BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 71(3), 2023, Article number: e144606
DOI: 10.24425/bpasts.2023.144606

SPECIAL SECTION

Ensemble of feature extraction methods to improve the structural damage classification in a wind turbine foundation

Jersson X. LEON-MEDINA [1,2]*, **Núria PARÉS** [3], **Maribel ANAYA** [4], **Diego A. TIBADUIZA** [4], and **Francesc POZO** [1,5]

[1] Control, Data, and Artificial Intelligence (CoDAlab), Department of Mathematics, Escola d'Enginyeria de Barcelona Est (EEBE), Campus Diagonal-Besòs (CDB), Universitat Politècnica de Catalunya (UPC), Eduard Maristany 16, 08019 Barcelona, Spain
[2] Programa de Ingeniería Mecatrónica, Universidad de San Buenaventura, Carrera 8H #172-20, Bogota, Colombia
[3] Laboratori de Càlcul Numèric (LaCàN), Department of Mathematics, Escola d'Enginyeria de Barcelona Est (EEBE), Campus Diagonal-Besòs (CDB), Universitat Politècnica de Catalunya (UPC), Eduard Maristany 16, 08019 Barcelona, Spain
[4] Departamento de Ingeniería Eléctrica y Electrónica, Universidad Nacional de Colombia, Cra 45 No. 26-85, Bogotá 111321, Colombia
[5] Institute of Mathematics (IMTech), Universitat Politècnica de Catalunya (UPC), Pau Gargallo 14, 08028 Barcelona, Spain

Abstract. The condition monitoring of offshore wind power plants is an important topic that remains open. This monitoring aims to lower the maintenance cost of these plants. One of the main components of the wind power plant is the wind turbine foundation. This study describes a data-driven structural damage classification methodology applied in a wind turbine foundation. A vibration response was captured in the structure using an accelerometer network. After arranging the obtained data, a feature vector of 58 008 features was obtained. An ensemble approach of feature extraction methods was applied to obtain a new set of features. Principal Component Analysis (PCA) and Laplacian eigenmaps were used as dimensionality reduction methods, each one separately. The union of these new features is used to create a reduced feature matrix. The reduced feature matrix is used as input to train an Extreme Gradient Boosting (XGBoost) machine learning-based classification model. Four different damage scenarios were applied in the structure. Therefore, considering the healthy structure, there were 5 classes in total that were correctly classified. Five-fold cross validation is used to obtain a final classification accuracy. As a result, 100% of classification accuracy was obtained after applying the developed damage classification methodology in a wind-turbine offshore jacket-type foundation benchmark structure.

Key words: structural health monitoring; wind turbine foundation; damage classification; machine learning; feature extraction; XGBoost.

1. INTRODUCTION

Structural health monitoring (SHM) of wind turbine foundations can consider the effects that marine waves and wind exert on the structure. The environmental and operational conditions to which these wind turbines are subjected can be extreme [1]. The approaches of using guided waves that propagate through the structure and classifying the damage that it may have compared to the healthy structure do not consider the effects of environmental conditions [2]. Therefore, an approach that uses the vibration response of the structure has been successfully used to classify damage to jacket-type wind turbine foundations [3]. To accomplish this, a shaker excites the structure and a series of accelerometers capture the vibration response of the structure. To perform the processing of the large amount of data obtained in the structure, different machine learning methods have been used, such as Support Vector Machines [4], Convolutional Neural Networks [5], k-Nearest Neighbors(k-NN) [6], Autoencoder Neural Network [7], Siamese Neural Networks [8], among others.

Some studies in the literature are found related to combine principal component analysis (PCA) with t-distributed stochastic neighbor embedding (t-SNE) to perform classification of structural changes in an aluminum plate using a piezoelectric sensor network and lamb waves. One approach was based on transforming the acquired signals into the frequency domain using the Fast Fourier Transform (FFT) algorithm [9]. A similar approach is used in the time domain [10] reaching high values of classification accuracy.

