26th ABCM International Congress of Mechanical Engineering
November 22-26, 2021. Florianópolis, SC, Brazil

COB-2021-0579
STATISTICAL INDICATORS-BASED MACHINE LEARNING METHOD FOR CLASSIFICATION OF VIBRATION SIGNALS

Gisele de Fátima Lima Camargo
Eurípedes Guilherme de Oliveira Nóbrega
State University of Campinas (UNICAMP)
School of Mechanical Engineering, Computacional Mechanics Department
e-mails: g208458@dac.unicamp.br, egon@fem.unicamp.br

Abstract. *This paper proposes a vibration analysis methodology, based on statistical indicators and Machine Learning algorithms, for feature selection, fault detection, diagnosis and localization. To assess the proposed methodology, it was initially applied to a rotating machine and then to a Lamb wave-based fault localization in an anisotropic structure. In the preprocessing phase, the statistical indicators are calculated from the signals and normalized. They are then analyzed through scatter and box plot diagrams. Features are represented by the calculated indicators, which are selected to best characterize the state of the respective monitored system, using unsupervised clustering in the hyperplane where the selected features are the coordinates. Each cluster is represented by its centroid. The centroids of each identified cluster, defined through an initial offline procedure, represent together the normal behavior of a system. Distances from the online acquired signals to the centroids are then calculated, in order to detect the eventual anomalies caused by faults. The methodology is first applied to the Case Western Reserve University rotating machine dataset and then to the experimental analysis of Lamb waves propagation in a flexible structure. For the rotating machine, an algorithm based on euclidean distance was developed for data classification. In the case of an anisotropic structure, a fully connected Artificial Neural Network was adopted to locate the faults, which also led to good classifications. The results suggest that the proposed methodology simplify the classification process yielding a good performance in both applications, permitting to expect successful application to monitor mechanical systems.*

Keywords: *Structural Health Monitoring, Machine Learning, Statistical Indicators, Condition Monitoring, Vibration Analysis*

1. INTRODUCTION

Structural Health Monitoring (SHM) methods rely on sensors and software to detect and analyze damages that may occur to parts of mechanical structures. Damages are usually the result of wear caused by operating time, due to natural causes or human intervention, which may lead to failures of the structure. They change the behavior of the system, compromising the quality of the final working, in addition to the possibility of generating catastrophic situations. To prevent damage from occurring, indicators of healthy behavior are used to monitor any change in the state of the system, in order to early identify damage and act to avoid unwanted situations. These methods have been successfully used in the areas of mechanical engineering, civil construction and aeronautics (Balageas *et al.*, 2010) (Farrar and Worden, 2010).

Mathematical modeling is traditional, but data-based methods are becoming the preferred form due to the complexity of the systems. In SHM, the math are in general based on modal properties to identify damage and estimate the remaining service life of the system (Chen and Ni, 2018) (Tibaduiza Burgos *et al.*, 2020). The application of these methods to monitor a real system requires a specialist interacting with some signal analyzers or a computational monitoring system, with a necessary expertise in areas such as mechanical engineering, programming language, signal processing, among others (Wu, 2013). In consequence, the data-driven model has been the most used method nowadays. The fault monitoring and identification systems are based on the observed changes in the vibration signals of the monitored system, which are collected in general from sensors fixed near the monitored area. Relevant articles on data-driven methods emerged in the last decades, starting when some researchers suggested that SHM is mainly about statistical patterns recognition (SPR) (Doebling *et al.*, 1996) (Staszewski *et al.*, 1997). For the present study, the method of recognizing statistical patterns is used to classify and localize faults in mechanical systems.

The principal motivations for using SHM are economic factors and the need to preserve lives. By maximizing the machinery life through continuous monitoring, for example, the wasting of the material is avoided, the production time of the machine is prolonged, product quality is improved and final costs in general are reduced. The benefits of SHM

techniques application to the aeronautic field is even greater because it is directly linked to the people safety. To preserve human health in civil construction, the monitoring of structures, such as buildings and bridges is used in conjunction with other security mechanisms (Farrar and Worden, 2013) (Alamdari *et al.*, 2017).

To solve increasingly complex problems in our society, innovative new mechanisms have been continuously designed, leading also to constant research in SHM. To meet the needs of new applications and economic viability, new materials have been developed. Composite materials, with anisotropic structure, have been used to replace metallic materials, due to their good mechanical characteristics and low specific weight. Different ways of vibration in a system are due to the application of new materials and the development of complex projects (Balageas, 2006).

In order to diagnose the damage to composite materials, sophisticated signal preprocessing techniques are generally used, obtaining satisfactory results (Khan *et al.*, 2019) and (Santos *et al.*, 2019). However, the difficulty and the technical labor required to implement these preprocessing techniques may generate high costs. This study presents faults classification and localization in mechanical systems, using a methodology based on statistical indicators of vibration signals and artificial intelligence algorithms, in which machine learning methods are applied twice, to feature selection and also classification. To fully assess the feasibility of the proposed method, two different mechanical systems are analyzed, respectively a rotating machine and an anisotropic flexible structure.

