Journal Paper

“Linear and Nonlinear Features and Machine Learning for Wind Turbine Blade Ice Detection and Diagnosis”

- Renewable Energy 2019 -

Alfredo Arcos Jiménez

Ingenium Research Group, Universidad de Castilla-La Mancha

alfredo.arcos@alu.uclm.es

Fausto Pedro García Márquez

Ingenium Research Group, Universidad de Castilla-La Mancha

FaustoPedro.Garcia@uclm.es

Victoria Borja Moraleda

Ingenium Research Group, Universidad de Castilla-La Mancha

victoria.borja@alu.uclm.es

Carlos Quiterio Gómez Muñoz

Universidad Europea, Spain

carlosquiterio.gomez@universidadeuropea.es

Cite as: Jiménez, A. A., Márquez, F. P. G., Moraleda, V. B., & Muñoz, C. Q. G. (2019). Linear and nonlinear features and machine learning for wind turbine blade ice detection and diagnosis. *Renewable Energy*, 132, 1034-1048.

D.O.I.: <https://doi.org/10.1016/j.renene.2018.08.050>

Linear and Nonlinear Features and Machine Learning for Wind Turbine Blade Ice Detection and Diagnosis

Alfredo Arcos Jiménez¹, Fausto Pedro García Márquez¹, Victoria Borja Moraleda¹,
Carlos Quiterio Gómez Muñoz²

¹ Ingenium Research Group, Castilla-La Mancha University, Spain.

alfredo.arcos@alu.uclm.es, faustopedro.garcia@uclm.es, victoria.borja@uclm.es

² Universidad Europea, Spain.

carlosquiterio.gomez@universidadeuropea.es

Abstract

The mass of ice on wind turbines blades is one of the main problems that energy companies have in cold climates. This paper presents a novel approach to detect and classify ice thickness based on pattern recognition through guided ultrasonic waves and Machine Learning. To successfully achieve a supervised classification, it is necessary to employ a method that allows the correct extraction and selection of features of the ultrasonic signal. The main novelty in this work is that the approach considers four feature extraction methods to validate the results, grouped by linear (AutoRegressive (AR) and Principal Component Analysis) and nonlinear (nonlinear-AR eXogenous and Hierarchical Non-Linear Principal Component Analysis), and feature selection is done by Neighbourhood Component Analysis. A supervised classification was performed through Machine Learning with twenty classifiers such as Decision tree, Discriminant Analysis, Support Vector Machines, K-Nearest Neighbours, and Ensemble Classifiers. Finally, an evaluation of the classifiers was done in single frequency and multi-frequency modes, obtaining accurate results.

Key Words: Feature Extraction, NARX, NLPCA, NCA, Machine Learning, Guided Waves, Ice, Wind Turbine Blade, Classifiers.

1 Abbreviation list

AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AR	Autoregressive
ARX	Autoregressive exogenous
AUC	Area Under Curve
COGSVM	Coarse Gaussian Support Vector Machine
COKNN	Coarse K-Nearest Neighbours
CSKNN	Cosine K-Nearest Neighbours
CKNN	Cubic K-Nearest Neighbours
CSVM	Cubic Support Vector Machine
DA	Discriminant Analysis
DT	Decision Tree
DTC	Decision Tree Complex
DTM	Decision Tree Medium
DTS	Decision Tree Simple
EBT	Ensemble Bagged Tree
ESD	Ensemble Subspace Discriminant
ESKNN	Ensemble Subspace K-Nearest Neighbours
FE	Features Extraction
FGSVM	Fine Gaussian Support Vector Machine
FKNN	Fine K-Nearest Neighbours
FS	Features Selection
GW	Guided Wave
h-NLPCA	hierarchical Nonlinear Principal Component Analysis
KNN	K-Nearest Neighbours
LDA	Linear Discriminant Analysis
LSVM	Linear Support Vector Machine
MFC	Macro Fibre Composites
MGsVM	Medium Gaussian Support Vector Machine
MKNN	Medium K-Nearest Neighbours
ML	Machine Learning
MSE	Mean Square Error
NARX	Nonlinear autoregressive with exogenous

NCA	Neighbourhood Component Analysis
NLPCA	Nonlinear Principal Components Analysis
PCA	Principal Components Analysis
QDA	Quadratic Discriminant Analysis
QSVM	Quadratic Support Vector Machine
SHM	Structural Health Monitoring
SRE	Squared Reconstruction Error
SVM	Support Vector Machine
WKNN	Weighted KNN
WT	Wind Turbine
WTB	Wind Turbine Blade

1. INTRODUCTION.

Renewable Energy requires the advances in the reliability, availability, maintainability and safety in order to be competitiveness in the electrical energy market [1,2]. It is important in cold areas, where the energy production can be affected by the technology employed [3]. Wind farms in cold climate zones are expected to increase 12 GW per year until 2020 in Scandinavia, North America, Europe, and Asia [4]. Ice on wind turbines in icy environments is one of the main challenges of the wind power industry [5,6]. The most significant problems caused by ice are power loss due to the reduction of aerodynamic efficiency [7,8], loads on turbines [9], the lifetime of the components [10], increased noise generated by blades [11], changes in blade surfaces by ice accretion [12], safety hazards [13], and measurement errors [14,15]. The mass of ice on Wind Turbines Blades (WTB) has special significance in the operation of the wind turbine [16], similar to other elements on the WTB surface, e.g. dirt and mud [17,18]. The wind turbine is stopped when a certain mass of ice is reached [19]. In post-icing, the last stage, the wind turbine remains out of order until the ice has disappeared [6,20].

Ice detection can be classified into two main categories[21]: The indirect and direct approach. The indirect approach provides measures that predict ice events, combining environmental conditions, e.g. temperature, humidity, etc., with empirical or deterministic models. In the direct approach, considered in this paper, the sensor measures the alteration of properties such as mass, conductivities, inductances, dielectric constants, condition of ice accumulation, and deformation, etc.

Nowadays, it is difficult to detect ice due to the need to find reliable and low cost sensors [15,22]. There are several techniques used for ice detection[23]. Ultrasonic Guided Waves (UGW) are one of the most efficient and reliable techniques since they can detect discontinuities in large areas [24,25], including defects on a surface [26,27], e.g. cracks. Tanget et al. carried out an experimental study of acoustic emission methodology for the in service condition monitoring of WTBs [28]. Recent studies show methodologies based on UGW condition monitoring employ different patterns of ice thickness [29].

UGWs have also been used as protection systems that eliminate the ice. For example, H. Habibi et al. [30] studied a dual de-icing system for WTB, employing high power UWG and low frequency forced vibrations. Goraj develops de-icing devices based on the stress created by the UGW in the junction of the two materials [31]. Zeng and Song researched

the experimental and numerical simulation of ultrasonic de-icing for WTB [32]. Wang et al. developed light lithium niobate transducers for the ultrasonic de-icing of WTBs [33].

Pattern recognition is employed in this research together with Structure Health Monitoring (SHM) [34] for fault detection and diagnosis [35]. Esu et al. considered the feasibility of a fully autonomous wireless monitoring system for a WTB [36]. Song et al. studied a wind turbine health state monitoring system based on a Bayesian data-driven approach [37]. Xiao et al. presented a reconfigurable tolerant control of uncertain mechanical systems with actuator faults [38]. The references do not consider multi-class classification for ice detection and diagnosis. In the last decade, multi-class classification using non-linear classifiers, e.g. Support Vector Machines (SVM) [39], has achieved excellent results in other cases. This paper employs linear and nonlinear models of Features Extraction (FE).

