classification performance.
3. The mapping function from oscillation space to feature space. In this paper we used three alternatives: Mean absolute error, mean square error and difference in specific time steps.

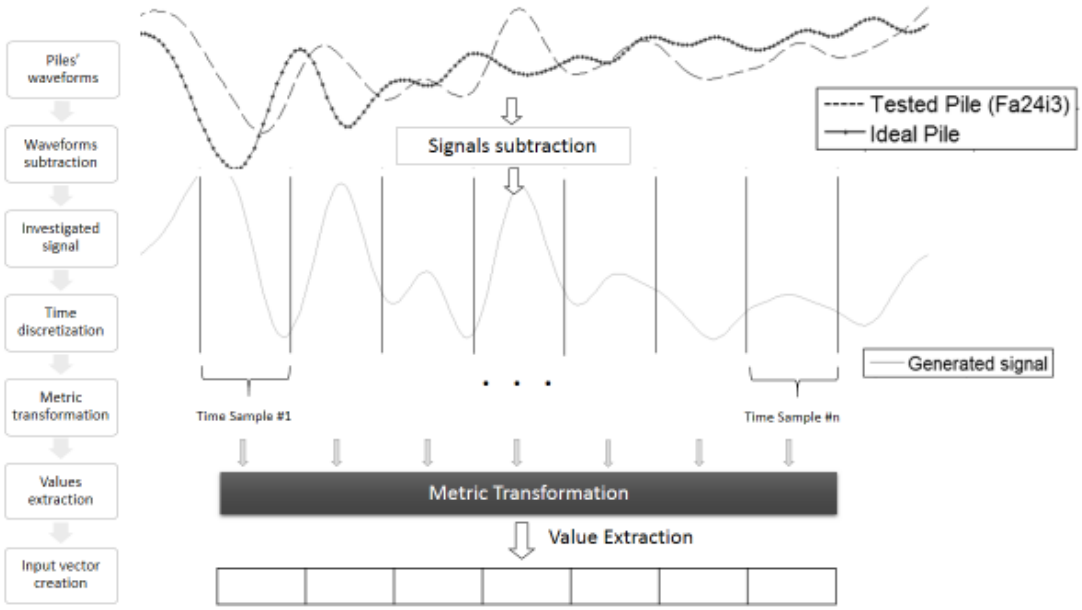

Figure 2. Input vector creation process for the neural detector. The process exploits the differences between the investigated and the ideal pile.

2.3 Graph Based Semi Supervised Classification

Graph-based semi-supervised methods define a graph over the entire data set, $\boldsymbol{X} = \boldsymbol{X}_L \cup \boldsymbol{X}_U$, where, $\boldsymbol{X}_L = \{(\boldsymbol{x}_1, \boldsymbol{y}_1), \dots, (\boldsymbol{x}_l, \boldsymbol{y}_l)\}$, is the labeled data set and $\boldsymbol{X}_U = \{\boldsymbol{x}_{l+1}, \dots, \boldsymbol{x}_{l+u}\}$ the unlabeled data set. Feature vectors, $\boldsymbol{x}_i \in \mathbb{R}^m, i = 1, \dots, l+u$, are available for all the observations and $\boldsymbol{y}_i \in \mathbb{R}^k, i = 1, \dots, l$, are the corresponding classes of the labeled ones, in a vector form; k denotes the available classes. The nodes represent the labeled and unlabeled, examples in the dataset; edges reflect the similarity among examples. These methods usually assume label smoothness over the graph. That is, if two instances are connected by a strong edge, their labels tend to be the same.