This study is divided into four sections. In Sec. 2. the main basic concepts adopted in this study are presented. In Sec. 3. the proposed methodology is exemplified to diagnose a bearing fault and its severity on a rotating machine. In Sec. 4. the proposed methodology is applied to localize a simulated damage on a composite plate. In Sec. 5. the conclusions and final considerations are presented.

2. BASIC FUNDAMENTALS

In order to assess the system state, vibration signals are processed and submitted to a pattern recognition method. Signals are recorded and analyzed for damage detection through changes in the system default vibrational mode (Avci *et al.*, 2021). This study follows the diagram presented in Fig. 1.

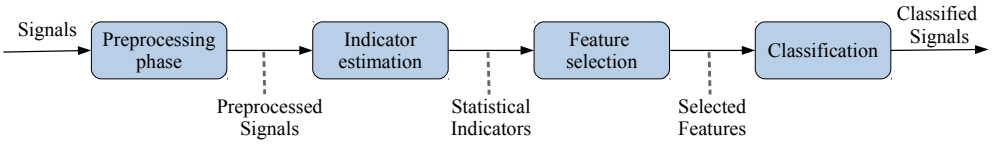


Figure 1: Commonly adopted block diagram for temporal series pattern classification.

As observed in Fig. 1, the preprocessing phase receives the sensors' raw signals that are prepared in order to calculate the statistical indicators in the next block. The indicators are then selected to achieve the good representation of the system behavior, to be submited to a classification process. In the next subsections the adopted techniques for each processing block in Fig. 1 are briefly described.

2.1 Signals preprocessing

The preprocessing module includes one or more steps like cleaning, transformation, dimensionality reduction and normalization. Cleaning includes noise reduction, filling in the missing values or detect some other problem that may occur in the acquisition or storage of the signals. In this process the signals may be truncated and subdivided into smaller signals, and then prepared to be processed. Processing consists generally in using mathematical tools to facilitate the signal analysis. It may transform a time domain signal into a frequency or time-frequency domain signal. The most common used transformations are the Fourier, Hilbert and Wavelet transforms. The Fast Fourier Transform (FFT) maps the vibration signal from time domain to a frequency domain signal, with the signals windowed in time (Caesarendra and Tjahjowidodo, 2017). Hilbert transform is defined as a quadrature linear filter which introduces a 90 degrees phase shift, used to detect envelopes for peak sequences or signal modulation (Wang *et al.*, 2018) (Gowda *et al.*, 2020). An integral transform that has a basis function, called mother wavelet, is used in the Wavelet Transform (WT) to transform a time domain signal into a time-frequency domain signal (Kankanamge *et al.*, 2020). Statistical indicators calculation have been traditionally used to extract attributes from vibration signals (Caesarendra and Tjahjowidodo, 2017). They are adopted as the main processing tool in the present work. The set of all statistical indicators used here are presented in Tab. 1.

For the indicators described in Tab. 1, consider X a finite-time discrete signal with n points such that $X = \{x_1, x_2, \ldots, x_n\}$. Considering $\overline{X}$ the average of X and σ its standard deviation, x_{max} and x_{min} are respectivelly the maximum and minimum value of X. Explanation for the meaning of each of these indicators may be found in Caesarendra and Tjahjowidodo (2017). A normalization is applied to the transformed signals. Normalization scales the values in the same variation range, facilitating comparison between different parameters. The simple scaling, the Min-Max and the Z-score are the most common methods of normalization (Jo, 2019). An approach in signal preprocessing techniques can be found

Table 1: Statistical indicators.

Root Mean Square	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$	(1)	Range	$x_{max} - x_{min}$	(2)
Shape Indicator	$\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}}{\frac{1}{n}\sum_{i=1}^{n} \lvert x_i \rvert}$	(3)	Impulse Indicator	$\frac{\lvert x_{max} \rvert}{\frac{1}{n}\sum_{i=1}^{n} \lvert x_i \rvert}$	(4)
Skewness	$\frac{\sum_{i=1}^{n}(x_i - \overline{X})^3}{(n-1)\sigma^3}$	(5)	kurtose	$\frac{\sum_{i=1}^{n}(x_i - \overline{X})^4}{(n-1)\sigma^4}$	(6)
Crest Indicator	$\frac{\lvert x_{max} \rvert}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}}$	(7)	Clearance Indicator	$\frac{\lvert x_{max} \rvert}{(\frac{1}{n}\sum_{i=1}^{n} \sqrt{\lvert x_i \rvert})^2}$	(8)

in Alasadi and Bhaya (2017).

2.2 Indicator estimation

Signals are visually analyzed through box and scatter diagrams after normalization. The box diagram is a statistical tool used to visualize data dispersion and asymmetry. It returns a numerical summary of the set, based on data median (Marmolejo-Ramos and Tian, 2010). Scatter plot is used for dispersion analysis. It characterizes a point with respect to two or three variables in cartesian coordinates, where it is possible to verify the correlation between them.

2.3 Feature selection

Data analysis and testing allows selecting the indicators that best represent them. For this study, Machine Learning (ML) algorithms are adopted for this purpose. ML is the ability of a machine to perform a specific task, without having been directly programmed to do so. The process is done by an algorithm that aims to learn a function to map an observation to its respective category. For this process, there are two main techniques: supervised and unsupervised machine learning. Supervised learning uses past information to predict new observations. Each information is passed to the algorithm with the category it belongs to. In unsupervised learning the information passed to the algorithm is not labeled. The algorithm separate data in groups with the same characteristics, according to some defined criteria.