FE is an efficient process for detecting faults when it is applied to pattern recognition [40], and is essential for setting the best FE model. Principal Component Analysis (PCA), AutoRegressive (AR) and AR eXogenous (ARX) have been employed with accuracy [41,42]. Recently, non-linear models, such as Hierarchical Non-Linear Principal Component Analysis (h-NLPCA), have been applied in fault classification processes [43]. In contrast, Nonlinear ARX (NARX) is not common in pattern recognition because it possesses a large number of features, which influence the performance of the classifier and increase the probability of overfitting. Neighbourhood Component Analysis (NCA) is a Feature Selection (FS) method that provides good performance, and can reduce the number of features and increase accuracy [44]. These methods have not been applied to the case study presented in this paper.

The approach based on Machine Learning (ML) could be applied in other cases studies, e.g. electrical grid by using extreme learning machines and self-adaptive evolutionary extreme learning machines to directly modelling prediction intervals [45]. ML in WTB has been employed in SHM [46]. Skrimpas et al. studied ice detection in WTB by vibration and power curve analysis, together with the K-Nearest Neighbours (KNN) method [47]. Sanggyun Lee et al. [48] applied two ML algorithms, Decision Trees (DT) and random forest, to estimate the thickness of ice in the Arctic Sea. ML has not yet been employed with UGW for ice detection and diagnosis in WTB.

The main novelty in this paper is the approach for multilevel detection of ice thickness by UGW and ML. The approach considers new methodologies for FE based on the non-linear method through NARX and h-NLPCA, together with linear as AR and PCA, and FS by

NCA. Different linear and non-linear classifiers are analysed by alternating different linear and non-linear FE methods to validate the results. The classification is also evaluated and validated by frequency and multifrequency levels. Finally, the best classifiers and FE model are selected according to the results.

2. CASE STUDY

2.1 Experimental procedure.

A case study was carried out in the laboratory using UGW to detect ice on the WTB. The purpose is to train the different classifiers with the experiments, and to study and validate the approach for ice detection.

Figure 1 shows the condition monitoring system employed in the experiments. The National Instruments PXIe-6124 is controlled by Matlab software, where the ultrasonic signals are generated with the number of cycles, frequency and curve shape desired. In this case, a Hanning window has been applied to the excitation pulses to improve the generation and transmission of the UGW. The electrical signal generated is amplified with the TREK 2100HF amplifier to reach $\pm 150V$, employed as input in the ultrasonic transducer. The transducers are Macro Fiber Composite (MFC) type with interdigitated piezoelectric fibers [49].

Fig. 1 Icing detection system

The technique used in the experiments is called ‘pitch and catch’, where a short ultrasonic pulse is emitted by the MFC transmitter (Tx), and then it is collected by the MFC receiver

(Rx). The mechanical wave travels through the blade, being sensible to the existence of elements on the surface. The receiver converts the ultrasonic wave into an electrical signal, which must be amplified to be acquired correctly by the same equipment PXIe-6124.

Five different scenarios were studied in the experiments to detect and diagnose the cases considered in ISO 12494, 2001 (Atmospheric icing of structures) [50]: The first experiment was carried out at room temperature; the second was done at freezing temperature but with no ice on the WTB; the rest of experiments were with frozen WTB and ice on the WTB, where the ice thickness was progressively increased in each scenario.

The experiment details are shown in Table I

TABLE I Scenarios and Delamination Size.

Scenario	Thickness of ice
1	0 mm without ice at room temperature
2	0 mm without ice (frozen WTB) - 22°C
3	1 mm ice (frozen WTB) -22°C
4	5 mm ice (frozen WTB) -22°C
5	15 mm ice (frozen WTB) -22°C

A six cycle Hanning pulse with four different excitation frequencies (20 kHz, 30 kHz, 50 kHz, and 100 kHz) were conducted for each scenario. 500 experiments under each excitation frequency were carried out. Figure 2 shows the signals for each scenario at 50 kHz.

Fig. 2 Set of signals for different scenarios, collected at 50 kHz.

2.2 Novel approach in WTB ice detection.

Figure 3 shows the flowchart for the novel approach presented in this paper.

Fig. 3 Flowchart for the SHM

The signals are firstly filtered and de-noised by Wavelet Transforms [29,51,52]. Sections 3.1 and 3.2 describe the FE of each ultrasonic signal extracted by linear (PCA and ARX) and non-linear models (h-NLPCA and NARX). Section 3.4 presents the FS by NCA. The pattern recognition was made by ML by supervised learning. The results were compared and analysed by: DT, Linear Discriminant Analysis, SVM, KNN, and Ensemble Classification. The confusion matrix was employed to evaluate the classification, resulting in a high level of precision.

3. METHODS

3.1 Linear Features

PCA and AR by Yule-Walker are used for FE in this paper. It has been applied in SHM in a large number of cases. PCA is a technique for mapping multidimensional data into lower dimensions with minimal loss of information [53]. Let $\mathbf{Y}$ a $n \times m$ matrix, where n is the

number of observations and m is the number of variables. PCA is an optimal factorization of $\mathbf{Y}$ into two matrices, $\bar{\mathbf{T}}^{n \times f}$ called score matrix, and $\bar{\mathbf{P}}^{m \times f}$ named loadings matrix, plus a matrix of residuals $\bar{\mathbf{E}}$, where f is the number of factors ($f < m$). Linear PCA is expressed by equation (1):

$$\mathbf{Y} = \bar{\mathbf{T}}\bar{\mathbf{P}}^T + \bar{\mathbf{E}} \quad (1)$$

PCA can also be expressed as a linear mapping of data from $\mathcal{R}^m$ to $\mathcal{R}^f$ given by equation (2):

$$\bar{\mathbf{T}}_j = \mathbf{Y}_j \bar{\mathbf{P}} \quad (2)$$

where $\mathbf{Y}_j$ is the j^{th} observation of $\mathbf{Y}$, $\bar{\mathbf{T}}_j$ is the corresponding row of $\bar{\mathbf{T}}$ and $\bar{\mathbf{P}}$ is the loading matrix that contains the coefficients for the transformation. The information lost due to transformation in this mapping can be evaluated by the measurement vector reconstruction by reversing the projection back to $\mathcal{R}^m$ by equations (3) and (4):

$$\mathbf{Y}'_j = \bar{\mathbf{T}}_j \bar{\mathbf{P}} \quad (3)$$

where $\mathbf{Y}'_j$ is the j^{th} observation of $\mathbf{Y}$ in $\mathcal{R}^m$. The residual error vector $\bar{\mathbf{E}}_j$ is given by equation (4):

$$\bar{\mathbf{E}}_j = \mathbf{Y}_j - \mathbf{Y}'_j \quad (4)$$

Figure 4 shows a flowchart of the linear approach.

Fig 4. Flowchart Linear Method

3.2 Nonlinear features

The nonlinear feature is extracted from the NARX and h-NLPCA models. The model extraction procedure is presented in this section, together with the FS using NCA. The flowchart of the nonlinear approach is showed in Figure 5.

Fig 5. Flowchart Non-Linear Method

3.2.1 NARX model

The nonlinear function in NARX model is done by a polynomial function. The polynomial NARX model with an order up to 3 is given by equation (5) [54].

$$\begin{aligned}
 y(t) = & \sum_{i_1=1}^M y(t-i_1) \theta_{i_1} + \sum_{i_1=1}^M \sum_{i_2=i_1}^M y(t-i_1) y(t-i_2) \theta_{i_1 i_2} \\
 & + \sum_{i_1=1}^M \dots \sum_{i_l=i_{l-1}}^M y(t-i_2) \dots y(t-i_l) \theta_{i_1 \dots i_l} + \xi(t)
 \end{aligned} \tag{5}$$

where $y(t)$ are data samples at time t , $t=1, \dots, N$, and $y(t-i_1)$ are the delayed data with lagged i_1 , $i_1 = 1, \dots, r$. The coefficient θ is the feature. Given a set of N training samples, equation (5) can be expressed in the matrix form as equation (6),

$$\mathbf{y} = \mathbf{P}\mathbf{\Theta} + \mathbf{\Xi} \tag{6}$$

where $\mathbf{y} = [y(1), \dots, y(N)]^T$ is the coefficient vector, $\boldsymbol{\theta} = [\theta_1, \dots, \theta_M]^T$ is the unknown parameter vector, $\boldsymbol{\xi} = [\xi(1), \dots, \xi(N)]^T$ is the residual vector, and $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_i, \dots, \mathbf{p}_M]$ is a N-by-M matrix, being $\mathbf{p}_i = [\mathbf{p}_i(x(1), v_j), \dots, \mathbf{p}_i(x(N), v_j)]^T$ [27].