In our case the harmonic function approach is adopted [18]. This approach estimates a function f on the graph which satisfies two conditions. Firstly, f has the same values as given labels on the labeled data, i.e. $f(\boldsymbol{x}_i) = \boldsymbol{y}_i, i = 1, \dots, l$. Secondly, f satisfies the weighted average property on the unlabeled data:

$$f(\boldsymbol{x}_j) = \frac{\sum_{k=1}^{l+u} w_{jk} f(\boldsymbol{x}_j)}{\sum_{k=1}^{l+u} w_{jk}}, j = l+1, \dots, l+u \tag{6}$$

where w_{ij} denotes the edge weight. Those two conditions lead to the following problem:

$$\min_{f: f(\boldsymbol{x}) \in \mathbb{R}} \sum_{i,j=1}^{l+u} w_{ij} \left(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j) \right)^2 \tag{7}$$

$$s.t.\ f(\boldsymbol{x}_i) = \boldsymbol{y}_i, i = 1, \dots, l$$

In the following we introduce some notations in order to present the close form solution of such problem. Let $\boldsymbol{W}$ be an $(l+u) \times (l+u)$ weight matrix, whose i,j-th element is the edge weight w_{ij}. Let $D_{ii} = \sum_{j=1}^{l+u} w_{ij}$ be the weighted degree of vertex i, i.e., the sum of edge weights connected to i. Then we create a diagonal matrix $\boldsymbol{D} \in \mathbb{R}^{(l+u) \times (l+u)}$ by placing D_{ii} on the diagonal. The unnormalized graph Laplacian matrix $\boldsymbol{L}$ is defined as: $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$. Matrix $\boldsymbol{L}$ is rearranged in the form:

$$\boldsymbol{L} = \begin{bmatrix} \boldsymbol{L}_{ll} & \boldsymbol{L}_{lu} \\ \boldsymbol{L}_{ul} & \boldsymbol{L}_{uu} \end{bmatrix} \tag{8}$$

Let $\mathbf{f} = \left(f\left(x_1\right), \ldots, f\left(x_{l+u}\right)\right)^T$ be the vector of f values on all vertices arranged in a way that $\mathbf{f} = (\mathbf{f}_\mathrm{l}, \mathbf{f}_\mathrm{u})$ and let $\boldsymbol{Y}_l = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_l)^T$. The harmonic solution is:

$$\begin{aligned} \mathbf{f}_\mathrm{l} &= \boldsymbol{Y}_l \\ \mathbf{f}_\mathrm{u} &= \boldsymbol{L}_{uu}^{-1}\boldsymbol{L}_{ul}\boldsymbol{Y}_l \end{aligned} \tag{9}$$

Thus, we are able to estimate (soft label) output vectors for all the edges of the graph. Each labeled edge, i, is guaranteed to have the output vector, $\boldsymbol{y}_i$, as it was provided by the expert. Note that weight matrix $\boldsymbol{W}$ has to be well-defined so that the graph Laplacian matrix $\boldsymbol{L}$ will be invertible. Generally, radial basis function kernel and features of good quality are able to support a well-defined weight matrix.

3 EXPERIMENTAL SETUP

The defect recognition can be seen as a classification problem. Assume that a pile is separated in q parts. Also, the available information is limited in few waveforms, recorded on the pile's surface. We have to classify each of these parts in one out of three categories. The "neck"-category indicates the existence of a "neck", i.e. smaller radious than expected. The "bulb"-category indicates the existence of a "bulb", i.e. greater radious than expected. There is, also, the "no-defect"-category, where there are neither "bulbs" nor "necks". Feature extraction and defect recognition routines are written in MatLab code.

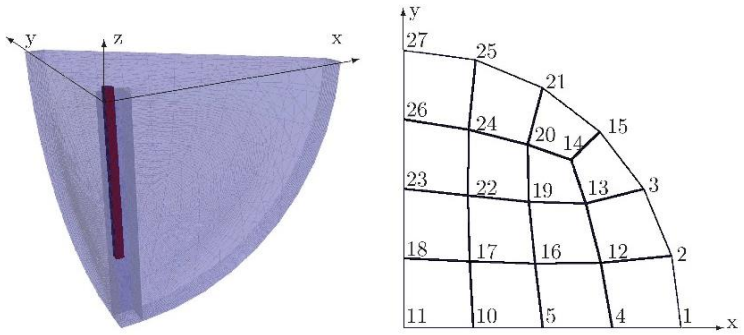

Figure 3. Left: The FEM near field discretization includes pile and surrounding soil; Right: Top view at the piles surface and the corresponding node numbers.