2.3.1 *K-Nearest Neighbor*

K-Nearest Neighbor (K-NN) is a supervised ML algorithm. It emerged from the idea that objects near from each other have the same characteristics. Consider a set of pairs (X_i,y_i), with $i = \{1, 2, ..., n\}$ and X_i being an n-dimensional point in the attribute space and y_i being the index indicating the category of that point. The classification process can be described as follows: first, the number of neighbors (k) is identified and passed to the algorithm; second, a distance definition is chosen for the calculation; third, distances from all points to new observation are calculated by the algorithm; fourth, the algorithm orders the calculated distances in ascending order and selects the first points that are associated with the quantity of neighbors chosen in the beggining of the process; finally it classifies the new point with the category that has more votes among the first k quantity neighbors.

2.3.2 *K-Means*

K-Means (KM) is an unsupervised ML algorithm frequently used in data mining. The goal is to separate objects with same characteristics into groups, according to the minimum distance criteria. To start the clustering process, the number (k) of clusters is usually provided to the algorithm. From this information, the whole process is done automatically. Randomly the algorithm selects the first prototypes (u_k) from the set of the objects. The term prototype is used here to refer to the centroids of the clusters; second, it associates each object with the closest prototype; third, it recalculates and updates the prototypes to restart a new iteration, using the same set of objects. These steps are repeated until the algorithm reaches some stopping criterion.

The minimum distance criterion is defined as follows: consider two n-dimensional points in the attribute space: $P_1 = (p_{1,1}, p_{1,2}, ..., p_{1,n})$ and $P_2 = (p_{2,1}, p_{2,2}, ..., p_{2,n})$. The calculation of the euclidean distance between these two points

can be obtained by Eq. 9. In the present study, the word distance is used to refer to the euclidean distance.

$$||P_1 - P_2|| = \sqrt{(p_{1,1} - p_{2,1})^2 + (p_{1,2} - p_{2,2})^2 + \ldots + (p_{1,n} - p_{2,n})^2} \tag{9}$$

Consider the set of objects $\{X\} = \{X_1, X_2, ..., X_m\}$. The elements of each subset belong to the same category: $C_1, C_2, ..., C_k$. The centroids for each subset are obtained by using Eq. 10:

$$u_k = \frac{1}{N_k} \sum_{X \in k} X, \tag{10}$$

with N_k equal to the number of points in the *k* cluster and u_k is the median or centroid of it.

2.4 Classification

Classification is the process of distinguishing and assigning objects to their respective classes and concepts through a model. In this process a percentage of the data is used for training and the percentage of remaining data is used to measure the classifier performance.

2.4.1 Artificial Neural Networks

Artificial Neural Network (ANN) is a computational model inspired by the functioning of nerve cells, the biological neurons, existing in living beings. It is constructed by small units that are connected to each other to form a complex neural network. These units are referred here as neurons. In 1943, Warren McCulloch and Walter Pitts presented in their work a neuron model that used a binary threshold activation function, using propositional logic as a way of functioning, true for 1 and false for 0. The model had two inputs which produced a binary output (McCulloch and Pitts, 1943). In 1958, Frank Rosenblatt proposed a probabilistic model that added weights (w) to the inputs of the neuron, which represented the importance they would reproduce in the output (Rosenblatt, 1958) (Hebb, 1949). This probabilistic model became known as *Perceptron*. An example of a Perceptron may be seen in Fig. 2.

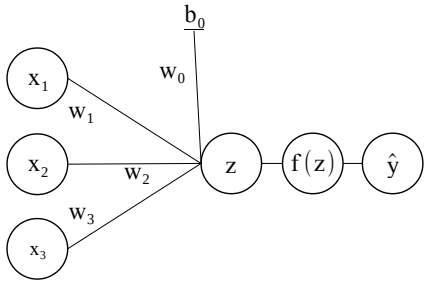


Figure 2: The basic neuron representation.

The Perceptron represented in Fig. 2 corresponds to a single artificial neuron with three inputs (x_1, x_2 and x_3), that are associated to their respective weights (w_1, w_2 and w_3), resulting in a weighted linear combination, to which the bias b_0 is added, also associated with the weight w_0, such that $b = w_0 b_0$. An activation function $f(z)$ is used to generate a nonlinear output from the artificial neuron $\hat{y}$ (Khanna, 1990). The Eq. 11a represents the variable *z*, which is the weighted sum for the three inputs of the neuron in Fig. 2, and $\hat{y}$ represented in Eq. 11b is the result of the activation function applied to this variable.

$$z = w_1 x_1 + w_2 x_2 + w_3 x_3 + b \tag{11a}$$

$$\hat{y} = f(z) = f(w_1 x_1 + w_2 x_2 + w_3 x_3 + b) \tag{11b}$$

Sigmoidal activation and the Rectified Linear Unit (ReLu) are examples of activation functions. A summary with the major activation functions recently used in ANN's can be found in Nwankpa *et al.* (2018).