This study shows the development of a damage classification methodology in jacket-type wind turbine foundations that is based on using an ensemble of feature extraction methods to reduce the high dimensionality of the acquired data and improve the final average classification accuracy. The novelty of this study is twofold. On one hand, with respect to data normalization, it is worth noting the application of the so-called *mean-centered unitary group-scaling method* (MCUGS), instead of the more standard approaches, such as the z-score normalization. On the other hand, the input of the classifier is the combination of a linear (principal component analysis, PCA) and

*e-mail: jersson.xavier.leon@upc.edu

Manuscript submitted 2022-09-07, revised 2022-11-29, initially accepted for publication 2022-12-01, published in June 2023.

a non-linear (Laplacian eigenmaps) feature extraction strategies. The combination of these two approaches is genuinely enriching: the transformation due to PCA maximizes the variability of the original data while the Laplacian eigenmaps provide locality-preserving properties and a natural connection to clustering. The remainder of this paper is described as follows: Section 2 shows the experimental setup. Next, Section 3 describes the damage classification methodology. Following, Section 4 illustrates the results and finally, Section 5 outlines the main conclusions of this study.

2. EXPERIMENTAL SETUP

A data-driven approach was performed in a laboratory-scaled jacket-type wind turbine foundation. This laboratory-scaled structure is composed of three parts from bottom to top: the jacket, the tower and the nacelle. The total height of the structure is 2.7 m. The jacket was made of steel bars and bolts. Figure 1 illustrates the structure. The experimental setup includes:

- An arbitrary function generator (GW INSTEK AFG-2005), to apply a white noise signal to the structure.
- The white noise signal is amplified and applied to an inertial shaker (GW-IV47 from Data Physics).
- This signal produces a vibration response that is captured using eight triaxial accelerometers (PCB Piezotronics, Model 356A17).
- The data acquisition process was performed using a cDAQ-9188 chassis from National Instruments and six NI-9234 modules, each of which has four channels.

The experimental procedure is detailed as follows: first an arbitrary function generator inserts a white noise signal. This signal is amplified with factors 0.5, 1, 2 and 3 and it is applied through an inertial shaker placed in the nacelle that produces vibrations in the structure. The vibration-response-only generated by the shaker is acquired by a set of 8 accelerometers, bonded to the structure using petro wax (PCB Piezotronics, model 080A109). The location of the accelerometers shown in Fig. 1a was strategically defined according to a modal study [3]. Eight triaxial accelerometers capture the data, thus 24 sensors in total acquired signals with a sampling frequency of 275 Hz during 8.789 s obtaining 2417 data points per sensor. These 24 signals are arranged one after the other following an unfolding procedure to finally obtain a feature vector in every experiment of 58 008 features. This high dimensionality must be treated to reduce the size of the feature matrix.

Figure 2 shows random samples of data collected before normalization. More precisely, two samples associated with the undamaged structure with minimum and maximum amplitudes are shown in Fig. 2.

The jacket-type wind turbine foundation was made of steel angle bars, also known as *L*-shaped profiles. In this work, a 5 mm crack is performed at four different bars of the jacket,