3.1 Piles Set Description

Different three-dimensional pile configurations are analyzed. All investigated piles are modelled as floating piles, since no bedrock is taken into account. One clean pile without defects, Figure 4 (ideal), is discretized. Length l_0 and radius r_0 are chosen as $2.1m$ and $0.1m$, respectively. The surface of the ground is defined at $0.0m$, the pile's head is located $+0.1m$ over the surface, while the piles' toe is at $-2.0m$ in the ground. Additional piles with defects are discretized as well, the geometries of these modified piles are shown in Figure 4. The finite element mesh is shown in Figure 3. The near field is discretized by 116974 tetrahedral elements and 28140 hexahedral elements and 151833 degrees of freedom. The attached far field is discretized by 549 quadrangular elements and 1794 degrees of freedom. All elements are using linear finite and scaled boundary finite elements. The minimum elements length is $l_{min} = 7.10347 \times 10^{-3}\ m$ and the maximum length is $l_{max} = 6.46424 \times 10^{-1}\ m$.

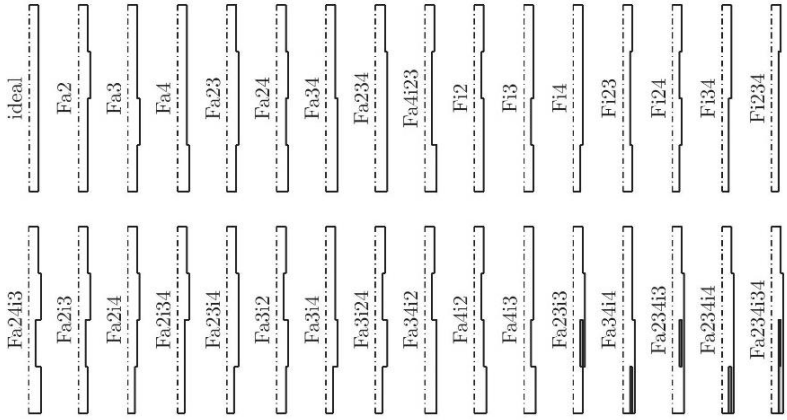

Figure 4. Pile geometries and naming of the cases, half the longitudinal section of the pile is pictured. The pile is divided into 4 sections enumerated from 1 to 4 top down. The "a" stands for additional pile material and the "i" inner distortion of pile cross section. So, "Fa2" gives additional material in section 2. The last five geometries in the second row have an internal distortion, which is represented with a soil inclusion in the concrete body of the pile.

3.2 Training, Validation and Evaluation Sets

Let us first focus on the data creation. The data, for the graph construction, originates from the subtraction between two signals (i.e. the ideal and the examined pile) in the process shown in Figure *2*. Although the signal duration is 2000 time steps (or 0.012 seconds), only the first 400 time steps of the transient period were utilized, since after that period the wave signal is backward propagated causing interference in the signal altitude. This way, we would create an input signal of size 400×1, which causes misclassification issues, due to the high dimensionality. To handle this problem, we equally downsample the input signal by 10. Thus, input vector is of size 40×1.

For every input vector there is a corresponding output vector of size $k \times 1$, where k denotes the number of pillar parts that are investigated (in our case $k = 4$). The number of parts was selected to facilitate the numerical simulations, in terms of computational complexity. However, division into greater number of parts is feasible, for simulating more complex structures. The first part is located above the ground.