A common architecture is the Multilayer Perceptron (MLP) that represents a universal nonlinear approximator, which is formed by three layers: one input layer, which does not have neurons, one so called hidden layer with nonlinear activation neurons, and an output layer with linear neurons (Ososkov and Goncharov, 2017). When the MLP has more than these three layers, it is called a *deep* network, which may present several recently introduced architectures.

The simplest way for an ANN to process information is in the feedforward direction. Training is accomplished with the well known backpropagation algorithm, which minimizes the error in the output layer. Backpropagation is a systematic way of updating the weights associated with the neurons in the network, with the goal of minimizing the difference between the result of the predicted and the actual values of the classes (Rumelhart *et al.*, 1986). The process takes place at the end of each iteration, when the performance of the algorithm is measured using a differentiable loss function. An iterative optimization algorithm to find local minimum (gradient descent method) is usually used to minimize the loss function (Vishwakarma *et al.*, 2020). A review of loss functions and the main types of ANN's and their characteristics can be found in Wang *et al.* (2020) and Emmert-Streib *et al.* (2020), respectively.

3. METHODOLOGY

3.1 Case 1: rotating system

The vibration signals are part of the database of Case Western Reserve University (CWRU, 2021) and were collected from the experimental system, observed in Fig. 3.

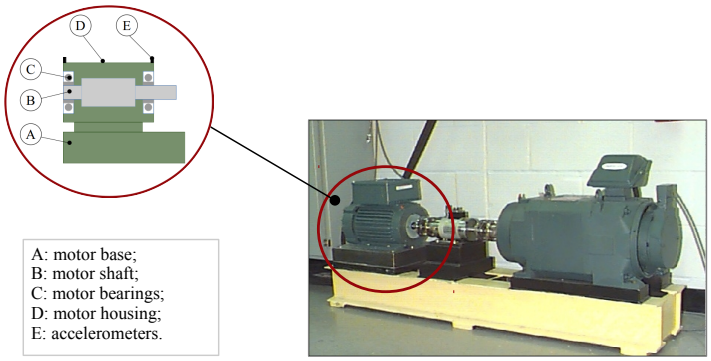


Figure 3: Rotating machine setup used to generate the CWRU dataset (reproduced from CWRU (2021) site).

Table 2: System operating characteristics.

motor load-Hp	motor speed-r.p.m.
1	1797
2	1772
3	1750
4	1730

The experimental rotating system consists of the motor, on the left, a transducer, in the center and a dynamometer, on the right. As observed in detail of Fig. 3, the accelerometers are attached to the motor frame, at the 12 o'clock position in relation to the axis.

Simulations were performed with different failure diameters, located on the bearing inner race. In addition, four loads with different speeds were considered, as observed in Tab. 2. The signals were collected by the accelerometer positioned above of the output shaft motor, denominated in the CWRU files as *Drive End*. Samples were collected at 12 kHz frequency. The failure diameters considered for the study were 0.178 and 0.530 mm, respectively, both with a depth of 0.279 mm. For each failure diameter, four signals with 120832 points were collected. A set of signals representing the health state of the bearing were also collected. Acceleration signals of the system is going to be analyzed in order to compare both normal and failure vibration signals.

3.2 Signals preprocessing

The signals were segmented into 1024 point signals,resulting in a vibration signal matrix M of dimension 1416×1024. Normal class (NM) were determined for healthy state signals; Small Fault (FP) and Large fault (FG) classes were determined for signals with bearing failure diameter of 0.178 mm and 0.530 mm, respectively. A total of eight statistical indicators were calculated for each sample, and the MinMax method was used for data normalization, in a interval of [0,1], as observed in Eq. 12, resulting in a statistical indicator matrix I of dimension 1416×8. Let X be the representation of a sample in the attribute space, with n being the number of attributes. Then: $X = \{ x_1, x_2, \dots, x_n \}$,

$$X_{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{12}$$

X_{norm} is the normalized value for X, x_{min} is the minimum value for the set and x_{max} is the maximum value of the set.

3.2.1 Indicator estimation

For each indicator, data were grouped into classes and analyzed using box and scatter diagrams, as observed in Fig. 4. As observed in Fig. 4, RMS indicator, on the left up localization, the abscissa represents the loads and their respective

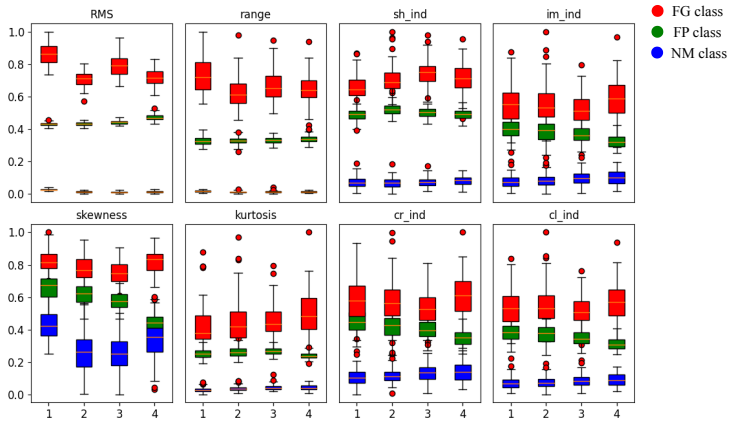


Figure 4: Normalized statistical indicators.

rotations, the ordinate represents the ranges of values from the indicators. The blue, green and red colors represent NM, FP and FG classes, respectively. The described structure is the same for the last seven indicators diagrams. In order to make the scatter diagram a usefull tool, the combination of three indicators was provided.

