Research Article/Mechanical Engineering, Mechatronics, and Materials Science

http://doi.org/10.15446/ing.investig.111646

Electromechanical Impedance-Based Damage Detection Using Machine Learning Approaches

Detección de daños basada en impedancia electromecánica mediante métodos de aprendizaje automático

Paulo Elias Carneiro Pereira[1], Stanley Washington Ferreira de Rezende[2], Bruno Pereira Barella[3], José dos Reis Vieira de Moura Júnior[4], and Roberto Mendes Finzi Neto[5]

ABSTRACT

Electromechanical impedance-based structural health monitoring has been the subject of extensive research in recent decades. The method's low cost and ability to detect minor structural damages make it an appealing alternative to other non-destructive techniques. Ongoing research on damage detection approaches continues to be a topic of interest in relation to the electromechanical impedance method. This work proposes the use of the K-Means, Decision Tree, and Random Forest algorithms to distinguish between four structural conditions in an aluminum beam. These techniques were applied to raw impedance data and a dataset reduced via principal components analysis. The findings revealed that the compressed dataset improved the accuracy of all models, except for the Random Forest approach, whose accuracy decreased by 2.9%. The K-Means algorithm was most affected by the reduction in dimensionality, with a 105.9% increase in accuracy. The Decision Tree and Random Forest methods yielded outstanding outcomes, comparable or superior to other state-of-the-art approaches. This makes them a compelling choice for detecting damage using electromechanical impedance data, even when using raw data as the input information.

Keywords: electromechanical impedance method, K-means algorithm, Decision Tree, Random Forest, structural health monitoring

RESUMEN

El monitoreo de la salud estructural basado en la impedancia electromagnetica ha sido objeto de investigacion exhaustiva en las ultimas decadas. El bajo coste del metodo y su capacidad para detectar daños estructurales menores lo convierten en una alternativa atractiva a otras tecnicas no destructivas. La investigacion actual sobre enfoques de deteccion de daños sigue siendo un tema de interes en lo que concierne al metodo de impedancia electromecanica. En este trabajo se propone utilizar los algoritmos K-Means, Decision Tree y Random Forest para diferenciar entre cuatro condiciones estructurales en una viga de aluminio. Estas tecnicas se aplicaron a datos de impedancia en bruto y a un conjunto de datos reducido mediante analisis de componentes principales. Los resultados revelaron que el conjunto de datos comprimido mejoro la precision de todos los modelos, excepto en el caso del metodo Random Forest, cuya precision disminuyo en un 2.9 %. El algoritmo K-Means fue el mas afectado por la reduccion de la dimensionalidad, con un aumento del 105.9 % en la precision. Los metodos Decision Tree y Random Forest produjeron resultados sobresalientes, comparables o superiores a otros enfoques de vanguardia. Esto los convierte en una opcion convincente para detectar daños a traves de datos de impedancia electromecanica, incluso cuando se utilizan datos en bruto como informacion de entrada.

Palabras clave: Metodo de impedancia electromecanica, Algoritmo K-Medias, Arbol de Decision, Bosque Aleatorio, Control del Estado Estructural.

Received: October 17th 2023
Accepted: May 29th 2024

Introduction

Among the non-destructive methods developed for the structural health monitoring (SHM) of structures and mechanical systems, the method based on electromechanical impedance (EMI) is particularly strategic due to its ability to detect minor structural damage, providing timely data for intervention (Giurgiutiu, 2014).

Damage detection through EMI data is a crucial aspect in the practical application of this technique. To address this issue, various approaches have been devised to enhance damage detection using EMI technology, as demonstrated by Kim and Wang (2019), Fan and Li (2020), Zhou *et al.* (2021), Wang *et al.* (2022), and Nguyen *et al.* (2023).

Machine learning (ML) approaches for damage detection have yielded promising outcomes across various scenarios, with research endeavors delving into both supervised and unsupervised methods to unravel and leverage their potential. Specifically, in the realm of supervised ML

[1]Mining engineer, Federal University of Goiás, Brazil. MSc Modeling and Optimization, Federal University of Goiás, Brazil. Affiliation: PhD Candidate, Mechanical Engineering, Federal University of Uberlândia, Brazil. Assistant professor, Federal University of Catalão, Brazil. E-mail: paulo_elias_carneiro@ufu.br, paulo_elias_carneiro@ufcat.edu.br
[2]Industrial mathematician, Federal University of Goiás, Brazil. MSc Modeling and Optimization, Federal University of Goiás , Brazil. Affiliation: PhD Candidate, Mechanical Engineering, Federal University of Uberlândia, Brazil. E-mail: stanley_washington@ufu.br
[3]Industrial mathematician, Federal University of Goiás, Brazil. MSc Modeling and Optimization, Federal University of Goiás , Brazil. Affiliation: PhD Candidate, Computer Science, Federal University of Uberlândia, Brazil. E-mail: brunobarella@ufu.br
[4]Mechanical engineer, Federal University of Uberlândia, Brazil. MSc Mechanical Engineering, Federal University of Uberlândia, Brazil. PhD Mechanical Engineering, Federal University of Uberlândia, Brazil. Affiliation: Assistant professor, Federal University of Uberlândia, Brazil. E-mail: zereis@ufu.br
[5]Electrical engineer, Federal University of Uberlândia, Brazil. MSc Electrical Engineering, Federal University of Uberlândia, Brazil. PhD Electrical Engineering, Federal University of Uberlândia, Brazil. Affiliation: Full professor, Federal University of Uberlândia, Brazil. E-mail: finzi@ufu.br

techniques for damage detection, the predominantly employed algorithms center around neural networks, notably convolutional neural networks (CNNs) (Oliveira *et al.*, 2018a; Rezende *et al.*, 2020, Li *et al.*, 2021; Ai and Cheng, 2023; Du *et al.*, 2023; Nguyen *et al.*, 2023). Other noteworthy neural network architectures include probabilistic neural networks (PNNs) (Na, 2021), deep residual networks (DRN) (Alazzawi and Wang, 2021), fuzzy ARTMAP networks (FAN) (Oliveira *et al.*, 2018b), and extreme learning machines (ELMs) (Djemana *et al.*, 2017). These applications have demonstrated significant efficacy in classifying damage levels, often incorporating considerations for temperature variations.

Non-neural network-based supervised approaches include the random forest (RF) method, which was applied, among other techniques, by Yan *et al.* (2022) to monitor the early hydration of cement mortar, with promising results. Another application of the RF method developed for EMI-based monitoring was the prediction of impedance signatures in steel structures (Parida *et al.*, 2023).

In parallel, in the realm of unsupervised ML approaches, principal components analysis (PCA) has emerged as a pivotal tool in EMI-based monitoring. The method has found application as a damage index (Malinowski *et al.*, 2021) and has been employed for extracting crucial features from original impedance or admittance signatures (Park *et al.*, 2008; Jiang *et al.*, 2021b; Jiang *et al.*, 2021c; Ai *et al.*, 2022).