Specific nodes were used to form the training data, while the remaining formed the evaluation data. The two different data sets are described in Table 1. For each node, waveforms of nine different cases were available:

1. x-axis: displacement (x-d), velocity (x-v) and acceleration (x-a)
2. y-axis: displacement (y-d), velocity (y-v) and acceleration (y-a)
3. z-axis: displacement (z-d), velocity (z-v) and acceleration (z-a).

However, due to modelling assumptions, regarding the boundary conditions of the FEM-SBFEM approach, specific nodes oscillation patterns had to be excluded from the data generation (as non-informative waveforms). These oscillations refer to x and y axes, but not to z-axis. Due to symmetry boundary conditions these x and y data should be equal to zero. If not they are practically zero.

Data set	Training nodes	Evaluation nodes
Ts1	{1, 15, 25 }	{ 2, 3, 4, 5, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 22, 23, 24, 26, 27 }
Ts2	{17, 25 }	{ 1, 2, 3, 4, 5, 10, 11, 12, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 26, 27 }
Ts3	{2, 19 }	{ 1, 3, 4, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 25, 26, 27 }
Ts4	{1, 21 }	{ 2, 3, 4, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 26, 27 }

Table 1. Training and evaluation nodes for each of the created data sets.

3.3 Classification Performance

The proposed approach has been evaluated against two well-known classification approaches: k nearest neighbors [19] (*k*NN) and ANNs [20]. All methods are characterized as soft labeled; corresponding outputs are not integers. Thus, we adopt a mapping procedure. For a given pile, a specific output vector, $\boldsymbol{P}_i = [p_1, p_2, p_3, p_4]$, is generated. The p_i values, $i = 1, \dots, 4$, correspond to a certain defect type, d_f , according to the following transformation:

$$d_f = \begin{cases} -1 & , p_i \in (-\infty, -0.5) \\ 0 & , p_i \in [-0.5, 0.5] \\ 1 & , p_i \in (0.5, \infty) \end{cases} \quad (10)$$

For the i-th part, value $df = -1$ suggests the existence of a "neck", while value $d_f = 1$ suggests the existence of a "bulb". Value $df = 0$ corresponds to non-detection of any defect. Range selection for the value intervals in eq. 10 stems from equal division of the detectors' interval range of $[-1, 1]$ into three examined defect types.

ANNs and kNN were available through MatLab toolboxes. Harmonic label propagation function and weight matrix creation functions where provided by [21] and [22] respectively. A random search among a variety of ANN topologies was performed. ANNs parameters' range is shown in Table 2. The number of nearest neighbors was set to 4, for both *k*NN and Harmonic function approaches. Indicative classification results are shown in Table 3.

Parameter's name	Min value	Max value
Hidden layers	1	3
Neurons / layer	5	20
Training epochs	100	400

Table 2: Parameters' range for the ANN classification approach.

	ANN			*k*NN			Harmonic		
descriptor	**MSE**	**MAE**	**Diff**	**MSE**	**MAE**	**Diff**	**MSE**	**MAE**	**Diff**
x_d	0.366	0.369	0.362	0.359	0.363	0.370	0.356	0.377	**0.382**
x_v	0.224	0.225	0.222	0.247	0.259	**0.265**	0.252	0.255	0.260
x_a	0.345	0.361	0.376	0.364	0.381	**0.400**	0.347	0.362	0.375
y_d	0.468	0.490	0.486	0.488	0.500	0.508	0.498	0.508	**0.515**
y_v	0.287	0.300	**0.310**	0.292	0.307	0.310	0.241	0.242	0.238
y_a	0.452	0.480	**0.502**	0.442	0.454	0.481	0.380	0.382	0.394
z_d	0.706	0.715	0.723	0.764	0.743	0.738	0.831	**0.839**	0.837
z_v	0.725	0.736	0.758	0.704	0.706	0.710	0.817	0.819	**0.830**
z_a	0.689	0.691	0.696	0.735	0.736	0.732	0.787	0.764	**0.789**

Table 3: Classification accuracy over the evaluation set. The simulation's results correspond to average values over the different data sets, as described in Table 1.