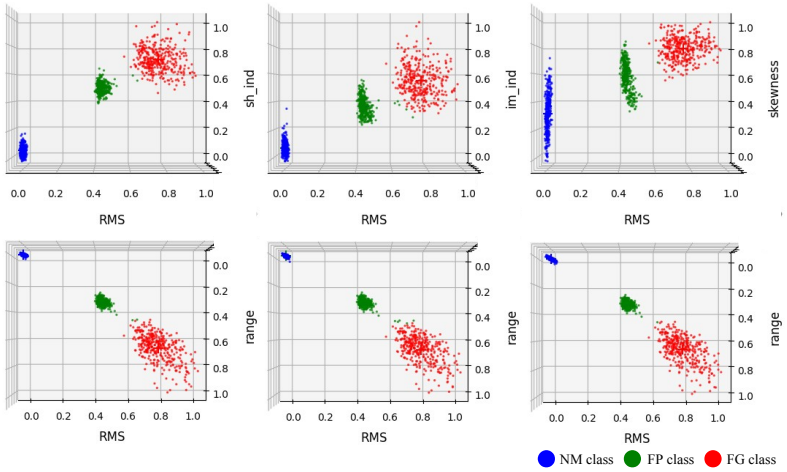


Figure 5: Orthogonal views of three indicators combination.

It may be observed in Fig. 5 that the RMS values presented well-defined ranges of values for each class, without intersection between them; range indicator presented a minimum intersection between FP and FG classes. Other parameters, except for the RMS, presented an intersection between FP and FG classes; range, Kurtosis, Shape Indicator and Clearance Indicator do not presented intersection between the NM and FP classes, they characterize the healthy and failure states of the rotating system; Impulse Indicator, Skewness Indicator and Crest Indicator presented intersection between classes NM and FP and FP and FG. The analysis and selection of the statistical indicators were performed using the concept and the operations between the sets. Which is considered a counting problem in a probabilistic approach, observed mainly in the points diagram.

3.3 Feature selection

By considering some intersection between classes, three indicators were select for training and testing the algorithms: range, shape indicator and skewness. These intersections allows the algorithms some desirable generalization. For each selected indicator, the centroides of the classes NM, FP and FG were calculated using the KM algorithm, and distances from prototypes to each cluster were calculated using the euclidean distance.

3.4 Classification

An algorithm was implemented to identify the minimum distances from each prototype to the clusters of the classes and to classify the data. To validate this developed algorithm, three algorithms from scikit-learn and keras libraries were used for detecting and classifying damages in the rotating system: K-Means, K-Nearest Neighbor and shallow neural network. The percentage of 75% of the total samples was separeted for training the algorithms and the remaining 25%

of the samples were separated for testing. The classification results are calculated using the confusion matrix and are presented in Tab. 3 to Tab. 6.

Table 3: Implemented algorithm classification.

classes	true	predict	FP	FN	CA	total
NM	112	112	0	0	-	112
FP	124	124	0	0	-	124
FG	118	117	0	1	FP	118

Table 4: K-Means classification.

classes	true	predict	FP	FN	CA	total
NM	112	112	0	0	-	112
FP	124	124	0	0	-	124
FG	118	117	0	1	NM	118

As observed in Tab. 3, column *true* are the correct quantity of samples per class. Column *predict* are the correct labels quantity predicted by the classifier. *FP* column means *False Positive*. *False Negative* is in column *FN*. The *CA* column is the *Class Attribution* of *False Negative* or *False Positive*. In *total* column is the total number of samples. The False Negative was assigned to class FP. The accuracy of the distance-based algorithm was 99.72%. As observed in Tab. 4, one False Negative was assigned to class NM. The accuracy of K-Means classifier was 99.72%.

Table 5: K-Nearest Neighbor classification.

classes	true	predict	FP	FN	CA	total
NM	112	112	0	0	-	112
FP	124	124	0	0	-	124
FG	118	117	0	1	FP	118

Table 6: Shallow neural networks classification.

classes	true	predict	FP	FN	CA	total
NM	112	112	0	0	-	112
FP	124	124	0	0	-	124
FG	118	117	0	1	FP	118

As observed in Tab. 5, one False Negative was assigned to class FP. The accuracy of K-Nearest Neighbor classifier was 99.72%. As observed in Tab. 6, one False Negative was assigned to class FP. The accuracy of shallow neural networks classifier was 99.72%. The performance of all methods may be considered equivalent for this dataset.

4. ANISOTROPIC PLATE

For the second application of the methodology an experimental setup based on a composite plate, represented in Fig. 6, was analysed using some configurations of data structures and neural network classifiers. The faults were simulated through deposition of a small pair of magnetic cubes in different regions of the plate. The problem is to detect and localize a fault estimating the region where it is located.