In the context of the latter, PCA has been employed in conjunction with various complementary methods, including K-Means (KM) clustering (Park *et al.*, 2008; Jiang *et al.*, 2021b), the support vector machine (SVM) method (Jiang *et al.*, 2021c), and artificial neural networks (ANN) (Ai *et al.*, 2022). These combined approaches have yielded promising results in EMI-based monitoring for damage detection.

Following the application of unsupervised methods in EMI-based monitoring, Perera *et al.* (2019) used hierarchical clustering and the KM method to detect damages in reinforced concrete (RC) beams.

PCA and KM show promise among the unsupervised methods used in EMI-based monitoring. However, there is a pressing need for more studies to explore the alternative applications of these techniques for effective damage detection. Moreover, the realm of supervised ML methods, including decision trees (DT) and RF, remains relatively uncharted or inadequately examined in EMI-based monitoring.

To address this research gap, our study offers a comprehensive comparative analysis of the KM, DT, and RF methods for damage detection using EMI signals, focusing on their capabilities in classifying four distinct structural conditions and utilizing two types of datasets: unprocessed raw impedance signatures and PCA-compressed datasets. Additionally, the study delves into the impact of dimension reduction on the models' efficacy in predicting damage occurrences.

To the authors' knowledge, the combined study of the DT and RF methods with EMI measurements for damage detection is limited or nonexistent. Hence, this work introduces a novel perspective, wherein the application of these methods can offer fresh insights into approaches for EMI-based damage detection.

Theoretical background

This section presents the theoretical concepts underlying the EMI method, followed by information on the supervised and unsupervised ML methods used to improve damage detection in this work.

Electromechanical impedance method

The EMI method is based on the electromechanical coupling between a structure and a piezoelectric patch, usually made of lead zirconate titanate (PZT), bonded with or embedded into the interrogated structure. The patch is excited by a sinusoidal voltage of varying frequency, resulting in a deformation of the PZT due to the reverse piezoelectric effect as a function of the application of the electric field, whose deformation extends through the local vicinity of the patch (Martowicz and Rosiek, 2013).

The deformation of the patch results in the local deformation of its surroundings, which is returned to the PZT in the form of an electrical signal due to the direct piezoelectric effect. The returned signal is a signature of the structure's health. Any changes in the impedance (or admittance) signals reveal changes in the structural conditions and indicate the occurrence of damage (Sun *et al.*, 1995; Park *et al.*, 2003).

Damage indices are typically used as a quantitative approach to identify damage and evaluate its severity. The developed damage indices include the root mean square deviation (RMSD), the mean absolute percentage deviation (MAPD), the correlation coefficient (CC), the correlation coefficient deviation (CCD), and the changes in peak frequencies (Giurgiutiu, 2014; Fan and Li, 2020; Nomelini *et al.*, 2020; Gonçalves *et al.*, 2021).

The information obtained from EMI-based structural monitoring guides the condition-based maintenance of structures and mechanical systems.

Supervised learning

During supervised learning, the algorithm uses two types of information to predict a new and unobserved state: the input vectors, which are the independent variables, and their corresponding labels; or target variables, which depend on the input vectors (Bishop, 2006; Simeone, 2018; Géron, 2019).

Many algorithms based on supervised learning have been developed. This work used the DT and RF methods to classify the health states in an aluminum beam.

The DT method is a recursive model in which a given number of units are divided into subgroups, and the impurity of a set is reduced by creating subgroups (Rokach and Maimon, 2009; Kim and Upneja, 2014), thereby establishing an effective model for solving classification and regression problems (Liu *et al.*, 2021).

The structure of the method is a tree-like hierarchy consisting of a root, internal or test nodes, and leaves (Liu *et al.*, 2021). The root corresponds to the initial node and has no incoming edges. The internal or test nodes have outgoing edges and form a set of nodes called a *branch*. The leaves correspond to a terminal node, have only one incoming edge, and refer to a subgroup that cannot be further divided,

representing decision outcomes (Koza, 1991; Rokach and Maimon, 2009; Kim and Upneja, 2014; Hong *et al.*, 2018).

During the classification process, the DT algorithm starts at the root node containing all objects in the training dataset, performs a test on an attribute, and generates child nodes according to the value of the objects. The process is repeated recursively in each node and ends when a stopping criterion is met (Jegadeeshwaran and Sugumaran, 2013; Chen *et al.*, 2018; Hong *et al.*, 2018; Fletcher and Islam, 2019; Barros *et al.*, 2023; Loyola-González *et al.*, 2023). The resulting leaf or external nodes represent the outcomes or predicted classes (Chen *et al.*, 2020; Biswal and Parida, 2022).

The DT method generates a single tree structure, which, however, may become trapped in a local optimum (Fletcher and Islam, 2019) or overfit the training set (Ning *et al.*, 2021).

This issue is overcome by using multiple DTs, *i.e.*, the RF method, an ensemble learning algorithm consisting of a set of DTs, each associated with a subsample taken from the training set using bootstrap sampling with replacement, resulting in uncorrelated samples (Chencho *et al.*, 2020; Tang *et al.*, 2021).

Bootstrapping randomly selects samples from a single input training set, generating multiple datasets to train the base classifiers (Loyola-González *et al.*, 2023). Each DT in the forest grows upon the basis of the corresponding bootstrapped dataset and the randomly selected features at each node, resulting in a collection of prediction results (Bergmayr *et al.*, 2023).

The results obtained from each tree are averaged or determined by majority voting and applied to regression and classification problems, respectively (Lim *et al.*, 2021; Dinh *et al.*, 2023). This process adds diversity to the DTs, preventing the model from finding only a local optimum and overfitting the training set (Fletcher and Islam, 2019).

Unsupervised learning

Unlike supervised ones, unsupervised methods can handle unlabeled data sets and learn the specific properties and mechanisms underlying data generation, thereby helping to understand the data structure (Simeone, 2018; Géron, 2019).

From a dataset with a given number of unlabeled observations, the goal is to learn valuable properties of the distribution, which depend on the application, as is the case of density estimation, clustering, dimensionality reduction, feature extraction, and new samples generation (Simeone, 2018).

Clustering is one of the most fundamental unsupervised learning problems. It consists of dividing the original data into clusters, each of which comprises similar items, while the clusters are different from each other (Alelyani *et al.*, 2014), thereby providing important insights into the structure of the dataset.

Among the clustering methods is the KM algorithm, a heuristic technique that groups points close to each other based on Euclidean distance, assigning them to one of the initially defined k clusters (Bishop, 2006; Simeone, 2018; Mayer, 2020).

According to the number of clusters defined, the method works as follows. First, the k centroids are randomly determined, and each sample is assigned to a cluster based on its Euclidean distance to each centroid. Based on the distance between the samples and the centroids, the latter are recalculated until the convergence criterion is reached (Harrison, 2020).

The resulting clusters contain the samples whose sum of the squares of the distances to the corresponding centroid represents the minimum value, implying an optimal configuration for the k clusters considered (Reddy and Vinzamuri, 2014).