SSL graph based approaches limitations do not apply in our case, given the application field. Although, graph based methods scale badly as the number of data raises (i.e. $O(n^2)$), we do not expect any implications given the limited amount of available waveforms (we expect no more than few hundreds of samples, even in real life cases). Another possible limitation is the transductive nature of the approach; graph based approaches are unable to handle new data. In such case, we have to recreate the graph and all the corresponding matrices.

The proposed approach performs better than the other commonly used techniques. Compared to the ANN, harmonic function needs a considerably smaller amount of labeled data and requires less heuristic approaches for the best topology definition. Generally, ANN performance highly depends on various parameters regarding the detector structure, such as number of hidden layers, neurons and training epochs. Compared to the *k*NN, Harmonic function exploits additional information of the unlabeled data. Thus, as labels propagate, edges' labels match the k closest ones from the entire data space.

Results suggest that bulb or neck cases are easily identifiable. However, normal radius is classified either as neck or bulb more frequently (see Figure 5). The misclassification of non-defective pile parts can be partially explain by the waveform, which has a specific form defined by the defect(s) type. Such form doesn't change significantly if the defective pile part is followed by a non-defective part. Thus, we have a waveform, indicating a defect, crossing a non-defective part. According to the similarity mechanism of the proposed methodology, such form is more likely to indicate a defection, resulting in wrong classification. The misclassification rates grow as more defects appear in the same pile.