The NARX coefficients given in equation (5) are obtained as:

Step 1. The matrix of first-order terms ($\sum_{i_1=1}^M x_{i_1}(t) \theta_{i_1}$) is

$$\mathbf{A}(M) = \begin{pmatrix} y(t) & y(t-1) & \cdots & y(t-M) \\ \vdots & \vdots & & \vdots \\ y(N-1) & y(N-2) & \cdots & y(N-M) \end{pmatrix} \quad (7)$$

Step 2. The matrix of second-order terms ($\sum_{i_1=1}^M \sum_{i_2=i_1}^M x_{i_1}(t) x_{i_2}(t) \theta_{i_1 i_2}$) is:

$$\mathbf{B}(M) = \begin{pmatrix} y(t)y(t-1) & \cdots & y(t)y(t-M) \\ y(t+1)y(t) & \cdots & y(t+1)y(t-M) & y(t)y(t) \\ \vdots & & \vdots & \\ y(N-1)y(N-2) & \cdots & y(N-1)y(N-M) & y(N-2)y(N-M) & \cdots & y(N-M)y(N-M) \end{pmatrix} \quad (8)$$

Step 3. For matrix of N variables, each term is the product of the N values $\sum_{i_1=1}^M \cdots \sum_{i_l=i_{l-1}}^M x_{i_1}(t) \cdots x_{i_l}(t) \theta_{i_1 \dots i_l}$.

Step 4. $\mathbf{P}$ matrix consists of the matrices involved ($\mathbf{A}$, $\mathbf{B}$)

Step 5. The solution of equation (6) is given by equation (9).

$$\boldsymbol{\theta} = \mathbf{P}^{-1} \mathbf{y} \quad (9)$$

where $\boldsymbol{\theta} = [\theta_1, \dots, \theta_M]^T$ is the extraction features parameter vector.

3.2.2 h-NLPCA

NLPCA is an auto-associative neural network based on the approach presented by Kramer [55,56]. It can be subdivided into two variations: normal NLPCA and hierarchical. The mapping into feature space is generalized to allow arbitrary nonlinear functions. Mapping in NLPCA is done by equation (10), similar to equation (2):

$$\bar{\mathbf{T}}_j = \bar{\mathbf{G}} \mathbf{Y}_j \quad (10)$$

where $\bar{\mathbf{G}}$ is nonlinear vector, composed of f individual nonlinear functions; $\bar{\mathbf{G}} = \{\bar{\mathbf{G}}_1, \bar{\mathbf{G}}_2, \dots, \bar{\mathbf{G}}_f\}$, analogous to the columns of $\bar{\mathbf{P}}$ in linear PCA.

Equation 10 can also be expressed in terms of the i^{th} factors by equation (11):

$$\bar{\mathbf{T}}_{j,i} = \bar{\mathbf{G}}_i \mathbf{Y}_j \quad (11)$$

where $\bar{\mathbf{T}}_{j,i}$ are the i^{th} elements of $\bar{\mathbf{T}}_j$, $\bar{\mathbf{G}}_i$ is the i^{th} nonlinear factor of the j^{th} observation.

the inverse transformation will be performed by nonlinear function $\bar{\mathbf{H}} = \{\bar{\mathbf{H}}_1, \bar{\mathbf{H}}_2, \dots, \bar{\mathbf{H}}_n\}$ in equation (12), similar to equation (3):

$$\mathbf{Y}'_j = \bar{\mathbf{H}} \bar{\mathbf{T}}_j \quad (12)$$

Figure 6 shows a two feed-forward neural network. The network is composed of five layers, three of which are hidden layers (mapping, bottle-neck and de-mapping), and one an input (first layer) and one an output layer (fifth layer). This architecture is designed to estimate nonlinear vector functions $\bar{\mathbf{G}}$ (mapping layer) and vector functions $\bar{\mathbf{H}}$ (de-mapping layer). The bottle-neck layer resolves the features $\bar{\mathbf{T}}_{j,f}$.

Fig. 6 Network layout for NLPCA.

The number of mapping nodes is related to the complexity of the nonlinear functions that can be generated by the network [57]. This network could provide low accuracy if the number of nodes is lower, and this may lead to overfitting if there are many mapping nodes [58]. Kramer proposes to restrict the number of weights in the network to a fraction of the number of restrictions imposed by the data set. A constrain is imposed by each output node for each data vector. The number of parameters that can be adjusted must

be less than $j \times n$. The number of adjustable parameters is given by equation (13) assuming all nodes have biases for the whole network:

$$(j + f + 1)(M_1 + M_2) + j + f \quad (13)$$

where M_1 and M_2 are the number of nodes in the mapping and demapping layers, implying the inequality by equation (14):

$$M_1 + M_2 \ll \frac{j(n - f)}{j + f + 1} \quad (14)$$

Equation (14) can be expressed by equation (15) for a relatively small number of factors ($f \ll j, n$):

$$M_1 + M_2 \ll n \quad (15)$$

Scholz and Vigário propose h-NLPCA that extends PCA into NLPCA through a hierarchical type of learning [59]. The h-NLPCA algorithm performs an identity mapping, where the output nonlinear $\mathbf{Y}'_{j,n}$ is obliged to be equal to the input nonlinear $\mathbf{Y}_{j,n}$. The aim is achieved by minimizing the Squared Reconstruction Error (SRE), defined by equation (16):

$$SRE = \frac{1}{jn} \sum_{j=1}^j \sum_{n=1}^n (\mathbf{Y}_{j,n} - \mathbf{Y}'_{j,n})^2 \quad (16)$$

while all components are treated equally by the NLPCA algorithm, i.e. there is no particular order. The h-NLPCA is ordered hierarchically, where its first n components indicate the maximum variance of the data that are included in an n -dimensional subspace.

The number of features is an unknown integer variable estimated from the dataset for AR and NARX. The Akaike Information Criterion (AIC) is employed in this paper because it is one of the most efficient techniques for order optimization [28]. The AIC is a measure of the goodness-of-fit of an estimated statistical model, based on the trade-off between fitting accuracy and the number of estimated parameters. AIC is given by equation (17):

$$AIC = N \log \left(\det \left(\frac{1}{N} \sum_{t=1}^N \boldsymbol{\varepsilon}(t, \hat{\theta}_N) \left(\boldsymbol{\varepsilon}(t, \hat{\theta}_N) \right)^T \right) \right) + 2n_p + N \left(n_y (\log(2\pi) + 1) \right) \quad (17)$$

where n_p is the number of estimated parameters, N is the number of predicted data points, n_y is the number of model outputs, $\boldsymbol{\varepsilon}$ is a $n_y - b_y - 1$ vector of the prediction errors, and θ_N represents the estimated parameters.

3.2.3 Feature Selection

NCA method has been applied in order to reduce dimensionality in AR and NARX [44,60,61]. Let $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n\}$ be the labelled features set, where $\mathbf{y}_i \in \mathbb{R}^d$ and the corresponding class labels are $\{\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n\}$.

The objective is to find a vector of weighting $\mathbf{w}$ to select the subset of characteristics. optimizing the classification of its nearest neighbours. W is the weighted distance between two samples y_i and y_j , given by equation (18):

$$D_w(y_i, y_j) = \sum_{l=1}^d w_l^2 |y_{il} - y_{jl}| \quad (18)$$

where w_l is a weight associated with l^{th} feature.