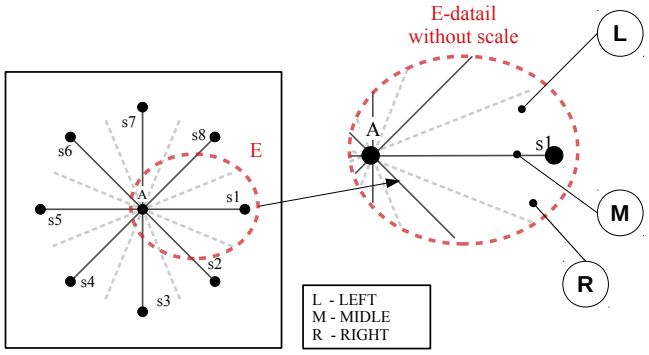


Figure 6: Flexible structure and datailed regions.

Table 7: Samples characteristics

regions	class	number of experiments
-	0	100
s1	1	100
s2	2	100
s3	3	100
s4	4	100
s5	5	100
s6	6	100
s7	7	100
s8	8	100

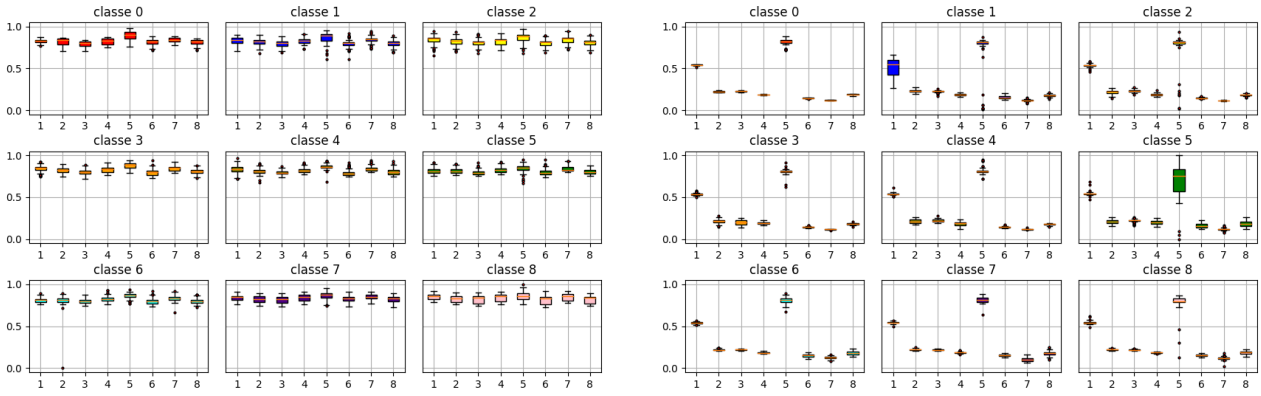
As may be seen in Fig. 6, nine circular piezoelectric transducers, with a diameter of 20 mm, are fixed on this anisotropic flexible structure with dimensions of 300 mm x 300 mm x 1 mm. A transducer was positioned in the center of the plate, to be used as an actuator to generate the Lamb wave, and the 8 other transducers, acting as sensors, were located to form a circle with a radius of 100 mm. As observed in the detail of Fig. 6, eight regions were determined by the sensors arrangement, limited by the dashed lines, each one divided into three sub-regions, Left (L), Right (R), and the Medium (M), when the cubes were colocated precisely in the sensor/actuator line. The 5 mm magnetic cubes were placed randomly on each region, one on each side of the plate to maintain their position. The actuator was excited with a burst of five high frequency sine wave cycles, multiplied by a *Hanning* window, with the respective propagated signals acquired from the eight sensors. Each sensor signal was labeled accordingly to the region where the two cubes were placed. The sampling frequency used was 1.21 MHz, based on the fact that high frequencies are capable of detecting very small damages (Worden *et al.*, 2007). Each set of sensor responses to excitation for the damage in each region, including also the case

without damage, is called an *experiment*. Considering that the signals were sampled with 1024 points equally spaced in time, each experiment corresponds to a vibration signal matrix containing 72 discretized signals with 1024 sample points. A total of 100 experiments were collected per plate region, however the number of tested sub-regions on each experiment is not the same. Nine classes were adopted for the initial supervised classification, according to Tab. 7, but afterwards 25 different classes were also classified, where the sub-regions were considered separately.

4.1 Feature extraction

The 900 collected experiments gave rise to a vibration signal matrix M of dimension 7200×1024. These signals are preprocessed using a Discrete Wavelet and a Hilbert Transforms, in order to generate the envelopes of the modulated signal acquired by each sensor, where the burst high frequency is the carrier of the propagated signal (Santos *et al.*, 2019). Eight statistical indicators are calculated for each signal, which resulted in a statistical indicator matrix of dimension 7200×8. The MinMax normalization method was used to normalize the parameter data.

The signals were grouped into classes and analyzed using box and scatter diagrams. As observed in Fig. 7, where data from two parameters are represented, the set that forms the intersection between the class objects is not empty and occurs in all diagrams. The presence of *outliers* were observed in all classes. Since these indicators presented a considerable set of intersections between classes, all the eight statistical indicators were considered for the preprocessing phase, representing the hiperdimensional attibute space.