(a) Sensor location in the structure

(b) Introduced damage in different links of the structure

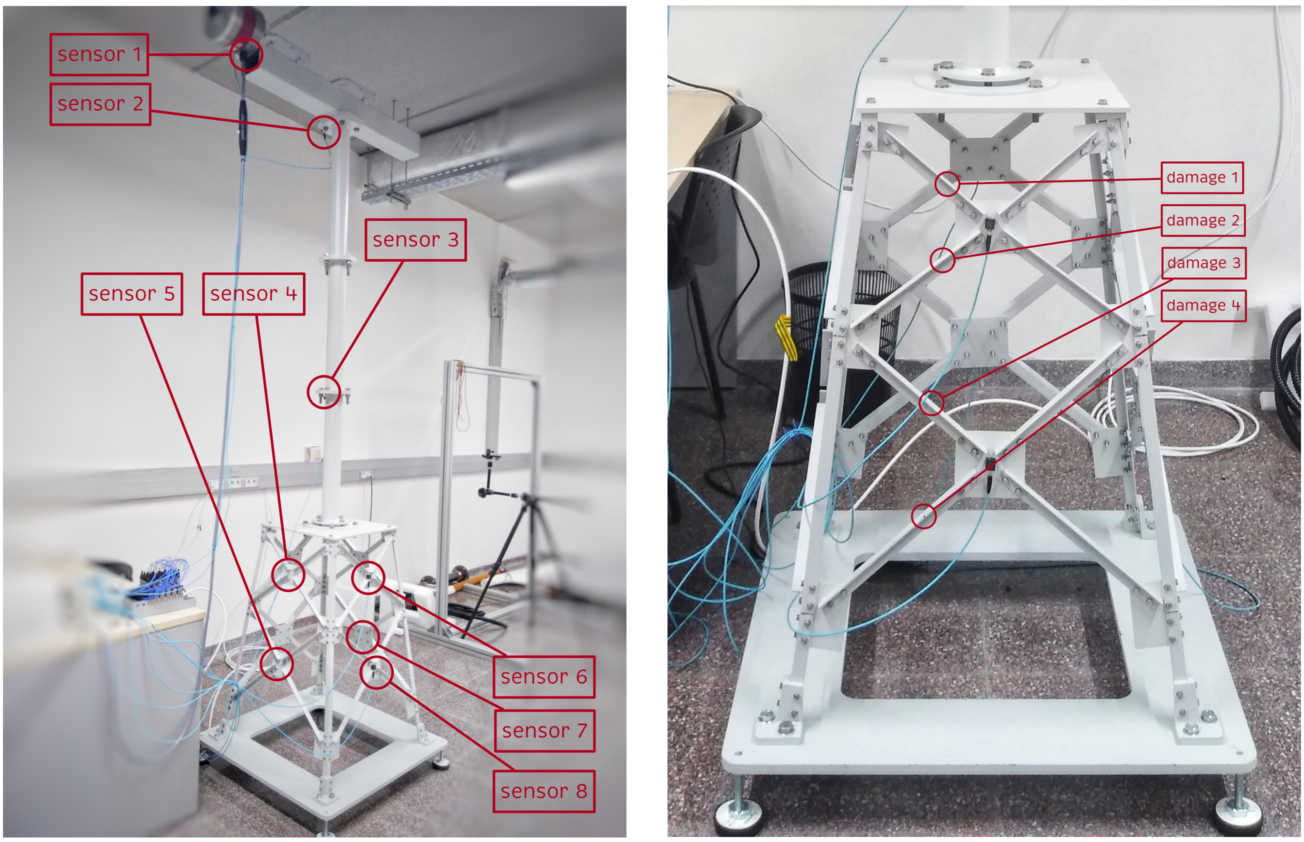

Fig. 1. Small-scale wind-turbine foundation structure

(a) Raw signal for an undamaged sample using an amplitude of 0.5

(b) Raw signal for an undamaged sample using an amplitude of 3

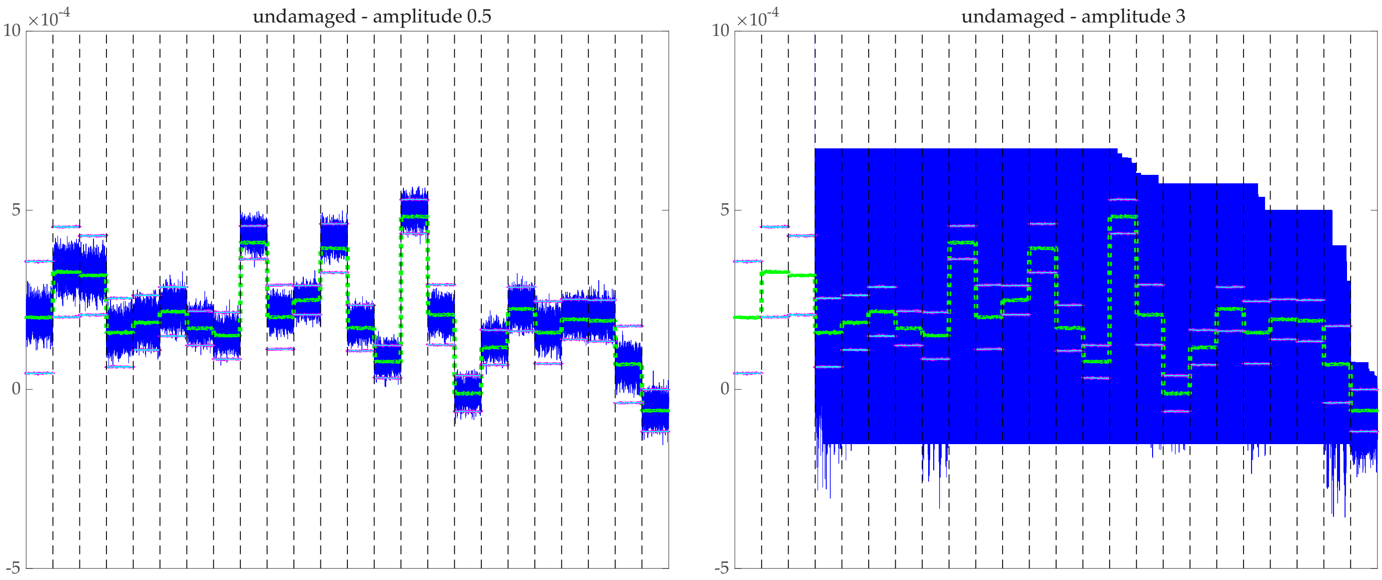

Fig. 2. Raw signals [3] (blue) are shown along with the mean values μ_j^k (green) and the representation of the standard deviations $\mu_j^k \pm \sigma_{\mathrm{MCUGS}}^k$ (cyan) and $\mu_j^k \pm \sigma_{\mathrm{MCGS}}^k$ (magenta). Dashed vertical lines separate the measures of the 24 sensors

one at a time. Figure 1b shows the jacket-type wind turbine foundation structure with the four different types of damage, applied one by one.

Experiments were performed following a number of different measurements for the healthy and damage structures. Particularly, 2460 measurements were obtained for the healthy structure and 820 measurements for every one of the four types of damages applied to a link of the wind turbine foundation.

3. DAMAGE CLASSIFICATION METHODOLOGY

A feature extraction-based approach using manifold learning algorithms is developed to improve the classification accuracy in a wind turbine foundation structural damage classification problem. The developed signal-processing methodology is composed of four stages: data unfolding, normalization using the mean centered unitary group scaling method (MCUGS) [11], ensemble feature extraction and classification through a XGBoost [12] model. This study aims to use an ensemble of feature extraction algorithms, which include principal component analysis (PCA) [13] as linear feature extraction method and Laplacian eigenmaps [14] as nonlinear manifold learning algorithm. Figure 3 illustrates the stages of the developed structural damage classification methodology

A sensitivity study of the k parameter of the Laplacian Eigenmaps algorithm is performed. A data set of five different structural states in a wind turbine foundation is used to validate the proposed methodology. A 5-fold cross-validation procedure was applied to obtain the final confusion matrix in the classification problem. The final confusion matrix obtained after applying the ensemble feature extraction procedure represents a classification accuracy of 100%.