The reference point in NCA is chosen randomly, where all points in $\mathbf{X}$ could be selected as the reference point. The point x_i selects the point x_j as its neighbour with a probability p_{ij} , receiving the class label from the neighbour selected. The probability p_{ij} is defined by equation (18):

$$\begin{cases} p_{ij} = \frac{\mathcal{K}(D_w(y_i, y_j))}{\sum_{k \neq i} \mathcal{K}(D_w(y_i, y_k))}; \text{ if } i \neq j \\ p_{ij} = 0; \text{ if } i = j \end{cases} \quad (19)$$

where $\mathcal{K}(z) = \exp(-\frac{z}{\sigma})$ is a Kernel function, and the Kernel with σ is an input parameter that influences the probability of each point to be selected as the reference point. The probability of the query point $\mathbf{x}_i$ to be correctly classified is given by equation (19):

$$p_i = \sum_j C_{ij} p_{ij} \quad (20)$$

$C_{ij} = 1$ if the class label $C_i = C_j$

$C_{ij} = 0$ if the class label $C_i \neq C_j$

Consequently, the true leave-one-out classification accuracy $\xi(w)$ is given in equation (21):

$$\xi(w) = \sum_i \sum_j c_{ij} p_{ij} - \lambda \sum_{l=1}^d w_l^2 \quad (21)$$

where λ is a regularization parameter that can be adjusted through cross validation. The regularization term $\lambda \sum_{l=1}^d w_l^2$ is introduced in equation (21) to perform FS and avoid overfitting.

A gradient rule can be applied by differentiating $\xi(w)$ with respect to w_l by equation (22):

$$\frac{\partial \xi(w)}{\partial w_l} = 2 \left(\frac{1}{\sigma} \sum_i \left(p_i \sum_{j \neq i} p_{ij} |y_{il} - y_{jl}| - \sum_j c_{ij} p_{ij} |y_{il} - y_{jl}| - \lambda \right) w_l \right) \quad (22)$$

3.3 Classification methods

3.3.1 Introduction

ML has been employed to classify the results through supervised learning. The classifiers applied were:

Decision Trees is a classifier and is used because a complex decision process is segregated into simpler decision processes [62]. Nodes and branches represent a classification tree. Each node symbolizes a decision about one of the FE, where it has two or more branches. The model is constructed from training input functions, where each branch stands for a result of the DT process. The classification rules are crucial for setting the path from the root of the tree root to its leaf-nodes. The advantage of this classifier is its flexibility and ability to be applied to nonlinear relationships between features and classes [63].

Discriminant Analysis is a common technique for multivariate analysis, based on finding the linear and quadratic combination of variables, optimising the classification of different classes by means of a discriminant function [64,65].

Support Vector Machines considers each feature as a point in a high-dimensional space, where the number of dimensions is equal to the number of rating levels[66]. An optimal hyperplane is defined to separate the dataset based on class membership. The Kernels function is used to transform the input dataset to a higher dimensional space, where a decision boundary can be designed. A training classifier is done to find w that maximizes the margin between the class boundary and the training patterns [67].

K-Nearest Neighbours classifier has been used for pattern classification and machine learning [68]. Given a set $\mathbf{X}$ of n points and a distance function, KNN search allows the k closest points in $\mathbf{X}$ to a query point, or set of points $\mathbf{Y}$, to be found. The advantage of this classifier as a search technique is its simplicity.

Ensemble Classification methods are learning algorithms that construct a set of classifiers whose individual decisions are combined in some way, generally by weighted or unweighted voting, to classify the set of features of each pattern [69]. Bagging, boosting, and random subspaces are general techniques that can be used with any type of base classifier.

3.3.2 Evaluation

The confusion matrix, has been employed to evaluate the classifications, obtaining the precision of classifier or tape rate, P , by equation (23),

$$P = tp/(tp + fn) \quad (23)$$

where tp are true positives correctly classified, and fn are false negatives in the confusion matrix.

There are two conventional methods for evaluating performance average across the two categories: macro-averaging and micro-averaging [70].

Macro-average (P_i^M): P^M is obtained by the average overall P_i^M , where M denotes Macro-average, and i is the scenario. Finally, they are calculated for each category, i.e. the precision value is evaluated locally P_i^M , and then globally P^M .

Micro-average (P_i^μ): The precision value is obtained as: i) tp_i , fp_i , fn_i are calculated for each scenario, where fp are false positives in the confusion matrix; ii) the value of tp , fp and fn are calculated as the sum of tp_i , fp_i , fn_i , and; iii) to apply the definition of the measure that corresponds to it.

3.3.3 Comparative analysis of classifiers

The Friedman test, recommended by Demšar [71], together with its modification made by Garcia and Herrera [72], are employed to perform the comparative analysis of different classifiers. The Friedman test assumes that all classifiers can achieve the same performance with the same classification accuracy; this is called the null hypothesis [73]. The Bonferroni-Dunn test is used to determine any significant differences between the top-ranked classifiers [74]. The Holm test is applied to contrast the results [75].

4. RESULTS

The case study has been designed to perform pattern recognition classification, employing signals at the same frequency and different frequencies (multi-frequency).

4.1 Filtering and de-noising signal

The process of signal filtering and de-noising is required for the characteristics extraction because redundant and non-essential data are eliminated, e.g. noise. Figure 7 shows the denoised signals at 50 kHz for each experiment, employing Discrete Wavelet Transforms (DWT) at level 7.

Fig. 7 Set of filtering and de-noising signals for different scenarios, collected at 50 kHz.

4.2 Features Extraction

FE was carried out using the non-linear methodologies NARX and h-NLPCA. Linear methods have been used to verify and compare both methodologies. Figure 8 shows the first three components applying FE by the h-NLPCA method. The same number of main components is chosen to compare the performance of the h-NLPCA with the linear PCA. The set of patterns for the first three components of each state is shown in a three-dimensional graph.

The architecture of the network has been tested in different configurations, 5 nodes being the optimal, according to the accuracy of the results. When the number of mapping layers and demapping is small, the results of the prediction accuracy are worse, and if it is high the network is susceptible to overfitting.

Fig. 8 Three-dimensional graphical representation of the first three components by the NLPCA methodology, signals at 50 kHz.

The network architecture consists of 12 hidden nodes in the mapping and demapping layers, and 5 nodes in the bottleneck layer, corresponding to the number of components chosen according to Kramer's recommendations (equations 14 and 15). The number of features will be 5 for all models, the first 5 components being for PCA and h-NLPCA due to the hierarchy of the models. The number of features has been selected by AIC in the case of AR and NARX. The results can be seen in Figure 9.

Fig. 9 AIC Curve Features.

According to the results given by AIC (Figure 9), $p = 15$ has been set for the Yule Walker method, and $p = 30$ for NARX because the first 15 features are of the first order.

4.3 Features Selection

The main features for AR and NARX method have been selected by NCA, and the weights of the non-relevant features are zero. According to the results shown in Figure 10.A, the results given by NCA, the relevant feature numbers for AR FS are: 2, 7, 9, 11, 12, 13 and 15 (Figure 10.A). The main features are 1, 2, 9, 11, 13 and 30 for NARX (Figure 10.B).

Fig. 10 Features selection by NCA AR (A) and NARX (B) at 50 kHz.

4.4 Analysis of the results regarding measurement precision

4.4.1 Models set and assumptions

The model inputs are the feature selection, being the outputs five scenarios labelled. The main assumptions were: The cross-validation technique has been used against overfitting and stimulating accuracy; the maximum number of splits in DT set the fine, medium and coarse tree models, being 100, 20 and 4 respectively. The index split has been set by the GDI criterion.