(a) Root Mean Square indicator. (b) Range indicator

Figure 7: RMS and range indicators box plots.

4.2 Tested configurations

Initially, the signals were preprocessed in order to extract the respective envelopes, which were adopted for the feature extraction process. For the nine fault conditions and respective eight sensor measurements, 72 clusters in a 8 dimensions parameter hyperspace, represented by its centroids, were calculated. A training set was calculated using distances for the centroids using 75 % of the 7200 matrices of dimension 8×8 dataset, where the rows are the sensors and the columns are the respective parameters. A classification deep neural network with four neural layers with alternating three dropout layers and a softmax output for 9 classes, was trained using 72 different models based on the 8 statistical indicators for the 9 regions. The training input was the data matrix with the respective euclidian distances to each centroid. It was tested with the next 25 % of the experiments. For the algorithm validation process, a total of 10% of the training set was considered.

A second configuration was adopted considering the 24 sub-regions and the no-fault region R0. However, there are 100 experiments concerning R0, 45 experiments for each region, for example, 45 for R1L (left side of each sensor in R1), and also 45 for the right side (R1R), but only 10 experiments for the medium line (R1M). That is the same for all regions, from R1 to R8. For this case, after expanding the medium line set to achieve balanced classes, a data matrix was generated, with dimension 1215 x 8192 signal samples, yielding a data matrix with 1215 experiments x 64 statistical parameters. It was trained under supervision 67 % of these experiments for a deep MLP with 64 inputs, 2 hidden layers respectively with 150 and 96 neurons and a softmax with 9 outputs.

A third configuration considering also the 25 regions and corresponding 45 experiments and 25 classes, generated a data matrix with 405 experiments x 8192 signal samples, adopting 85 % of them for the training subset. An MLP network with 8192 inputs, 596 neurons in the hidden layer and a softmax output layer with 25 outputs was trained. In a second moment, an output of 9 classes was adopted.

4.3 Discussion

To locate the failures in the flexible structure, three configurations were tested, where the first adopted the statistical parameters as the analysed features. The first configuration, a deep neural network with 8 neuronal layers, achieved an accuracy of 76.3 %. The second configuration, also adopting statistical indicators as features and a deep network, with only three neuronal layers, achieved an accuracy of 75.5 %, besides having considered the three subregions separated however with only the 9 classes. Finally, the third configuration processed the original modulated signal, a shallow network using only two neuronal classes, achieving an accuracy of 78.7 % taking into account the 25 classes, considering the subregions. However, adopting an output of 9 classes, where each class includes the respective subregions, this number increased to 89.7 %, representing the best accuracy of the all tested configurations.

5. CONCLUSIONS

A methodology for classification and localization of damage in two mechanical systems was developed and validated using statistical indicators and ML algoritms for feature extraction and a neural network and ML algorithms for the classification process. For that, eight statistical indicators from vibration signals were calculated and normalized, then data were analyzed and validated through statistical diagrams. The rotating machine classification was performed utilizing the minimum distance criteria. For the anisotropic structure the damage localization process used the statistical parameters through finding unsupervised clusters in a hyperplane, and its respective classifications using neural networks. In both cases, the proposed methodology achieved good results, however a configuration whose input was only the modulated signals of the Lamb wave propagation through the anisotropic plate achieved the better accuracy.

Time domain feature extraction, as statistics indicators calculation, is a traditional process to extract relevant information from a signal that could be explored for mechanical damage studies. The possibility of achieving good results for unsupervised configurations is of extreme value when the labels of the data patterns are not known. The research of the statistical parameters and indicators and its application to mechanical systems is indeed in a early stage, but it is a promising area of machine learning application.

The crescent utilization of artificial intelligence methods to solve SHM problems has generated a great demand for specialized people to perform technical tasks. The possibility of utilizing a standardized method can be an opportunity for quickly solving SHM problems, leading to economic and life security advantages.

The proposed methodology proved to be useful in several phases of fault detection and localization process. Data analysis leads to the appropriate algorithm selection for the classification task. The use of calculating distances and centroids can represent sufficient information for damage classification.

6. ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and by the Instituto de Pesquisas Eldorado. The authors would like to thank these institutions for providing the financial support.

7. REFERENCES

Alamdari, M.M., Rakotoarivelo, T. and Khoa, N.L.D., 2017. "A spectral-based clustering for structural health monitoring of the sydney harbour bridge". *Mechanical Systems and Signal Processing*, Vol. 87, pp. 384–400.

Alasadi, S.A. and Bhaya, W.S., 2017. "Review of data preprocessing techniques in data mining". *Journal of Engineering and Applied Sciences*, Vol. 12, No. 16, pp. 4102–4107.

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M. and Inman, D.J., 2021. "A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications". *Mechanical Systems and Signal Processing*, Vol. 147, p. 107077.

Balageas, D., 2006. *Introduction to Structural Health Monitoring*, John Wiley & Sons, Ltd, chapter 1, pp. 13–43.

Balageas, D., Fritzen, C.P. and Güemes, A., 2010. *Structural health monitoring*, Vol. 90. John Wiley & Sons.