For some problems, the number of features can reduce the model's ability to predict the classes correctly. This can be mitigated by performing dimensionality reductions via the PCA method.

PCA converts a dataset with a given number of variables into a more concise one, in which the variables or features are contained within principal components orthogonal to each other (Davis, 2002). The method projects the original high-dimensional data to a low-dimensional subspace based on the eigenvectors of the covariance matrix (Liu and Han, 2014).

The principal components define a hyperplane that, when projected onto the data, preserves as much variance as possible, so that the selected hyperplane loses less information than the other possible hyperplanes (Géron, 2019).

The first principal component accounts for most of the variance explained, and so on until the last principal component. Thus, the method can reduce the dimensionality of the dataset while preserving the behavior of the data as much as possible.

Materials and methods

The experimental setup consisted of (i) an aluminum beam; (ii) a PZT patch; (ii) an Analog Devices AD5933 acquisition board; and (iv) a personal computer. The first step was to glue the PZT patch to the 300 x 25 x 3 mm aluminum beam (Figure 1).

Figure 1. Aluminum beam and PZT patch used in the experiments
Source: Authors

The procedure for gluing the PZT patch to the structure was of great importance. Since the impedance measurement was systemic (*i.e.*, a measurement of the electro-mechanical impedance of the structure plus the PZT patch), it was essential to ensure a thin layer that was sufficient for the adhesion of the two surfaces and was not thick enough

to induce other unwanted mechanical behaviors (Islam and Huang, 2014; Albakri and Tarazaga, 2017; Bari and Moharana, 2024).

This gluing procedure consisted of applying a thin layer of epoxy resin onto the PZT patch and mounting it onto the surface of the specimen. Next, a thin 2 mm layer of EVA foam (ethylene vinyl acetate) and a weight of 1 kg were placed to ensure the homogeneity of the cure for the glue layer.

As the technique used was a differential and not an absolute procedure, *i.e.*, the damage metrics were calculated concerning the reference integrity state, the effect of variations in the glue layer was disregarded.

Using wires, the PZT patch was connected to the AD5933 acquisition board to obtain the impedance signatures. The signal gathered from the AD5933 was read by the acquisition board's software on the personal computer.

The real part of an electrical impedance is known as the *resistance* (R), and the imaginary part is the *reactance* (X). Resistance represents the opposition to the flow of electrical current due to factors such as the resistance of conductors, circuit components, and, in this experiment, the mechanical movement of the structure. Reactance represents the opposition to the flow of current which arises from reactive elements such as capacitors (capacitive reactance) and inductors (inductive reactance). In this experiment, the reactance is capacitive because of the PZT patch. Electrical impedance is then represented as a combination of these two terms, where the real part is directly associated with the mechanical properties of the studied specimen.

In this work, the resistance was employed since it is the most sensitive to the presence of damage and is less affected by environmental conditions (Na and Lee, 2013; Hamzeloo *et al.*, 2020; Meher *et al.*, 2022).

The acquisition system was set to cover a frequency interval varying from 60 to 70.2 kHz in steps of 20 Hz, resulting in 511 points in the frequency domain. The selected frequency interval was determined through trial and error, as it exhibited a high degree of sensitivity to the damage being assessed. Additionally, the interval range was selected to accommodate the resolution limitations of the AD5933 acquisition board.

The experiment involved four scenarios: one with the beam in its original state (baseline) and three others with damage. The damage was simulated by attaching a 10 x 10 x 10 mm magnet to the beam's surface in three different positions. The first damage position (Damage 1) was far from the PZT patch, whereas, in the subsequent scenarios (Damage 2 and Damage 3), the magnet was gradually brought closer to the PZT patch. For each scenario, 30 samples were collected, based on an average of 20 measurements in order to reduce random variation.

The collected data were analyzed in Microsoft's Visual Studio Code using the Python language (version 3.8.17). The datasets for each condition (Healthy State, Damage 1, Damage 2, and Damage 3) were imported using the Scipy package (Virtanen *et al.*, 2020) and consisted of two matrices. One dataset was related to the frequency data, which was presented as a one-row, 511-column vector, making it a vector of the frequency values, and the other corresponded to the real-part values of the impedance, with 30 rows and 511 columns, where each row represented a sample. Each column corresponded to a point within the frequency range in both datasets.

From the imported data set, the median of the resistance values along the frequency domain was calculated for the baseline condition and the three damage states, aiming to analyze the damage-related variations in the impedance signatures. All numerical calculations were performed using the Numpy package (Harris *et al.*, 2020), and plotting was performed using Matplotlib (Caswell *et al.*, 2023).

In the sequence, all the resistance matrices, each with 30 rows and 511 columns, were merged into a single one by concatenating rows, resulting in a 120-row, 511-column matrix. The first 30 rows were attributed to the healthy condition, the next 30 to Damage 1, and so on. The resistance values were then normalized to an interval between 0 and 1 to avoid undesirable effects caused by outliers once these signatures had been characterized by the presence of peaks, which could lead to significant differences in the magnitude of the values.

A one-row, 511-column vector of labels was created for each condition. The number 0 was assigned to the pristine condition, number 1 to Damage 1, and so on. The vectors containing the labels were merged using the same process employed for the real-part matrices of the impedance, resulting in a 4-row, 511-column matrix.

The resulting dataset was split into two parts, 70,0 % for training the model and 30 % for testing, in order to verify the models' ability to correctly predict the classes, which were all previously known: one baseline condition and three different damage levels. The dataset was partitioned using the Scikit-learn package (Pedregosa *et al.*, 2011).

The training dataset was used to obtain predictive models based on the KM, DT, and RF algorithms of the Scikit-learn package (Pedregosa *et al.*, 2011). In this step, the training data were based on the raw signature dataset of 511 variables.

In determining the KM model, the number of clusters was an important parameter. Based on the Elbow method, this number was set to 4, and it was selected as the first point from which there were no significant changes in the within-cluster sum of squares (WCSS).

The models based on the DT and RF methods were created using entropy, *i.e.*, Equation (1), as a criterion for defining the nodes, as well as based on the $p(X)$ value, which is a fraction of examples belonging to a given class. Entropy is a measure of the purity of a group, wherein low values indicate high homogeneity or a high level of purity, and higher values indicate a lower level of purity.

$$\text{Entropy} = -\sum p(X)\log p(X) \tag{1}$$

The ability of the models to correctly predict the classes was determined by using the test dataset and submitting these values to the model. Accuracy, *i.e.*, Equation (2), was used as a parameter to quantify the capability of each model.

$$\text{Accuracy}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n-1} 1(\hat{y}_i = y_i) \quad (2)$$

Accuracy corresponds to the mean of the correctly predicted classes, where y is the actual class, and $\hat{y}$ is the predicted class (Géron, 2019; Scikit-Learn Team, 2023). This verification is performed along the test vector, which refers to the number of samples evaluated (n). The total number of times that the true values are equal to the corresponding predicted ones is divided by the total number of items in the test vector, resulting in an accuracy value.