The full covariance structure has been applied for linear and quadratic discriminant. The linear, polynomial and gaussian radial basis Kernel functions have been used to modify the different Kernel parameters in SVM. The Sequential Minimal Optimization (SMO) method has been applied to set the training algorithm and find the separating hyperplane. The tolerance and the fraction of variables applied to Karush-Kuhn-Tucker conditions have been also employed for the SMO training method. The algorithm automatically centres the data points at the mean and scales them to a unit standard deviation before training. The box constraint level is used to control the maximum penalty to the

observations to prevent excessive adjustment (regulation). The best box constraint level found was 1. The Kernel scale parameter is selected based on a heuristic procedure. For Gaussian Radial Basis Kernel Classifiers have been employed manual Kernel scale, being 0.56 for FGSVM, 2.2 for MGSVM and 8.9 for COGSVM. One against one method has been implemented for multiclass classification.

Euclidean distance metric has been used in KNN classifiers, being 1,10 and 100 the neighbours for FKNN, MKNN and COKNN. Cosine distance metric, cubic distance metric by Minkowski and square inverse distance weigh are the distance options for CSKNN, CKNN and WKNN.

The numbers of learners in EBT and ESD were 30, while the maximum number of split were 1999 and 20 for EBT and ESD respectively. The subspace dimension was 3 for ESD and ESKNN and learning rate is 0.1 for all ensembles classifiers.

4.4.2 Analysis at each frequency

The AR, NARX, PCA and h-NLPCA methods are compared for each frequency with different classifiers and for each scenario. The results are shown in Table II. The row indicates a classifier, grouped into five main blocks of classifiers: DT, DA, SVM, KNN and Ensemble Classification. The columns show the accuracy of all scenarios and for each FE method at the frequency of the first row.

Figure 11 shows a comparative analysis of the FE. The average accuracy is more than 90.9 % for all cases. The best accuracy is obtained by Quadratic Support Vector Machine (QSVM), 99.6%. h-NLPCA improves the results of PCA in most of the cases, and NARX to the AR results. The best results are at 50 KHz.

TABLE II Results Precision all scenarios

	<i>f=20 KHz</i>				<i>f=30 KHz</i>				<i>f=50 KHz</i>				<i>f=100 KHz</i>				Average success classifiers
	AR	NARX	PCA	h-NLPCA	AR	NARX	PCA	h-NLPCA	AR	NARX	PCA	h-NLPCA	AR	NARX	PCA	h-NLPCA	
DECISION TREE																	
Complex Tree (CT)	0.978	0.958	1.000	0.998	0.956	0.926	1.000	1.000	0.982	0.984	1.000	1.000	0.944	0.962	1.000	1.000	0.981
Medium Tree (MT)	0.978	0.958	1.000	0.998	0.956	0.926	1.000	1.000	0.982	0.984	1.000	1.000	0.944	0.962	1.000	1.000	0.981
Simple Tree (ST)	0.956	0.726	1.000	0.998	0.914	0.772	1.000	1.000	0.952	0.962	1.000	1.000	0.930	0.902	1.000	1.000	0.946
DISCRIMINANT ANALYSIS																	
Linear DA (LDA)	0.940	0.838	1.000	1.000	0.810	0.808	1.000	0.980	0.956	0.966	1.000	1.000	0.912	0.886	1.000	1.000	0.947
Quadratic DA (QDA)	0.948	0.882	1.000	0.998	0.874	0.904	1.000	1.000	0.970	0.984	1.000	1.000	0.936	0.924	1.000	1.000	0.965
SUPPORT VECTOR MACHINES																	
Linear SVM (LSVM)	0.998	0.996	1.000	1.000	0.974	0.950	1.000	1.000	0.988	0.994	1.000	1.000	0.990	0.992	1.000	1.000	0.993
Quadratic SVM (QSVM)	0.992	0.994	1.000	1.000	0.998	0.972	1.000	0.998	0.998	0.996	1.000	1.000	0.990	0.996	1.000	1.000	0.996
Cubic SVM (CSVM)	0.996	0.994	1.000	1.000	0.988	0.954	1.000	0.998	1.000	0.996	1.000	1.000	0.988	0.996	0.980	1.000	0.993
Fine Gaussian SVM (FGSVM)	0.982	0.990	0.988	0.994	0.976	0.942	0.960	0.988	0.996	0.984	0.966	1.000	0.978	0.972	1.000	1.000	0.982
Medium Gaussian SVM (MGSVM)	0.992	0.996	0.998	0.998	0.974	0.960	1.000	1.000	0.988	0.994	0.998	1.000	0.988	0.990	0.998	1.000	0.992
Coarse Gaussian SVM (COGSVM)	0.970	0.960	1.000	1.000	0.896	0.910	1.000	0.996	0.982	0.990	1.000	1.000	0.974	0.976	1.000	1.000	0.979
K-NEAREST NEIGHBOURS																	
Fine KNN (FKNN)	0.988	0.994	1.000	1.000	0.984	0.960	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.984	1.000	1.000	0.993
Medium KNN (MKNN)	0.988	0.980	1.000	1.000	0.954	0.938	0.994	0.998	0.988	0.988	0.998	1.000	0.984	0.984	0.996	1.000	0.987
Coarse KNN (COKNN)	0.918	0.828	0.968	0.990	0.698	0.698	0.806	0.910	0.952	0.960	0.954	1.000	0.876	0.858	1.000	0.928	0.898
Cosine KNN (CSKNN)	0.986	0.974	1.000	1.000	0.962	0.914	1.000	1.000	0.988	0.986	1.000	1.000	0.984	0.984	1.000	1.000	0.986
Cubic KNN (CKNN)	0.988	0.978	0.984	1.000	0.956	0.932	0.956	1.000	0.988	0.988	0.994	1.000	0.984	0.984	0.992	1.000	0.983
Weigthed KNN (WKNN)	0.992	0.992	1.000	1.000	0.976	0.964	1.000	1.000	0.996	1.000	0.998	1.000	0.998	0.988	1.000	1.000	0.994
ENSEMBLE CLASSIFICATION																	
Ensemble Bagged Trees (EBT)	0.988	0.984	1.000	1.000	0.976	0.964	1.000	0.994	0.994	0.998	1.000	1.000	0.980	0.984	1.000	1.000	0.991
Ensemble Subspace Discriminant (ESD)	0.968	0.914	1.000	1.000	0.922	0.860	1.000	0.996	0.996	0.962	1.000	1.000	0.964	0.964	1.000	1.000	0.972
Ensemble Subspace KNN (ESKNN)	0.988	0.986	1.000	1.000	0.974	0.926	1.000	1.000	1.000	1.000	1.000	1.000	0.988	0.982	1.000	1.000	0.990
Average success in each method all classifiers	0.977	0.946	0.997	0.999	0.936	0.909	0.986	0.993	0.985	0.986	0.995	1.000	0.966	0.964	0.998	0.996

Fig. 11 Bar graph results

The scripts have been developed in Matlab® 2017b. The training run times are shown in Table (III).