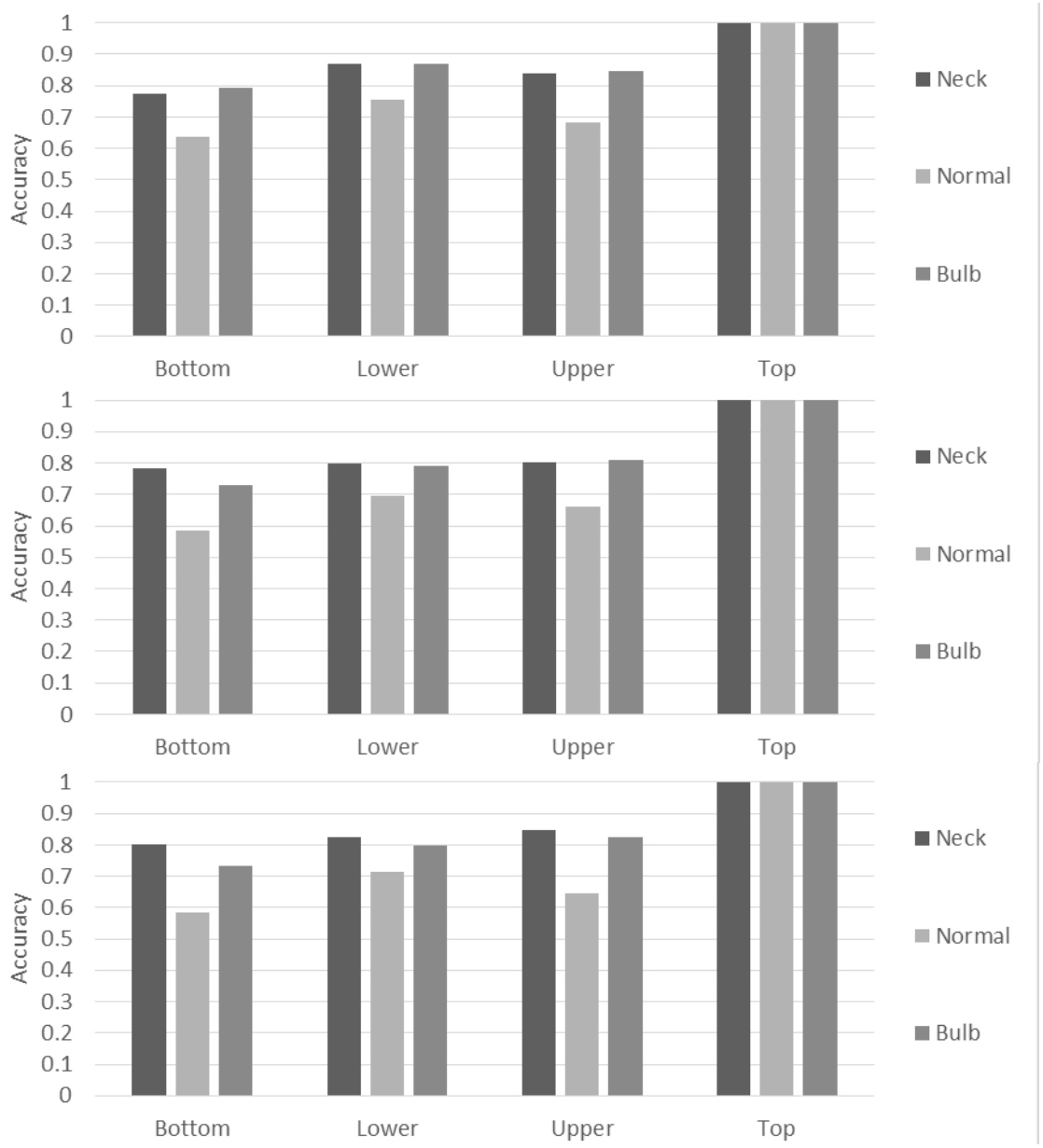

Figure 5. Defect Classification accuracy for each one of the 4 pile's parts. Results are based on displacement observations in x,y,z axes using: (a) Difference in values (b) MAE and (c) MSE as quality metric.

4 CONCLUSIONS

One method to assess the behavior of a pile is to apply non-destructive testing through the use of low strain integrity tests in time domain. The wave is generated by a hand held hammer blow impact and the response, as a function of time, is picked up by multiple accelerometers, placed on pile head and around it, on a circle base, close to the location of hammer blow. Then we need to apply signal processing methods on the waveforms generated in order to detect the defects.

Initially, appropriate features were extracted in order to map piles' waveforms to meaningful short-length signals. Then, these features form a graph, where the labels propagate among the edges utilizing both labeled and unlabeled data. The performed experiments provide very promising results; the defect recognition rate is above 80%, when z-axis observations are used. On the contrary, the performance based on x and y axes behavior patterns is severely low. The problem formulation can easily be expanded in piles with more divisions (i.e. more than 4) and varying defects' diameters. Although, a greater misclassification error is likely, as long as we deal with "neck"

or “bulb” detection, the performance is expected to remain high.

Finally, we observe that high detection rates are achievable using only a handful set of samples for training. The detection rates are also affected by the shape and the depth; deeper the defect harder to locate. Piles with more complex structure will be evaluated in future work. Furthermore, new adaptation strategies and detection techniques will be investigated to handle non-stationary waveforms cases.

ACKNOWLEDGEMENTS

The work has been partially supported by IKY Fellowships of excellence for postgraduate studies in Greece-Siemens program. Partial support from a Greek-German scientific and technological cooperation project IKYDA 2012 is gratefully acknowledged.