The number of variables was decreased using PCA to evaluate the effect of reducing the dimensionality of the dataset on the models' ability to predict the classes correctly.

The PCA was selected over other dimensionality reduction techniques such as multidimensional scaling (MDS), uniform manifold approximation and projection (UMAP), and t-distributed stochastic neighbor embedding (t-SNE), given its simplicity and computational processing capabilities when dealing with data in high dimensions. Additionally, the PCA is less susceptible to the influence of outliers.

Moreover, the PCA method exhibits superior computational efficiency (it is a low complexity, covariance matrix-based linear method) when compared to nonlinear approaches. While the dataset utilized in this study may not be extensive, thereby yielding similar processing times across various dimensionality reduction methodologies, it is crucial to recognize that, within an automated system operating at scale and encompassing diverse structural configurations and sensor arrays over extended durations, this efficiency advantage can yield tangible reductions in energy consumption and decision-making times.

In this sense, the original training and testing data were transformed into principal components and used in the KM, DT, and RF algorithms.

To build the KM model with the number of principal components and clusters, a routine was defined to obtain the optimal combination of parameters to maximize the model's accuracy.

Figure 2 shows the accuracy values for each tested combination. The results show that the highest accuracy value corresponds to 25 principal components and four clusters.

The execution of KM clustering – and the other methods – using the raw data and PCA information resulted in two prediction models, the first of which was based on 511 variables and the other on 25 variables (PCAs), with the latter representing a reduction of 95.1% in the number of features to be analyzed.

To allow for comparisons between the ML methods, the dataset consisting of 25 principal components was used to train the DT and RF models. Entropy was also used as a criterion to define the groups, and the parameters were the same as those used in the other models.

The resulting models were tested using the test set (30,0% of the data). Accuracy was also used as a parameter to quantify the models' ability to correctly predict the classes, and the results were used to make comparisons between the studied approaches.

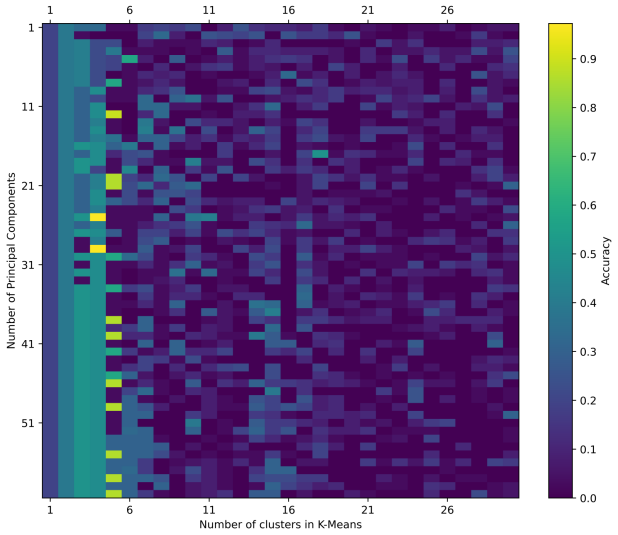

Figure 2. Matrix chart of the accuracy values as a function of the number of principal components and clusters
Source: Authors

Results and discussion

The analysis of the predictions generated by the KM models (Figure 3) reveals a notable discrepancy in performance. Specifically, the KM model relying on raw data exhibits many inaccuracies, as depicted in Figure 3a. This discrepancy is particularly pronounced for the first and second damage scenarios, wherein all predictions were erroneously classified. Conversely, accurate predictions were achieved for the baseline condition and the third damage scenario.

Figure 3 also shows that the same portion of the third damage scenario was misclassified as the second class, which may be due to the characteristics of the database itself – probably the number of samples.

The PCA method's improvement in the classification ability of the KM algorithm can also be observed in the WCSS results (Figure 4). Considering the four clusters used, the WCSS for the PCA-based approach was 48,7% lower than that attributed to the raw data-based KM model.

WCSS values indicate how compact clusters are, wherein high values reflect a high variability within the cluster. In this sense, the lower values obtained by the model based on principal components reflect a better classification.

The results obtained with the DT method (Figure 5) show that the model based on the raw signature database yielded predictions that agreed with the corresponding true values (Figure 5a). Misclassification occurred in the second and third damage scenarios, with only small proportions of incorrect predictions.

Furthermore, misclassification occurred mainly between consecutive damage scenarios, probably due to the small differences between impedance signatures. The exception was the 2.78% of the third damage scenario that was misclassified as Damage 1.

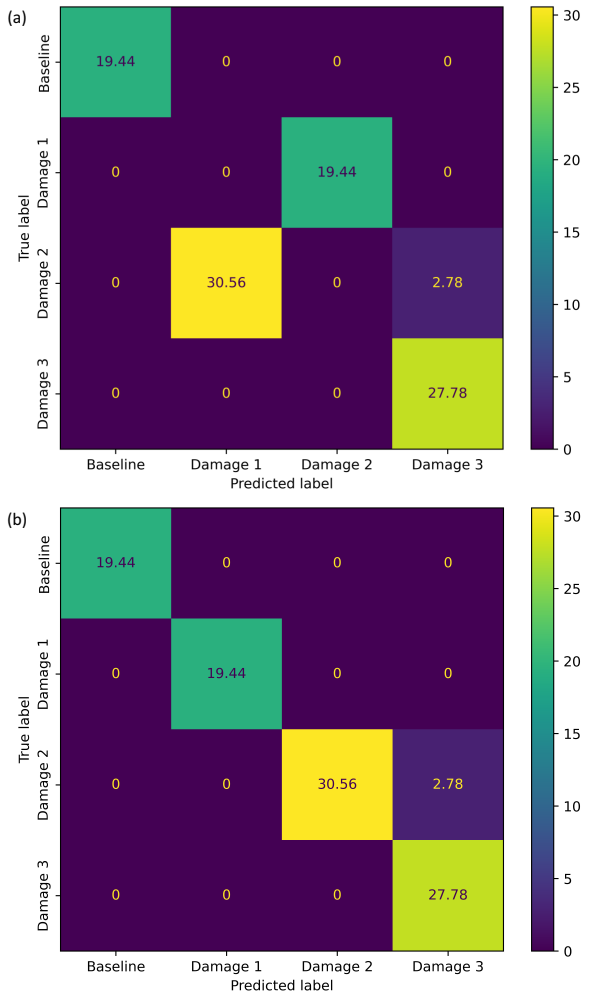

Figure 3. Confusion matrix for the KM clustering models based on raw data (a) and PCA information (b)
Source: Authors

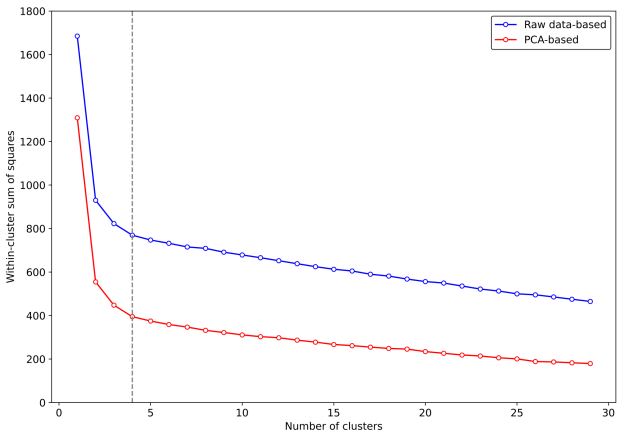

Figure 4. Plot of the WCSS obtained for the KM method using raw data and PCA information
Source: Authors

The results obtained via the DT model based on PCA information (Figure 5b) showed that all predictions matched the corresponding accurate labels, achieving a hit rate of 100%. Therefore, this model could differentiate all damaged scenarios based only on impedance signatures.