Table III. Training run time $f=50$ KHz

CLASSIFIER	Training run time (s)			
	AR	PCA	NARX	h-NLPCA
DECISION TREE (DT)				
Complex Tree (CT)	0.31237	0.24064	0.267280	0.25887
Medium Tree (MT)	0.099489	0.089818	0.098891	0.093025
Simple Tree (ST)	0.10249	0.094509	0.096468	0.093513
DISCRIMINANT ANALYSIS				
Linear DA (LDA)	0.10725	0.097699	0.10985	0.10432
Quadratic DA (QDA)	0.11785	0.10972	0.11803	0.11677
SUPPORT VECTOR MACHINES				
Linear SVM (LSVM)	0.55322	0.48512	0.49120	0.47714
Quadratic SVM (QSVM)	0.5294	0.53176	0.51552	0.48967
Cubic SVM (CSVM)	0.52246	0.47971	0.51628	0.51242
Fine Gaussian SVM (FGSVM)	0.57013	0.4893	0.55022	0.50574
Medium Gaussian SVM (MGSVM)	0.51553	0.46132	0.47309	0.50118
Coarse Gaussian SVM (COGSVM)	0.56037	0.46395	0.49189	0.55321
K-NEAREST NEIGHBOURS				
Fine KNN (FKNN)	0.13334	0.10443	0.11104	0.10664
Medium KNN (MKNN)	0.14968	0.10563	0.11082	0.10725
Coarse KNN (COKNN)	0.1276	0.11743	0.11554	0.13603
Cosine KNN (CSKNN)	0.13388	0.11577	0.11758	0.11416
Cubic KNN (CKNN)	0.22064	0.11939	0.12255	0.11477
Weighted KNN (WKNN)	0.12732	0.11122	0.11666	0.10836
ENSEMBLE CLASSIFICATION				
Ensemble Bagged Trees (EBT)	1.143	1.0238	1.1065	1.0577
Ensemble Subspace Discriminant (ESD)	1.1263	1.0645	1.0430	1.0031
Ensemble Subspace KNN (ESKNN)	1.2997	1.2175	1.2800	1.2972

4.4.3 Multi-frequency Analysis

Table IV shows the precision measure for all classifiers applied to a multi-frequency analysis. Discriminant Analysis classifiers provide lower accuracy than classifiers. In contrast, KNN obtain a high level of accuracy, with a success rate of more than 98%, where only Cosine KNN (CSKNN) provides a low percentage of 91.95%.

The first main column in Table IV shows the scenarios and precision parameters. The rest of the columns represent different classifiers, and each one is subdivided into four sub-columns. The first sub-column shows the FE results employing AR by Yule-Walker, the second the NARX results, and the third and fourth columns the precision of the PCA and h-NLPCA methods. The values of micro and macro precision accuracies are shown in the sixth and seventh rows. The coefficient given by Friedman tests and their classification are shown in the last rows.

TABLE IV. Results Precision all scenarios multi-frequency

	AR	NARX	PCA	h-NLPCA	Average success classifiers
DECISION TREE					
Complex Tree (CT)	0.960	0.944	0.999	0.999	0.9755
Medium Tree (MT)	0.908	0.877	0.999	0.999	0.9457
Simple Tree (ST)	0.699	0.720	0.745	0.699	0.7157
DISCRIMINANT ANALYSIS					
Linear DA (LDA)	0.635	0.650	0.550	0.502	0.5842
Quadratic DA (QDA)	0.650	0.661	0.760	0.791	0.7155
SUPPORT VECTOR MACHINES					
Linear SVM (LSVM)	0.870	0.899	0.870	0.897	0.8840
Quadratic SVM (QSVM)	0.980	0.970	1.000	1.000	0.9875
Cubic SVM (CSVM)	0.990	0.984	1.000	1.000	0.9935
Fine Gaussian SVM (FGSVM)	0.985	0.987	1.000	1.000	0.9930
Medium Gaussian SVM (MGSVM)	0.964	0.951	1.000	1.000	0.9787
Coarse Gaussian SVM (COGSVM)	0.845	0.862	0.850	0.800	0.8392
K-NEAREST NEIGHBOURS					
Fine KNN (FKNN)	0.987	0.985	1.000	1.000	0.9930
Medium KNN (MKNN)	0.978	0.969	1.000	1.000	0.9867
Coarse KNN (COKNN)	0.835	0.843	1.000	1.000	0.9195
Cosine KNN (CSKNN)	0.972	0.965	1.000	1.000	0.9842
Cubic KNN (CKNN)	0.975	0.969	1.000	1.000	0.9860
Weighted KNN (WKNN)	0.986	0.980	1.000	1.000	0.9915
ENSEMBLE CLASSIFICATION					
Ensemble Bagged Trees (EBT)	0.980	0.972	1.000	0.999	0.9877
Ensemble Subspace Discriminant (ESD)	0.980	0.741	0.528	0.469	0.6795
Ensemble Subspace KNN (ESKNN)	0.895	0.974	1.000	1.000	0.9672
Average success in each method all classifiers	0.9037	0.89515	0.91505	0.90775	

The Demšar recommendations have been employed to identify the significant differences between the classifiers when ranking them. The test results for the significant differences between classifiers and for each experiment show that the Friedman test for all methods did not reject the null hypothesis ($p\text{-value} = 1,09\text{e-}09 \leq 0,05$). The Bonferroni-Dunn test, see Figure 12, rejects the null hypothesis for $p \leq 0,05$, with $\alpha = 0.05$ for the classifier 5 (QDA for AR), 6 (QDA for NARX) and 8 (QDA for h-NLPCA). The Holm test rejects the null hypothesis for the classifiers 2-4.

Fig. 12 Bonferroni-Dunn test for all methods with $\alpha=0.05$.

The results for AR and NARX indicate that the classifier with more precision is QSVM (Table V). It is observed that all classifiers by the PCA method obtain the best results. Finally, the test indicates that FKNN and QSVM classifiers are the best classifiers for h-NLPCA and PCA.

TABLE V RESULTS PRECISION

	DECISION TREE COMPLEX				QDA				QSVM				FINE KNN				ENSEMBLE BAGGED TREE			
	P_t (AR)	P_t (NARX)	P_t (PCA)	P_t (h-NLPCA)	P_t (AR)	P_t (NARX)	P_t (PCA)	P_t (h-NLPCA)	P_t (AR)	P_t (NARX)	P_t (PCA)	P_t (h-NLPCA)	P_t (AR)	P_t (NARX)	P_t (PCA)	P_t (h-NLPCA)	P_t (AR)	P_t (NARX)	P_t (PCA)	P_t (h-NLPCA)
1	1,0000	0,9975	1,0000	0,9925	0,9925	0,9900	1,0000	0,9901	0,9926	0,9950	1,0000	1,0000	0,9926	0,9876	1,0000	1,0000	0,9926	0,9975	1,0000	1,0000
2	0,9075	0,9050	1,0000	0,2350	0,2350	0,3975	1,0000	0,7500	0,9725	0,9650	1,0000	1,0000	0,9550	0,9650	1,0000	1,0000	0,9375	0,9375	1,0000	0,9950
3	0,9400	0,9500	0,9925	0,6975	0,6975	0,6350	1,0000	0,5125	0,9825	0,9700	1,0000	1,0000	0,9825	0,9850	1,0000	1,0000	0,9825	0,9675	1,0000	0,9975
4	0,9575	0,8975	1,0000	0,9125	0,9125	0,9150	1,0000	0,7500	0,9950	0,9875	1,0000	1,0000	0,9950	0,9800	1,0000	1,0000	0,9825	0,9625	1,0000	1,0000
5	0,9925	0,9700	1,0000	0,4150	0,4150	0,3675	1,0000	0,9450	1,0000	1,0000	1,0000	1,0000	1,0000	0,9975	1,0000	1,0000	1,0000	0,9950	1,0000	1,0000
P^μ	0,960	0,944	0,999	0,999	0,650	0,661	0,760	0,791	0,980	0,970	1,000	1,000	0,987	0,985	1,000	1,000	0,980	0,972	1,000	0,999
P^M	0,960	0,944	0,999	0,999	0,650	0,661	0,760	0,791	0,980	0,970	1,000	1,000	0,987	0,985	1,000	1,000	0,980	0,972	1,000	0,999
RANKING	6,90	4,90	14,90	14,40	2,60	2,20	16,00	2,20	10,10	9,50	16,00	16,00	9,50	7,10	16,00	16,00	8,70	6,80	16,00	14,20
CLASSIFIC.	9	11	2	3	12	13	1	13	5	6	1	1	6	8	1	1	7	10	1	4