Caesarendra, W. and Tjahjowidodo, T., 2017. "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing". *Machines*, Vol. 5, No. 4, p. 21.

Chen, H.P. and Ni, Y.Q., 2018. *Structural Damage Identification Techniques*, John Wiley & Sons, Ltd, chapter 4, pp. 69–90.

CWRU, 2021. "Case western reserve university bearing data center website". https://csegroups.case.edu/bearingdatacenter/.

Doebling, S., Farrar, C., Prime, M. and Shevitz, D., 1996. "Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: A literature review". Technical report, Los

Alamos National Lab., NM (United States).

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S. and Dehmer, M., 2020. "An introductory review of deep learning for prediction models with big data". *Frontiers in Artificial Intelligence*, Vol. 3, p. 4.

Farrar, C.R. and Worden, K., 2010. *An Introduction to Structural Health Monitoring*, Springer Vienna, Vienna.

Farrar, C.R. and Worden, K., 2013. *Introduction*, John Wiley & Sons, Ltd, chapter 1, pp. 1–16.

Gowda, V.D., Pai, G.N., Sridhara, S., Shashidhara, K. *et al.*, 2020. "Signal analysis and filtering using one dimensional hilbert transform". In *Journal of Physics: Conference Series*. IOP Publishing, Vol. 1706, p. 012107.

Hebb, D.O., 1949. *The organization of behavior: a neuropsychological theory*. J. Wiley; Chapman & Hall.

Jo, J.M., 2019. "Effectiveness of normalization pre-processing of big data to the machine learning performance". *The Journal of the Korea institute of electronic communication sciences*, Vol. 14, No. 3, pp. 547–552.

Kankanamge, Y., Hu, Y. and Shao, X., 2020. "Application of wavelet transform in structural health monitoring". *Earthquake Engineering and Engineering Vibration*, Vol. 19, No. 2, pp. 515–532.

Khan, A., Kim, N., Shin, J., Kim, H.S. and Youn, B.D., 2019. "Damage assessment of smart composite structures via machine learning: a review". *JMST Advances*, Vol. 1, pp. 1–18.

Khanna, T., 1990. "Foundations of neural networks".

Marmolejo-Ramos, F. and Tian, T.S., 2010. "The shifting boxplot. a boxplot based on essential summary statistics around the mean." *International Journal of Psychological Research*, Vol. 3, No. 1, pp. 37–45.

McCulloch, W.S. and Pitts, W., 1943. "A logical calculus of the ideas immanent in nervous activity". *The bulletin of mathematical biophysics*, Vol. 5, No. 4, pp. 115–133.

Nwankpa, C., Ijomah, W., Gachagan, A. and Marshall, S., 2018. "Activation functions: Comparison of trends in practice and research for deep learning". *CoRR*, Vol. abs/1811.03378, pp. 1–20.

Ososkov, G. and Goncharov, P., 2017. "Shallow and deep learning for image classification". *Optical Memory and Neural Networks*, Vol. 26, No. 4, pp. 221–248.

Rosenblatt, F., 1958. "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review*, Vol. 65 6, pp. 386–408.

Rumelhart, D., Hinton, G.E. and Williams, R.J., 1986. "Learning representations by back-propagating errors". *Nature*, Vol. 323, pp. 533–536.

Santos, R.B.M., Souza, P.R., Inocente-Junior, N.R. and Nóbrega, E.G.O., 2019. "Composite plate automatic damage isolation based on support vector machine classification of lamb wave signals". In *Proceedings of ECCOMAS SMART 2019*. Paris, France.

Staszewski, W., Worden, K. and Tomlinson, G., 1997. "Time-frequency analysis in gearbox fault detection using the wigner-ville distribution and pattern recognition". *Mechanical Systems and Signal Processing*, Vol. 11, No. 5, pp. 673–692.

Tibaduiza Burgos, D.A., Gomez Vargas, R.C., Pedraza, C., Agis, D. and Pozo, F., 2020. "Damage identification in structural health monitoring: A brief review from its implementation to the use of data-driven applications". *Sensors*, Vol. 20, No. 3.

Vishwakarma, G.K., Paul, C. and Elsawah, A., 2020. "An algorithm for outlier detection in a time series model using backpropagation neural network". *Journal of King Saud University-Science*, Vol. 32, No. 8, pp. 3328–3336.

Wang, Q., Ma, Y., Zhao, K. and Tian, Y., 2020. "A comprehensive survey of loss functions in machine learning". *Annals of Data Science*, Vol. OnlineFirst, pp. 1–26.

Wang, Y., Liu, Z., Jiang, C. and Zhang, S., 2018. "Motion induced phase error reduction using a hilbert transform". *Opt. Express*, Vol. 26, No. 26, pp. 34224–34235. doi:10.1364/OE.26.034224. URL `http://www.opticsexpress.org/abstract.cfm?URI=oe-26-26-34224`.

Worden, K., Farrar, C.R., Manson, G. and Park, G., 2007. "The fundamental axioms of structural health monitoring". *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 463, No. 2082, pp. 1639–1664.

Wu, J.S., 2013. *Analytical and numerical methods for vibration analyses*. John Wiley & Sons.