When comparing the raw data and the PCA-based results of the DT method, only a slight improvement was observed

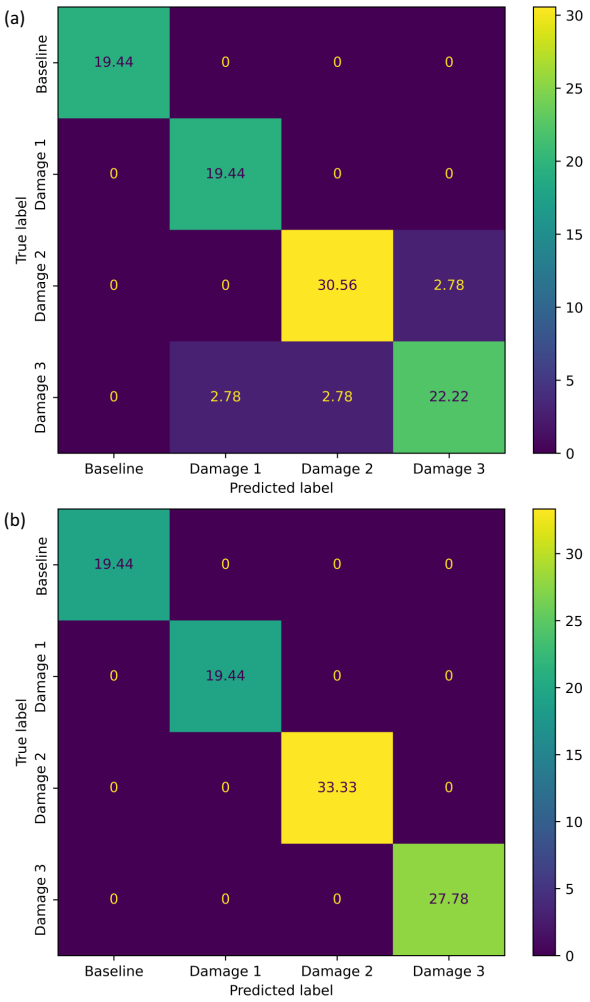

Figure 5. Confusion matrix for the DT models based on raw data (a) and PCA information (b)
Source: Authors

on one occasion. Even in the model based on the raw data, it was possible to obtain reliable results, with little deviations. This indicates that the DT method is a reliable and robust option when classifying structural conditions based on impedance signatures.

The confusion matrices obtained from the RF method are shown in Figure 6. The results show that the model based on raw data yielded predictions with minor deviations from the actual labels (Figure 6a), with only 2.78% of the third damage scenario misclassified.

The results of the PCA-based RF model (Figure 6b) show a slightly higher deviation than that based on raw data, which may be due to the fact that the method is characterized as an ensemble. This feature can be better explored when dealing with large data.

However, the misclassification reported by the PCA-based RF model was only 2.78% higher than the results based on raw data, so the former does not represent significant changes concerning the latter. In this sense, the model based on PCA information can provide reliable results.

The accuracy obtained by each model (Figure 7) shows that KM clustering was the most sensitive to the dimensionality reduction of the dataset. In this case, reducing the number

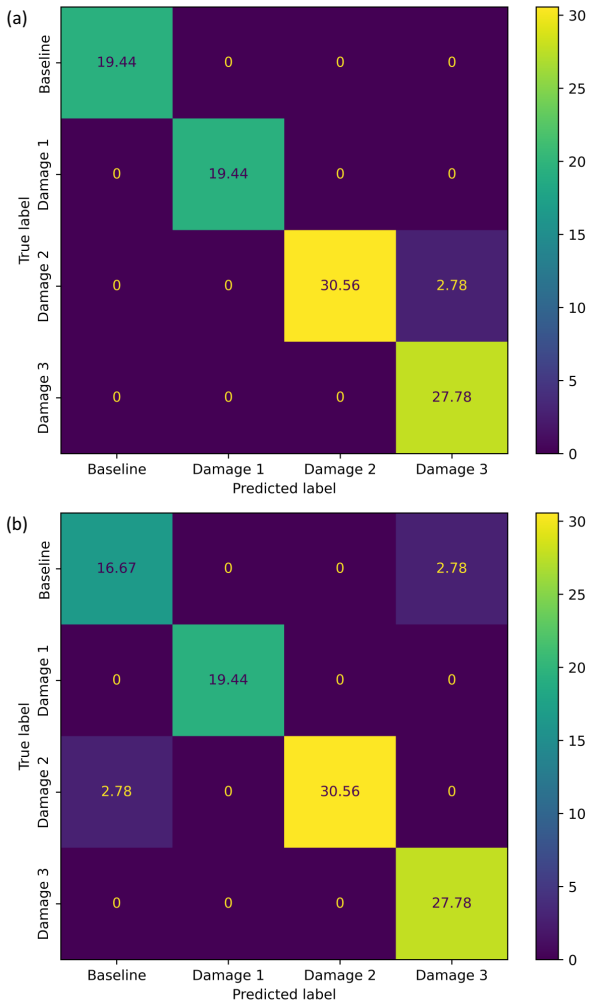

Figure 6. Confusion matrix for the RF models based on raw data (a) and PCA information (b)
Source: Authors

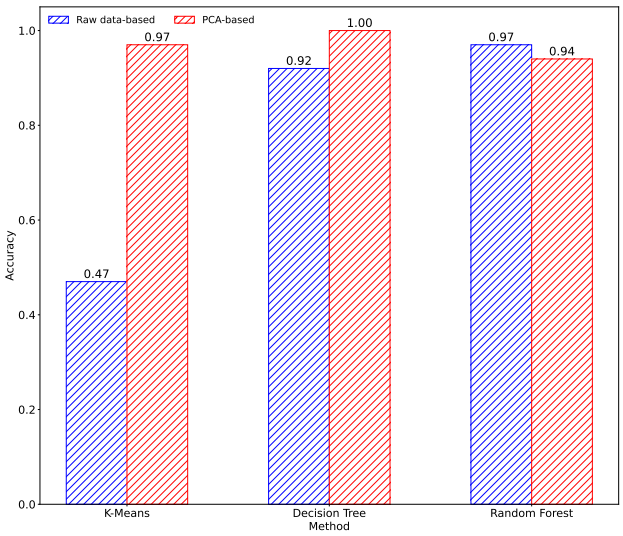

Figure 7. Accuracy results for each method in the two scenarios considered
Source: Authors

of variables from 511 to 25 principal components resulted in a 105.9% increase in accuracy.