Table VI shows the training run times

Table VI. Multifrequency: Training run time

CLASSIFIER	Training run time (s)			
	AR	PCA	NARX	h-NLPCA
DECISION TREE (DT)				
Complex Tree (CT)	0.28475	0.24852	0.28783	0.25358
Medium Tree (MT)	0.11333	0.10124	0.11142	0.10054
Simple Tree (ST)	0.10705	0.10173	0.10832	0.10562
DISCRIMINANT ANALYSIS				
Linear DA (LDA)	0.10296	0.097479	0.097389	0.10035
Quadratic DA (QDA)	0.1106	0.12441	0.11385	0.13059
SUPPORT VECTOR MACHINES				
Linear SVM (LSVM)	0.87659	0.7236	0.85842	0.9797
Quadratic SVM (QSVM)	1.3317	0.63218	1.562	0.63315
Cubic SVM (CSVM)	4.7772	0.61657	3.1046	0.623
Fine Gaussian SVM (FGSVM)	0.83077	0.65262	0.86261	0.65006
Medium Gaussian SVM (MGSVM)	0.73478	0.68321	0.77566	0.71435
Coarse Gaussian SVM (COGSVM)	0.8356	0.84961	0.89452	0.94119
K-NEAREST NEIGHBOURS				
Fine KNN (FKNN)	0.11832	0.1117	0.11301	0.11508
Medium KNN (MKNN)	0.13987	0.12502	0.12698	0.13726
Coarse KNN (CKNN)	0.16902	0.17791	0.17111	0.17556
Cosine KNN (CSKNN)	0.1568	0.15545	0.18048	0.15576
Cubic KNN (CKNN)	0.17589	0.15884	0.18016	0.15962
Weighted KNN (WKNN)	0.13133	0.13229	0.12998	0.13231
ENSEMBLE CLASSIFICATION				
Ensemble Bagged Trees (EBT)	1.7012	1.2088	1.5687	1.2641
Ensemble Subspace Discriminant (ESD)	1.3176	1.1575	1.1526	1.1166
Ensemble Subspace KNN (ESKNN)	1.3829	1.4705	1.3987	1.3559

4.5 Discussion.

Table II shows that the best results are found at 50 kHz in all classifiers and FEs. The best accuracy is 100% found by AR, NCA FS and FKNN classifiers, the general accuracy ranging from 95.2% to 100%. 100% accuracy is achieved for NARX FE, NCA FS and the FKNN, WKNN and ESKNN classifiers, i.e. the non-linear NARX provides the best results by KNN. The NARX FE accuracy, from 96% to 100%, is better than AR FE. Most classifiers provide 100% accuracy in FE PCA techniques, 95.4% being the worst. All classifiers give 100% accuracy for the h-NLPCA. In conclusion, nonlinear FE provide better results than linear FE. The best classifier for FE and FS techniques is the Quadratic SVM with an average accuracy of 99.6%. The minimal accuracy found is 89.8%.

The best classification accuracy results were found at 50 KHz, being 100% in h-NLPCA FE and, KNN and SVM classifiers. Table IV shows that the best classifier is QSVM employing PCA and h-NLPCA FE to identify the levels of ice.

5. Conclusions

A new methodological approach presented in this paper has been applied to ultrasonic signals to classify distinct levels of ice in wind turbine blades. Signal processing employs twenty linear and non-linear classifiers of Machine Learning. A pre-processing signal was done by a pre-filter and de-noising by wavelet transforms. Features Extraction was performed using linear, Auto Regressive and Principal Component Analysis, and non-linear AR eXogenous and Hierarchical Non-Linear methods to reduce the dimensionality of the signal data. Feature Selection was tested in h-NLPCA with different configurations of components to avoid overfitting of the Neural Network employed. Feature Selection was made by Neighbourhood Component Analysis for Auto Regressive and non-linear AR eXogenous. Different scenarios were tested. The classification was performed by comparing the signals of each scenario at the same frequency and at multi-frequency mode, showing that:

1. This novel methodological approach obtains ice detection results in different scenarios with excellent prediction accuracy, achieving in many cases 100%, in both single frequency and multi-frequency cases.
2. In general, with all Feature Extraction methods, 89.5% is obtained as the mean results of prediction accuracy in the worst case, and 91.50% in the best. This margin is narrow due to the linear proportionality of the ice thickness and the temperature.
3. The non-linear classifiers Support Vector Machine and K-Nearest Neighbours provide an accuracy better than 95%.

Acknowledgement

The work reported herewith has been financially by the Spanish Ministerio de Economía y Competitividad, under Research Grant Ref.: DPI2015-67264-P.

References

1. Zhang, S. International competitiveness of china's wind turbine manufacturing industry and implications for future development. *Renewable and Sustainable Energy Reviews* **2012**, *16*, 3903-3909.
2. Afshar, K.; Ghiasvand, F.S.; Bigdeli, N. Optimal bidding strategy of wind power producers in pay-as-bid power markets. *Renewable Energy* **2018**, *127*, 575-586.
3. Zanon, A.; De Gennaro, M.; Kühnelt, H. Wind energy harnessing of the nrel 5 mw reference wind turbine in icing conditions under different operational strategies. *Renewable Energy* **2018**, *115*, 760-772.
4. Karlsson, T. Iea task 19: Cold climate wind power market study 2015-2020. http://winterwind.se/wp-content/uploads/2015/08/9_1_24_Karlsson_IEA_Task_19_-_Cold_climate_wind_power_market_study_2015-2020_Pub_v1.pdf (10 November 2017),
5. Márquez, F.P.G.; Pérez, J.M.P.; Marugán, A.P.; Papaelias, M. Identification of critical components of wind turbines using fta over the time. *Renewable Energy* **2016**, *87*, 869-883.
6. Pliego Marugán, A.; García Márquez, F.P.; Lev, B. Optimal decision-making via binary decision diagrams for investments under a risky environment. *International Journal of Production Research* **2017**, 1-16.
7. Rahimi, H.; Schepers, G.; Shen, W.Z.; García, N.R.; Schneider, M.; Micallef, D.; Ferreira, C.S.; Jost, E.; Klein, L.; Herráez, I. Evaluation of different methods for determining the angle of attack on wind turbine blades with cfd results under axial inflow conditions. *arXiv preprint arXiv:1709.04298* **2017**.
8. Jaunet, V.; Braud, C. Experiments on lift dynamics and feedback control of a wind turbine blade section. *Renewable Energy* **2018**.
9. Zuo, Y.; Montesano, J.; Singh, C.V. Assessing progressive failure in long wind turbine blades under quasi-static and cyclic loads. *Renewable Energy* **2017**.
10. Meng, H.; Lien, F.-S.; Li, L. Elastic actuator line modelling for wake-induced fatigue analysis of horizontal axis wind turbine blade. *Renewable Energy* **2018**, *116*, 423-437.
11. Lee, J.-K.; Park, J.-Y.; Oh, K.-Y.; Ju, S.-H.; Lee, J.-S. Transformation algorithm of wind turbine blade moment signals for blade condition monitoring. *Renewable energy* **2015**, *79*, 209-218.
12. Haselbach, P.U.; Bitsche, R.; Branner, K. The effect of delaminations on local buckling in wind turbine blades. *Renewable Energy* **2016**, *85*, 295-305.
13. Rezaei, M.M.; Behzad, M.; Moradi, H.; Haddadpour, H. Modal-based damage identification for the nonlinear model of modern wind turbine blade. *Renewable Energy* **2016**, *94*, 391-409.
14. Shajee, S.; Pao, L.Y.; Wagner, P.N.; Moore, E.D.; McLeod, R.R. In *Direct ice sensing and localized closed-loop heating for active de-icing of wind turbine blades*, American Control Conference (ACC), 2013, 2013; IEEE: pp 634-639.
15. Pérez, J.M.P.; Márquez, F.P.G.; Hernández, D.R. Economic viability analysis for icing blades detection in wind turbines. *Journal of Cleaner Production* **2016**, *135*, 1150-1160.
16. Márquez, F.P.G.; Tobias, A.M.; Pérez, J.M.P.; Papaelias, M. Condition monitoring of wind turbines: Techniques and methods. *Renewable Energy* **2012**, *46*, 169-178.
17. Jiménez, A.A.; Muñoz, C.Q.G.; Márquez, F.P.G. Dirt and mud detection and diagnosis on a wind turbine blade employing guided waves and supervised learning classifiers. *Reliability Engineering & System Safety* **2018**.
18. Jiménez, A.A.; Muñoz, C.Q.G.; Márquez, F.P.G. Machine learning for wind turbine blades maintenance management. *Energies* **2017**, *11*, 1-16.
19. Pliego Marugán, A.; García Márquez, F.P.; Pinar Pérez, J.M. Optimal maintenance management of offshore wind farms. *Energies* **2016**, *9*, 46.
20. Kraj, A.G.; Bibeau, E.L. Phases of icing on wind turbine blades characterized by ice accumulation. *Renewable Energy* **2010**, *35*, 966-972.
21. Muñoz, C.Q.G.; Márquez, F.P.G.; Tomás, J.M.S. Ice detection using thermal infrared radiometry on wind turbine blades. *Measurement* **2016**, *93*, 157-163.
22. Fortin, G.; Perron, J. In *Wind turbine icing and de-icing*, 47th AIAA Aerospace Sciences Meeting including The New Horizons Forum and Aerospace Exposition, 2009; p 274.
23. Zhao, X.; Rose, J.L. Ultrasonic guided wave tomography for ice detection. *Ultrasonics* **2016**, *67*, 212-219.
24. Gao, H.; Rose, J.L. Ice detection and classification on an aircraft wing with ultrasonic shear horizontal guided waves. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control* **2009**, *56*, 334-344.
25. Gómez Muñoz, C.Q.; García Márquez, F.P. A new fault location approach for acoustic emission techniques in wind turbines. *Energies* **2016**, *9*, 40.