REFERENCES

1. Garnier, C., Pastor, M.-L., Eyma, F., Lorrain, B.: The detection of aeronautical defects in situ on composite structures using Non Destructive Testing. Compos. Struct. 93, 1328–1336 (2011).
2. Zhu, X., Goldberg, A.B.: Introduction to Semi-Supervised Learning. Morgan & Claypool Publishers (2009).
3. Huang, Y.-H., Ni, S.-H.: Experimental study for the evaluation of stress wave approaches on a group pile foundation. NDT E Int. 47, 134–143 (2012).
4. Hola, J., Schabowicz, K.: State-of-the-art non-destructive methods for diagnostic testing of building structures – anticipated development trends. Arch. Civ. Mech. Eng. 10, 5–18 (2010).
5. Haddad, A.: Numerical Simulation for Wave Propagation in Pile Integrity Test. Deep Foundations and Geotechnical In Situ Testing. pp. 224–231. American Society of Civil Engineers (2010).
6. Huang, Y.-H., Ni, S.-H., Lo, K.-F., Charng, J.-J.: Assessment of identifiable defect size in a drilled shaft using sonic echo method: Numerical simulation. Comput. Geotech. 37, 757–768 (2010).
7. Chai, H., Phoon, K., Zhang, D.: Effects of the Source on Wave Propagation in Pile Integrity Testing. J. Geotech. Geoenvironmental Eng. 136, 1200–1208 (2010).
8. Stavroulakis, G.E.: Inverse and crack identification problems in engineering mechanics. Springer (2000).
9. Stavroulakis, G., Bolzon, G., Waszczyszyn, Z., Ziemianski, L.: Inverse analysis. Compr. Struct. Integr. 3, 685–718 (2003).
10. Stavroulakis, G.E., Engelhardt, M., Likas, A., Gallego, R., Antes, H.: Neural network assisted crack and flaw identification in transient dynamics. J. Theor. Appl. Mech.-Wars.-. 42, 629–650 (2004).
11. Tam, C.., Tong, T.K.., Lau, T.C.., Chan, K..: Diagnosis of prestressed concrete pile defects using probabilistic neural networks. Eng. Struct. 26, 1155–1162 (2004).
12. Zhang, C., Zhang, J.: Application of Artificial Neural Network for Diagnosing Pile Integrity Based on Low Strain Dynamic Testing. In: Yuan, Y., Cui, J., and Mang, H.A. (eds.) Computational Structural Engineering. pp. 857–862. Springer Netherlands (2009).
13. Schauer, M., Langer, S.: Numerical Simulations of Pile Integrity Tests Using a Coupled FEM SBFEM Approach. PAMM. 12, 547–548 (2012).
14. Schauer, M., Roman, J.E., Quintana-Ortí, E.S., Langer, S.: Parallel Computation of 3-D Soil-Structure Interaction in Time Domain with a Coupled FEM/SBFEM Approach. J Sci Comput. 52, 446–467 (2012).
15. Wang, F., Zhang, C.: Label Propagation through Linear Neighborhoods. IEEE Trans. Knowl. Data Eng. 20, 55–67 (2008).
16. Wolf, J.P., Wolf, J.P., Wolf, J.P.: The scaled boundary finite element method. Wiley Chichester (W. Sx.) (2003).
17. Lipponen, J.A., Tarvainen, M.P.: Advanced maternal ECG removal and noise reduction for application of fetal QRS detection. Computing in Cardiology Conference (CinC), 2013. pp. 161–164 (2013).
18. Zhu, X.: Semi-supervised learning using gaussian fields and harmonic functions. Proceedings of the 20th International Conference on Machine learning (ICML-2003). p. 912 (2003).
19. Bhatia, N., Vandana: Survey of Nearest Neighbor Techniques. ArXiv10070085 Cs. (2010).
20. Tan, G., Liu, H., Cheng, Y., Liu, B., Zhang, Y.: Prediction method for the deformation of deep foundation pit based on neural network algorithm optimized by particle swarm. 2011 International Conference on Transportation, Mechanical, and Electrical Engineering (TMEE). pp. 1407–1410 (2011).
21. Zhu, X., Lafferty, J., Ghahramani, Z.: Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. ICML 2003 workshop on the continuum from labeled to unlabeled data in machine learning and data mining. pp. 58–65 (2003).
22. Liu, W., Chang, S.-F.: Robust multi-class transductive learning with graphs. IEEE Conference on Computer Vision and Pattern Recognition, 2009. pp. 381–388 (2009).