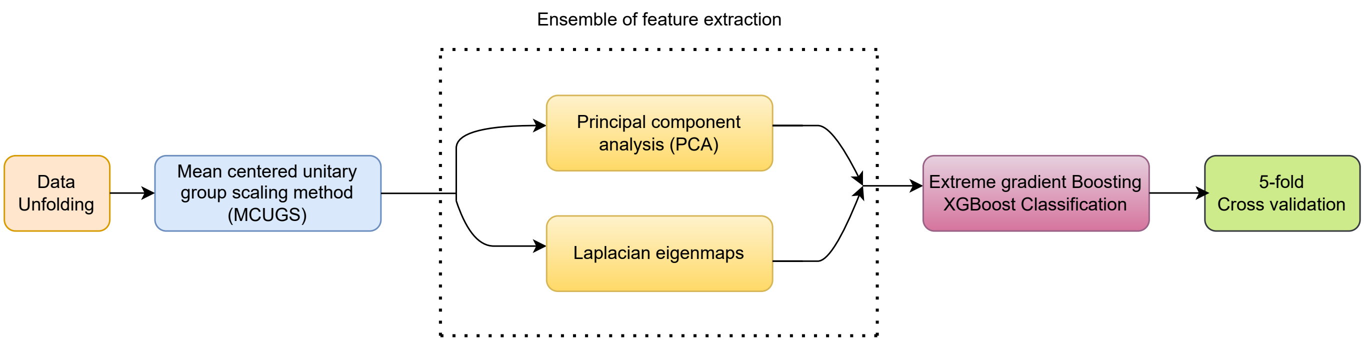

Fig. 3. Stages of the developed structural damage classification methodology

4. RESULTS

4.1. Classification performance measure

The average classification accuracy is selected as performance measure in this study. The following equation represents its calculation by each one of the 5 classes – tp: true positives, tn: true negatives, fn: false negatives and fp: false positives in the confusion matrix

$$\text{accuracy}_i = \frac{tp_i + tn_i}{tp_i + tn_i + fn_i + fp_i}.$$

The final macro average is calculated for the accuracy as follows:

$$\text{accuracy} = \frac{1}{l}\sum_{i=1}^{l}\text{accuracy}_i.$$

4.2. XGBoost parameters

In this study, a XGBoost classifier is used as a machine learning algorithm to classify the different classes in the wind turbine foundation dataset. The hyperparameters of the XGBoost classifier were tuned using a GridSearchCV function. The selected parameters in the XGBoost classifier algorithm are described in Table 1.

Table 1
Parameters for the XGBoost classifier method

Parameter	Value
learning_rate	0.1
n_estimators	1000
max_depth	8
min_child_weight	1
gamma	0
subsample	0.8
colsample_bytree	0.8
objective	multi:softprob
nthread	4
scale_pos_weight	1
seed	27

4.3. PCA results

The results obtained when using PCA as the feature extraction method are shown below in Fig. 4. The first 8 principal components were selected to create a 5740 × 8 size feature matrix.

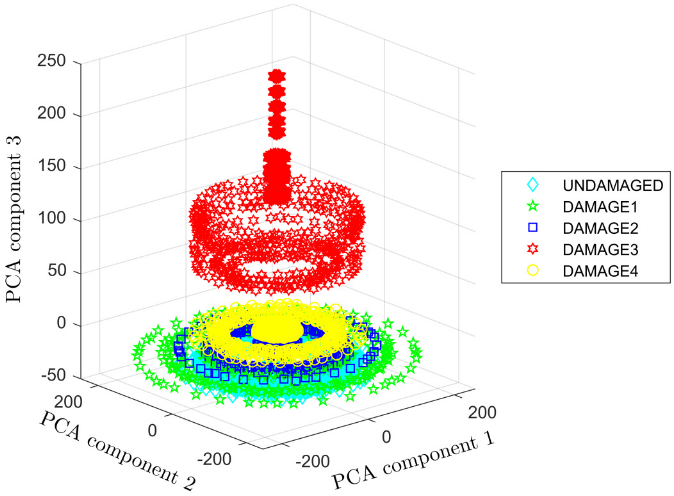

Fig. 4. 3D embedding obtained from PCA method, first three components are plotted

The resulting average classification accuracy obtained after applying the combination of the PCA method with XGBoost was 0.9992. Before the PCA method, the MCUGS method was used to perform the scaling. The confusion matrix that expresses the average accuracy value of 0.9992 obtained by using the PCA and XGBoost methods is shown in Table 2. From Table 2 is evident the perfect classification of DAMAGE 3 and 4 classes, while there were a few mistakes for the other classes as the UNDAMAGED class had 3 mistakes, the DAMAGE 1 class had 5 mistakes and finally, the DAMAGE 2 class had 3 mistakes.

4.4. Laplacian eigenmaps eesults

Table 3 displays the average classification accuracy results for the 5740 × 8 size matrix obtained after applying the Laplacian eigenmaps dimensionality reduction method. The values were obtained by varying the parameter k proper to the Laplacian eigenmaps method. It proves that as k increases the accuracy value increases up to a maximum of 0.9995 when $k = 100$.