On the other hand, the DT method reported an improvement of only 9.1% when comparing the PCA-based results with the raw data-based ones, and the RF method showed a 2.9% decrease in accuracy when using PCA information. This behavior can be explained by the characteristics of the RF method, which can be advantageous in situations with a large number of variables.

When it came to using raw data for model training, the RF method yielded the best results and the KM model the worst. When the PCA data set was used, the DT method performed the best, correctly classifying all classes.

The KM method utilizing PCA compression outperformed the probability neural network (PNN) (Na, 2021), which achieved an accuracy rate of 94.4%. Furthermore, the KM approach exhibited comparable outcomes to the CNN technique (de Rezende *et al.*, 2020), which reported a minimum accuracy rate of 97%.

The DT method showed the highest level of accuracy when using PCA information. It outperformed the PNN (Na, 2021) and CNN (de Rezende *et al.*, 2021; Du *et al.*, 2023) approaches and was comparable to other cutting-edge ML techniques for damage detection (Alazzawi and Wang, 2021; Jiang *et al.* 2021a; Ai and Cheng, 2023). Additionally, the DT method utilizing PCA compression yielded superior results compared to the DT-based approach presented by Jiang *et al.* (2021a), making it a noteworthy strategy for detecting damage using EMI data.

The results obtained with the RF approach were either comparable to or better than those obtained from the PNN method (Na, 2021). Additionally, they closely matched the outcomes of other related studies (de Rezende *et al.*, 2020; Jiang *et al.*, 2021a). Therefore, the RF method is a viable means to detect damage accurately. Moreover, it employs a single-step process and does not necessitate PCA compression to enhance the model's capabilities.

The results obtained herein show that reducing the number of variables generally entails an increase in accuracy, except for the RF method, without invalidating the use of the model.

The evaluation shows that the DT and RF methods are the most appropriate for damage detection in SHM applications using only a dataset of impedance signatures. When using PCA information, the former showed the highest accuracy, and the latter was the most stable, with acceptable accuracy values.

Conclusions

Identifying the presence of damage is a critical task in SHM systems. When using the EMI technique, the information derived from the impedance signature aids in detecting damage. When considering real-world applications, it is necessary to use adaptable techniques to provide reliable results in general situations.

This work used a dataset with impedance signatures of four classes in the KM, DT, and RF methods. In addition, to study the effect of reducing dataset dimensionality, PCA was applied, resulting in two scenarios for each method (*i.e.*, raw data- and PCA-based).

The results show that, in general, reducing the number of variables in the model through PCA caused an increase in the accuracy of the models, with an exception attributed to

the RF model. However, the decrease was not pronounced in this case, and the accuracy obtained was acceptable.

An analysis of the results showed the DT and RF methods to be the most suitable methods for damage detection based only on an impedance signature dataset. The former exhibited the highest accuracy value and good stability, resulting in a close and acceptable accuracy. On the other hand, the KM method managed to provide reliable results only by applying PCA to reduce dimensionality.

In addition, the approaches based on the DT and RF methods yielded results comparable to or better than other state-of-the-art solutions for damage detection using EMI.

According to the results obtained, applying supervised or unsupervised ML techniques for damage detection using EMI yields promising outcomes. Based only on a dataset of impedance signatures, an accurate classification was possible. Furthermore, by using the DT method, it was possible to achieve an accuracy of 100%.

The methods used herein are attractive options when it comes to damage detection using EMI data. Furthermore, the studied approaches do not require any pre-processing step before establishing the models.

In the experimental procedures, damage of the same severity was induced at varying distances from the PZT patch. Other works conducted by the authors have explored scenarios involving defects of different severity levels at the same distance from the patch. This proposal only considered the first scenario to check for feasibility, but different ML model responses for discerning different damage levels can be further explored.

Additionally, future research may focus on considering scenarios with different temperatures to enhance the use of unsupervised techniques for damage detection, bringing the approaches used herein closer to shop-floor industrial applications.

Acknowledgements

P. E. C Pereira would like to express his gratitude to the Federal University of Catalão for the provision of the license to pursue doctoral studies at the Federal University of Uberlândia.

CRediT author statement

P. E. C. Pereira: formal analysis, methodology, writing (original draft). S. W. F. de Rezende: data curation, writing (review and editing). B. P. Barella: data curation. J. R. V. de Moura Jr.: conceptualization, methodology, writing (review and editing). R. M. Finzi Neto: supervision, writing (review and editing).

Conflicts of interest

The authors declare no competing interests relevant to the content of this article.

References

Ai, D., and Cheng, J. (2023). A deep learning approach for electromechanical impedance-based concrete structural damage quantification using a two-dimensional convolutional neural network. *Mechanical Systems and Signal Processing, 183*, 109634. https://doi.org/10.1016/j.ymssp.2022.109634

Ai, D., Mo, F., Yang, F., and Zhu, H. (2022). Electromechanical impedance-based concrete structural damage detection using principal component analysis incorporated with a neural network. *Journal of Intelligent Material Systems and Structures, 33*(17), 2241–2256. https://doi.org/10.1177/1045389X221077440

Alazzawi, O., and Wang, D. (2021). Damage identification using the PZT impedance signals and residual learning algorithm. *Journal of Civil Structural Health Monitoring, 11*, 1225–1238. https://doi.org/10.1007/s13349-021-00505-9

Albakri, M. I., and Tarazaga, P. A. (2017). Dynamic analysis of a piezoelectric augmented beam system with adhesive bonding layer effects. *Journal of Intelligent Materials Systems and Structures, 28*(2), 178–194. https://doi.org/10.1177/1045389X16648426

Alelyani, S., Tang, J., and Liu, H. (2014). Feature Selection for Clustering: A Review. In C. C. Aggarwal and C. K. Reddy (Eds.), *Data Clustering: Algorithms and Applications* (pp. 29–60). CRC Press.