The best results obtained when using Laplacian eigenmaps ($k = 100$) as a feature extraction algorithm are shown below in Fig. 5. Figure 5 clearly shows the separation between classes and interclass grouping. The output dimensionality after non-

Table 2
Confusion matrix obtained with the methods PCA + XGBoost, average accuracy = 0.9992

		Predicted Classl				
		UNDAMAGED	DAMAGE 1	DAMAGE 2	DAMAGE 3	DAMAGE 4
Actual Class	UNDAMAGED	2457	3	0	0	0
	DAMAGE 1	5	815	0	0	0
	DAMAGE 2	0	3	817	0	0
	DAMAGE 3	0	0	0	820	0
	DAMAGE 4	0	0	0	0	820

Table 3
Average accuracy results obtained after varying the k parameter of the Laplacian eigenmaps (LE) method when constructing its neighborhood graph

k neighbors in LE	Average accuracy
20	0.9615
30	0.9755
40	0.9894
50	0.9911
60	0.9935
70	0.9917
80	0.9954
90	0.9982
100	0.9995

linear feature extraction was fixed to 8 dimensions to create a feature matrix of size 5740×8.

The confusion matrix that expresses the average accuracy value of 0.9995 obtained by using the Laplacian eigenmaps ($k = 100$) and XGBoost methods with the MCUGS scaling is shown in Table 4. The perfect classification of DAMAGE 1 and 2 classes is evident from Table 4, while there were a few mistakes for the other classes as the UNDAMAGED class had 2 mistakes, the DAMAGE 3 class had 2 mistakes and finally, the DAMAGE 4 class had 3 mistakes.

It is worth mentioning that for large values of k in the calculation of the neighborhood graph in the Laplacian eigenmaps method, results are obtained that are close to those obtained by the PCA method. As an example to illustrate the above, in Fig. 6 the results of the first three dimensions of the embedding produced by the Laplacian eigenmaps method are shown when

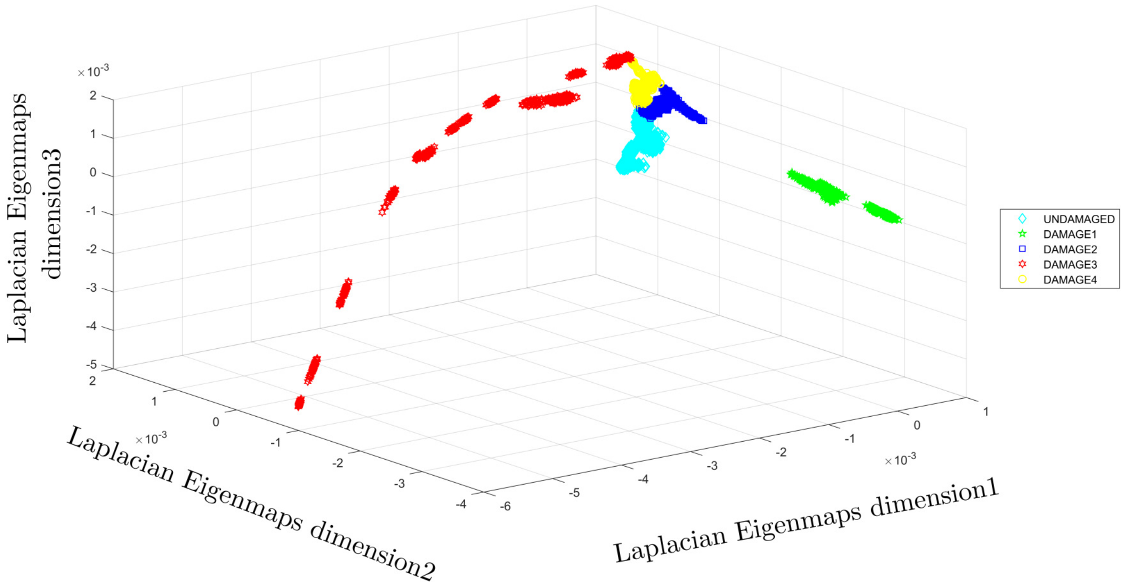

Fig. 5. 3D Embedding obtained from the Laplacian eigenmaps method, $k = 100$, first three dimensions are plotted

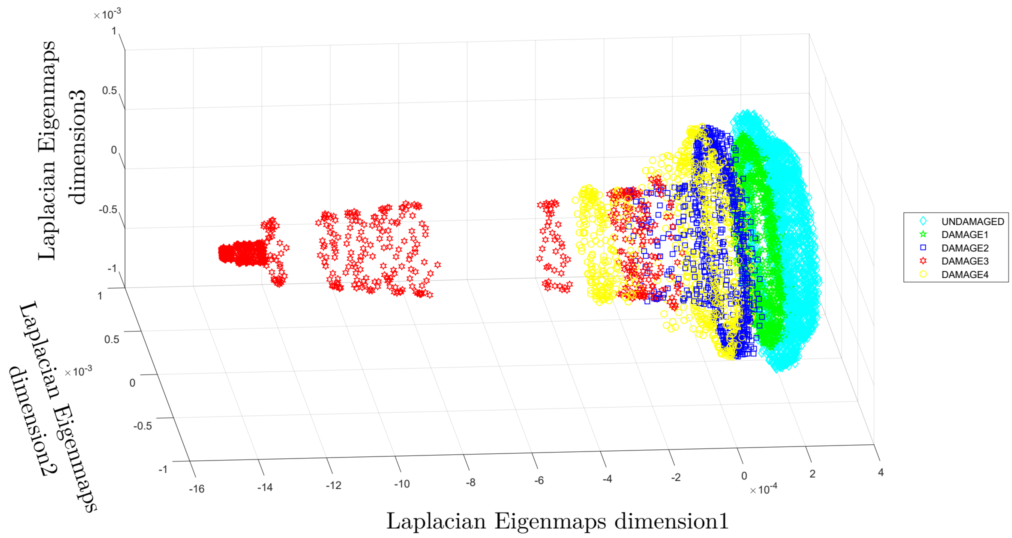

Fig. 6. Representation of the first three dimensions obtained with the Laplacian Eigenmaps method, $k = 700$

Table 4
Confusion matrix obtained with the Laplacian Eigenmaps and XGBoost methods, average accuracy = 0.9995, scaling with MCGS

		Predicted Classl				
		UNDAMAGED	DAMAGE 1	DAMAGE 2	DAMAGE 3	DAMAGE 4
Actual Class	UNDAMAGED	2458	0	2	0	0
	DAMAGE 1	0	820	0	0	0
	DAMAGE 2	0	0	820	0	0
	DAMAGE 3	0	0	0	818	2
	DAMAGE 4	0	0	2	1	817

Table 5
Confusion matrix obtained with the ensemble PCA and Laplacian Eigenmaps using XGBoost classifier method

		Predicted Classl				
		UNDAMAGED	DAMAGE 1	DAMAGE 2	DAMAGE 3	DAMAGE 4
Actual Class	UNDAMAGED	2460	0	0	0	0
	DAMAGE 1	0	820	0	0	0
	DAMAGE 2	0	0	820	0	0
	DAMAGE 3	0	0	0	820	0
	DAMAGE 4	0	0	0	0	820

$k = 700$. As seen in Fig. 6 the classes form rings but in particular, in the resulting representation, the Damage2, Damage3 and Damage4 classes are mixed together.

4.5. Ensemble of PCA and Laplacian Eigenmaps for feature extraction

Taking advantage of the first 8 principal components of the PCA method and the 8 dimensions resulting of applying the Laplacian Eigenmaps method. It is developed an ensemble of the aforementioned features for a total of $8+8 = 16$ features. The final feature matrix has a size of 5740×16. This feature matrix is used to train and validate a XGBoost classifier using a 5-fold cross validation. The final average classification accuracy reaches a value of 100% and its confusion matrix is showed in Table 5.

5. CONCLUSIONS

In this study, a methodology for damage classification in a wind turbine foundation was developed. The experimental test considered a only vibration response of the structure. It was excited using a shaker and its response was measured with 8 triaxial accelerometers. The following insights were found in the development of the data processing methodology:

Data unfolding: the acquired data is arranged in a two-dimensional matrix. Its rows are composed of the experimental trials, and its columns are the product of time instant signals multiplied by the number of sensors.

Data normalization: The MCUGS method was used to preprocess the data to consider the differences in magnitude of the signals acquired by different accelerometers.