Bari, Sk. A., and Moharana, S. (2024). A novel approach to alloy-based bonding for piezoelectric sensor integrated in impedance-based structural health monitoring. *Journal of Asian Architecture and Building Engineering, 23*(3), 898–920. https://doi.org/10.1080/13467581.2023.2257279

Barros, B., Conde, B., Cabaleiro, M., and Riveiro, B. (2023). Design and testing of a decision tree algorithm for early failure detection in steel truss bridges. *Engineering Structures, 289*, 116243. https://doi.org/10.1016/j.engstruct.2023.116243

Bergmayr, T., Holl, S., Kralovec, C., and Schagerl, M. (2023). Local residual random forest classifier for strain-based damage detection and localization in aerospace sandwich structures. *Composite Structures, 304*(1), 116331. https://doi.org/10.1016/j.compstruct.2022.116331

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Biswal, T., and Parida, S. K. (2022). A novel high impedance fault detection in the micro-grid system by the summation of accumulated difference of residual voltage method and fault event classification using discrete wavelet transforms and a decision tree approach. *Electric Power Systems Research, 209*, 108042. https://doi.org/10.1016/j.epsr.2022.108042

Caswell, T. A., Sales de Andrade, E., Lee, A., Droettboom, M., Hoffmann, T., Klymak, J., Hunter, J., Firing, E., Stansby, D., Varoquaux, N., Nielsen, J. H., Gustafsson, O., Root, B., May, R., Sunden, K., Elson, P., Seppanen, J. K., Lee, J.-J., Dale, D., ... Silvester, S. (2023). *matplotlib/matplotlib: REL: v3.7.3* (v3.7.3). Zenodo. https://doi.org/10.5281/zenodo.8336761

Chen, D., Montano, V., Huo, L., and Song, G. (2020). Depth detection of subsurface voids in concrete-filled steel tubular (CFST) structure using percussion

and decision tree. *Measurement, 163*, 107869. https://doi.org/10.1016/j.measurement.2020.107869

Chen, W., Zhang, S., Li, R., and Shahabi, H. (2018). Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naıve Bayes tree for landslide susceptibility modeling. *Science of The Total Environment, 644*, 1006–1018. https://doi.org/10.1016/j.scitotenv.2018.06.389

Chencho, J., Li, H., Hao, R., Wang, L., and Li, L. (2021). Development and application of random forest technique for element level structural damage quantification. *Structural Control and Health Monitoring, 28*(3), e2678. https://doi.org/10.1002/stc.2678

Davis, J. C. (2002). *Statistics and Data Analysis in Geology* (3rd ed.). John Wiley & Sons.

de Rezende, S. W. F., de Moura Jr., J. R. V., Finzi Neto, R. M., Gallo, C. A., and Steffen Jr., V. (2020). Convolutional neural network and impedance-based SHM applied to damage detection. *Engineering Research Express, 2*(3), Article 035031. https://doi.org/10.1088/2631-8695/abb568

Dinh, T. P., Pham-Quoc, C., Thinh, T. N., Nguyen, B. K. D., and Kha, P. C. (2023). A flexible and efficient FPGA-based random forest architecture for IoT applications. *Internet of Things, 22*, 100813. https://doi.org/10.1016/j.iot.2023.100813

Djemana, M., Hrairi, M., and Al Jeroudi, Y. (2017). Using Electromechanical impedance and extreme learning machine to detect and locate damage in structures. *Journal of Nondestructive Evaluation, 36*, 39. https://doi.org/10.1007/s10921-017-0417-5

Du, F., Wu, S., Xu, C., Yang, Z., and Su, Z. (2023). Electromechanical Impedance Temperature Compensation and Bolt Loosening Monitoring Based on Modified Unet and Multitask Learning. *IEEE Sensors Journal, 23*(5), 4556–4567. https://doi.org/10.1109/JSEN.2021.3132943

Fan, X., and Li, J. (2020). Damage identification in plate structures using sparse regularization based electromechanical impedance technique. *Sensors, 20*(24), 7069. https://doi.org/10.3390/s20247069

Fletcher, S., and Md. Islam, Z. (2019). Decision tree classification with differential privacy: A survey. *ACM Computing Surveys, 52*(4), 83. https://doi.org/10.1145/3337064

Géron, A. (2019). *Mãos à obra aprendizado de máquina com Scikit-Learn and TensorFlow: Conceitos, ferramentas e técnicas para a construcão de sistemas inteligentes*. Alta Books.

Giurgiutiu, V. (2014). *Structural health monitoring with piezoelectric wafer active sensors* (2nd ed.). Academic Press. https://doi.org/10.1016/C2013-0-00155-7

Gonçalves, D. R., de Moura Jr., J. R. V., Pereira, P. E. C., Mendes, M. V. A., and Diniz-Pinto, H. S. (2021). Indicator kriging for damage position prediction by the use of electromechanical impedance-based structural health monitoring. *CR Mécanique, 349*(2), 225–240. https://doi.org/10.5802/crmeca.81

Hamzeloo, S. R., Barzegar, M., and Mohsenzadeh, M. (2020). Damage detection of L-shaped beam structure with a crack by electromechanical impedance response: Analytical approach and experimental validation. *Journal of Nondestructive Evaluation, 39*(2), 47. https://doi.org/10.1007/s10921-020-00692-3

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature, 585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2

Harrison, M. (2020). *Machine learning: Guia de referência rápida*. Novatec Editora.

Hong, H., Liu, J., Bui, D. T., Pradhan, B., Acharya, T. D., Pham, B. T., Zhu, A.-X., Chen, W., and Ahmad, B. B. (2018). Landslide susceptibility mapping using J48 Decision Tree with AdaBoost, bagging and Rotation Forest ensembles in the Guangchang area (China). *CATENA, 163*, 399–413. https://doi.org/10.1016/j.catena.2018.01.005

Islam, M. M., and Huang, H. (2014). Understanding the effects of adhesive layer on the electromechanical impedance (EMI) of bonded piezoelectric wafer transducer. *Smart Materials and Structures, 23*(12), 125037. https://doi.org/10.1088/0964-1726/23/12/125037

Jegadeeshwaran, R., and Sugumaran, V. (2013). Comparative study of decision tree classifier and best first tree classifier for fault diagnosis of automobile hydraulic brake system using statistical features. *Measurement, 46*(9), 3247–3260. https://doi.org/10.1016/j.measurement.2013.04.068

Jiang, X., Zhang, X., Tang, T., and Zhang, Y. (2021c). Electromechanical impedance-based self-diagnosis of piezoelectric smart structure using principal component analysis and LibSVM. *Scientific Reports, 11*, 11345. https://doi.org/10.1038/s41598-021-90567-y

Jiang, X., Zhang, X., and Zhang, Y. (2021a). Establishment and optimization of sensor fault identification model based on classification and regression tree and particle swarm optimization. *Materials Research Express, 8*(8), 085703. https://doi.org/10.1088/2053-1591/ac1cae

Jiang, X., Zhang, X., and Zhang, Y. (2021b). Piezoelectric active sensor self-diagnosis for electromechanical impedance monitoring using K-means clustering analysis and artificial neural network. *Shock and Vibration, 2021*, 5574898. https://doi.org/10.1155/2021/5574898

Kim, J., and Wang, K.-W. (2019). Electromechanical impedance-based damage identification enhancement using bistable and adaptive piezoelectric circuitry. *Structural Health Monitoring, 18*(4), 1268–1281. https://doi.org/10.1177/1475921718794202

Kim, S. Y., and Upneja, A. (2014). Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models. *Economic Modelling, 36*, 354–362. https://doi.org/10.1016/j.econmod.2013.10.005

Koza, J. R. (1990). Concept formation and decision tree induction using the genetic programming paradigm. In H.-P. Schwefel and R. Manner (Eds.), *PPSN I: Proceedings of the 1st Workshop on Parallel Problem Solving from Nature* (pp. 124{-128). Springer-Verlag.