Ensemble of feature extraction: a combination of the feature matrices obtained after applying the PCA method and after applying the Laplacian eigenmaps method was used to improve the average classification accuracy.

Data classification: The XGBoost classifier method was used satisfactorily as machine learning algorithm, a 5-fold cross validation was performed obtaining a 100% of average classification accuracy.

As future work it is desirable to test the developed methodology in a real scale wind turbine foundation structure.

ACKNOWLEDGEMENTS

This work has been partially funded by the Spanish Agencia Estatal de Investigación (AEI) – Ministerio de Economía, Industria y Competitividad (MINECO), and the Fondo Europeo de Desarrollo Regional (FEDER) through the research projects PID2021-122132OB-C21 and TED2021-129512B-I00; and by the Generalitat de Catalunya through the research projects 2017-SGR-388 and 2017-SGR-1278. Jersson X. Leon-Medina thanks Universidad de San Buenaventura sede Bogota with the financing of the research project ID= FI-016-006, entitled: Sistema de Navegación Inercial y Visual para Robot Móviles en Espacios Confinados.

REFERENCES

[1] E. Papatheou, N. Dervilis, A.E. Maguire, C. Campos, I. Antoniadou, and K. Worden, "Performance monitoring of a wind turbine using extreme function theory," *Renew. Energy*, vol. 113, pp. 1490–1502, 2017.

[2] B. Asgarian, V. Aghaeidoost, and H.R. Shokrgozar, "Damage detection of jacket type offshore platforms using rate of signal energy using wavelet packet transform," *Mar. Struct.*, vol. 45, pp. 1–21, 2016.

[3] J.X. Leon-Medina, M. Anaya, N. Parés, D.A. Tibaduiza, and F. Pozo, "Structural damage classification in a jacket-type wind-turbine foundation using principal component analysis and extreme gradient boosting," *Sensors*, vol. 21, no. 8, p. 2748, 2021.

[4] Y. Vidal, G. Aquino, F. Pozo, and J.E.M. Gutiérrez-Arias, "Structural health monitoring for jacket-type offshore wind turbines: experimental proof of concept," *Sensors*, vol. 20, no. 7, p. 1835, 2020.

[5] B. Puruncajas, Y. Vidal, and C. Tutivén, "Vibration-response-only structural health monitoring for offshore wind turbine jacket foundations via convolutional neural networks," *Sensors*, vol. 20, no. 12, p. 3429, 2020.

[6] E. Hoxha, Y. Vidal, and F. Pozo, "Damage diagnosis for offshore wind turbine foundations based on the fractal dimension," *Appl. Sci.*, vol. 10, no. 19, p. 6972, 2020.

[7] M. d. C. Feijóo, Y. Zambrano, Y. Vidal, and C. Tutivén, "Unsupervised damage detection for offshore jacket wind turbine foundations based on an autoencoder neural network," *Sensors*, vol. 21, no. 10, p. 3333, 2021.

[8] J. Baquerizo, C. Tutivén, B. Puruncajas, Y. Vidal, and J. Sampietro, "Siamese neural networks for damage detection and diagnosis of jacket-type offshore wind turbine platforms," *Mathematics*, vol. 10, no. 7, p. 1131, 2022.

[9] D. Agis and F. Pozo, "A frequency-based approach for the detection and classification of structural changes using t-sne," *Sensors*, vol. 19, no. 23, p. 5097, 2019.

[10] D. Agis, D.A. Tibaduiza, and F. Pozo, "Vibration-based detection and classification of structural changes using principal component analysis and-distributed stochastic neighbor embedding," *Struct. Control Health Monit.*, vol. 27, no. 6, p. e2533, 2020.

[11] J.X. Leon-Medina, N. Parés, M. Anaya, D.A. Tibaduiza, and F. Pozo, "Data classification methodology for electronic noses using uniform manifold approximation and projection and extreme learning machine," *Mathematics*, vol. 10, no. 1, p. 29, 2022.

[12] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[13] R. Bro and A.K. Smilde, "Principal component analysis," *Anal. Methods*, vol. 6, no. 9, pp. 2812–2831, 2014.

[14] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.