Li, H., Ai, D., Zhu, H., and Luo, H. (2021). Integrated electromechanical impedance technique with convolutional neural network for concrete structural damage quantification under varied temperatures. *Mechanical Systems and Signal Processing, 152*, 107467. https://doi.org/10.1016/j.ymssp.2020.107467

Lim, D. K., Mustapha, K. B., and Pagwiwoko, C. P. (2021). Delamination detection in composite plates using random forests. *Composites Structures, 278*, 114676. https://doi.org/10.1016/j.compstruct.2021.114676

Liu, J., and Han, J. (2014). Spectral clustering. In C. C. Aggarwal and C. K. Reddy (Eds.), *Data Clustering: Algorithms and Applications* (pp. 177{-200). CRC Press.

Liu, R., Li, S., Zhang, G., and Jin, W. (2021). Depth detection of void defect in sandwich-structured immersed tunnel using elastic wave and decision tree. *Construction and Building Materials, 305*, 124756. https://doi.org/10.1016/j.conbuildmat.2021.124756

Loyola-González, O., Ramírez-Sáyago, E., and Medina-Pérez, M. A. (2023). Towards improving decision tree induction by combining split evaluation measures. *Knowledge-Based Systems, 277*, 110832. https://doi.org/10.1016/j.knosys.2023.110832

Malinowski, P. H., Wandowski, T., and Singh, S. K. (2021). Employing principal component analysis for assessment of damage in GFRP composites using electromechanical impedance. *Composite Structures, 266*, 113820. https://doi.org/10.1016/j.compstruct.2021.113820

Martowicz, A., and Rosiek, M. (2013). Electromechanical impedance method. In T. Stepinski, T. Uhl, and W. Staszewski (Eds.). *Advanced Structural Damage Detection: From Theory to Engineering Applications* (ch. 6, pp. 141–176). John Wiley & Sons. https://doi.org/10.1002/9781118536148.ch6

Mayer, C. (2020). *Python one-liners: Write concise, eloquent Python like a professional*. No Starch Press, Inc.

Meher, U., Mishra, S. K., and Sunny, M. R. (2022). Impedance-based looseness detection of bolted joints using artificial neural network: An experimental study. *Structural Control and Health Monitoring, 29*(10), e3049. https://doi.org/10.1002/stc.3049

Na, S., and Lee, H. K. (2013). A multi-sensing electromechanical impedance method for non-destructive evaluation of metallic structures. *Smart Materials and Structures, 22*(9), 095011. https://doi.org/10.1088/0964-1726/22/9/095011

Na, W. S. (2021). Bolt loosening detection using impedance-based non-destructive method and probabilistic neural network technique with minimal training data. *Engineering Structures, 226*, 111228. https://doi.org/10.1016/j.engstruct.2020.111228

Ning, F., Cheng, Z., Meng, D., and Wei, J. (2021). A framework combining acoustic features extraction method and random forest algorithm for gas pipeline leak detection and classification. *Applied Acoustics, 182*, 108255. https://doi.org/10.1016/j.apacoust.2021.108255

Nomelini, Q. S. S., da Silva, J. V., Gallo, C. A., Finzi Neto, R. M., Tsuruta, K. M., and de Moura Jr., J. R. V. (2020). Non-parametric inference applied to damage detection in the electromechanical impedance-based health monitoring. *International Journal of Advanced Engineering Research and Science, 7*(9), 73–79. https://doi.org/10.22161/ijaers.79.9

Oliveira, M. A. de, Monteiro, A. V., and Vieira Filho, J. (2018a). A new structural health monitoring strategy based on PZT sensors and convolutional neural network. *Sensors, 18*(9), 2955. https://doi.org/10.3390/s18092955

Parida, L., Moharana, S., and Giri, S. K. (2023). Machine learning approach for predicting impedance signatures of construction steel structures in various tensile pull action. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2023.03.741

Park, G., Sohn, H., Farrar, C. R., and Inman, D. J. (2003). Overview of piezoelectric impedance-based health monitoring and path forward. *Shock and Vibration Digest, 35*(6), 451–463. https://doi.org/10.1177/05831024030356001

Park, S., Lee, J.-J., Yun, C.-B., and Inman, D. J. (2008). Electro-mechanical impedance-based wireless structural health monitoring using PCA-data compression and K-means clustering algorithms. *Journal of Intelligent Material Systems and Structures, 19*(4), 509–520. https://doi.org/10.1177/1045389X07077400

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830

Perera, R., Torres, L., Ruiz, A., Barris, C., and Baena, M. (2019). An EMI-based clustering for structural health monitoring of NSM FRP strengthening systems. *Sensors, 19*(17), 3775. https://doi.org/10.3390/s19173775

Reddy, C. K., and Vinzamuri, B. (2014). A survey of partitional and hierarchical clustering algorithms. In C. C. Aggarwal and C. K. Reddy (Eds.). *Data Clustering: Algorithms and Applications* (pp. 87–110). CRC Press

Rokach, L., and Maimon, O. (2009). Classification trees. In L. Rokach and O. Maimon (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 165–192). Springer. https://doi.org/10.1007/978-0-387-09823-4_9

Scikit-Learn Team (2023). *Model selection and evaluation*. https://scikit-learn.org/stable/model_selection.html

Simeone, O. (2018). A brief introduction to machine learning for engineers. *Foundations and Trends® in Signal Processing, 12*(3-4), 200-431. https://doi.org/10.1561/2000000102

Sun, F. P., Chaudhry, Z., Liang, C., and Rogers, C. A. (1995). Truss structure integrity identification

using PZT sensor-actuator. *Journal of Intelligent Materials Systems and Structures, 6*(1), 134–139. https://doi.org/10.1177/1045389X9500600117

Tang, X., Gu, X., Rao, L., and Lu, J. (2021). A single fault detection method of gearbox based on random forest hybrid classifier and improved Dempster-Shafer information fusion. *Computers and Electrical Engineering, 92*, 107101. https://doi.org/10.1016/j.compeleceng.2021.107101

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., ... SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, L., Yuan, B., Xu, Z., and Sun, Q. (2022). Synchronous detection of bolts looseness position and degree based on fusing electro-mechanical impedance. *Mechanical Systems and Signal Processing, 174*, 109068. https://doi.org/10.1016/j.ymssp.2022.109068

Yan, Q., Liao, X., Zhang, C., Zhang, Y., Luo, S., and Zhang, D. (2022). Intelligent monitoring and assessment on early-age hydration and setting of cement mortar through an EMI-integrated neural network. *Measurement, 203*, 111984. https://doi.org/10.1016/j.measurement.2022.111984

Zhou, L., Chen, S.-X., Ni, Y.-Q., and Choy, A. W.-H. (2021). EMI-GCN: A hybrid model for real-time monitoring of multiple bolt looseness using electromechanical impedance and graph convolutional networks. *Smart Materials and Structures, 30*(3), 035032. https://doi.org/10.1088/1361-665X/abe292