26. Gómez Muñoz, C.Q.; Arcos Jimenez, A.; García Marquez, F.P.; Kogia, M.; Cheng, L.; Mohimi, A.; Papaelias, M. Cracks and welds detection approach in solar receiver tubes employing electromagnetic acoustic transducers. *Structural Health Monitoring* **2017**, 1475921717734501.
27. Muñoz, C.Q.G.; Marquez, F.P.G.; Lev, B.; Arcos, A. New pipe notch detection and location method for short distances employing ultrasonic guided waves. *Acta Acustica united with Acustica* **2017**, 103, 772-781.
28. Tang, J.; Soua, S.; Mares, C.; Gan, T.-H. An experimental study of acoustic emission methodology for in service condition monitoring of wind turbine blades. *Renewable Energy* **2016**, 99, 170-179.
29. Muñoz, C.Q.G.; Jiménez, A.A.; Márquez, F.P.G. Wavelet transforms and pattern recognition on ultrasonic guides waves for frozen surface state diagnosis. *Renewable Energy* **2017**.
30. Habibi, H.; Cheng, L.; Zheng, H.; Kappatos, V.; Selcuk, C.; Gan, T.-H. A dual de-icing system for wind turbine blades combining high-power ultrasonic guided waves and low-frequency forced vibrations. *Renewable Energy* **2015**, 83, 859-870.
31. Goraj, Z. In *An overview of the deicing and anti-icing technologies with prospects for the future*, 24th International Congress of the Aeronautical Sciences, 2004.
32. Zeng, J.; Song, B. Research on experiment and numerical simulation of ultrasonic de-icing for wind turbine blades. *Renewable Energy* **2017**, 113, 706-712.
33. Wang, Z.; Xu, Y.; Su, F.; Wang, Y. A light lithium niobate transducer for the ultrasonic de-icing of wind turbine blades. *Renewable Energy* **2016**, 99, 1299-1305.
34. Papaelias, M.; Cheng, L.; Kogia, M.; Mohimi, A.; Kappatos, V.; Selcuk, C.; Constantinou, L.; Muñoz, C.Q.G.; Marquez, F.P.G.; Gan, T.-H. Inspection and structural health monitoring techniques for concentrated solar power plants. *Renewable Energy* **2016**, 85, 1178-1191.
35. Martinez-Luengo, M.; Kolios, A.; Wang, L. Structural health monitoring of offshore wind turbines: A review through the statistical pattern recognition paradigm. *Renewable and Sustainable Energy Reviews* **2016**, 64, 91-105.
36. Esu, O.O.; Lloyd, S.D.; Flint, J.A.; Watson, S.J. Feasibility of a fully autonomous wireless monitoring system for a wind turbine blade. *Renewable Energy* **2016**, 97, 89-96.
37. Song, Z.; Zhang, Z.; Jiang, Y.; Zhu, J. Wind turbine health state monitoring based on a bayesian data-driven approach. *Renewable Energy* **2018**, 125, 172-181.
38. Xiao, B.; Yin, S.; Gao, H. Reconfigurable tolerant control of uncertain mechanical systems with actuator faults: A sliding mode observer-based approach. *IEEE Transactions on Control Systems Technology* **2017**.
39. Chong, J.W.; Kim, Y.; Chon, K.H. Nonlinear multiclass support vector machine-based health monitoring system for buildings employing magnetorheological dampers. *Journal of Intelligent Material Systems and Structures* **2014**, 25, 1456-1468.
40. Farrar, C.R.; Worden, K. *Structural health monitoring: A machine learning perspective*. John Wiley & Sons: 2012.
41. Oh, C.K.; Sohn, H. Damage diagnosis under environmental and operational variations using unsupervised support vector machine. *Journal of Sound and Vibration* **2009**, 325, 224-239.
42. Gomez Munoz, C.; De la Hermosa Gonzalez-Carrato, R.; Trapero Arenas, J.; Garcia Marquez, F. A novel approach to fault detection and diagnosis on wind turbines. *Global Nest Journal* **2014**, 16, 1029-1037.
43. Sierra-Pérez, J.; Torres-Arredondo, M.A.; Güemes, A. Damage and nonlinearities detection in wind turbine blades based on strain field pattern recognition. Fbgs, obr and strain gauges comparison. *Composite Structures* **2016**, 135, 156-166.
44. Yang, W.; Wang, K.; Zuo, W. Neighborhood component feature selection for high-dimensional data. *JCP* **2012**, 7, 161-168.
45. Mahmoud, T.; Dong, Z.; Ma, J. An advanced approach for optimal wind power generation prediction intervals by using self-adaptive evolutionary extreme learning machine. *Renewable Energy* **2018**, 126, 254-269.
46. Regan, T.; Canturk, R.; Slavkovsky, E.; Niezrecki, C.; Inalpolat, M. In *Wind turbine blade damage detection using various machine learning algorithms*, ASME 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2016; American Society of Mechanical Engineers: pp V008T010A040-V008T010A040.
47. Skrimpas, G.A.; Kleani, K.; Mijatovic, N.; Sweeney, C.W.; Jensen, B.B.; Holboell, J. Detection of icing on wind turbine blades by means of vibration and power curve analysis. *Wind Energy* **2015**.
48. Lee, S.; Im, J.; Kim, J.; Kim, M.; Shin, M.; Kim, H.-c.; Quackenbush, L.J. Arctic sea ice thickness estimation from cryosat-2 satellite data using machine learning-based lead detection. *Remote Sensing* **2016**, 8, 698.

49. García Márquez, F.P.; Muñoz, G.; Quiterio, C.; Papelias, M.; Arcos Jiménez, A. A heuristic method for detecting and locating faults employing electromagnetic acoustic transducers. **2015**.
50. Foder, M.H. In *Iso 12494" atmospheric icing of structures" and how to use it*, The Eleventh International Offshore and Polar Engineering Conference, 2001; International Society of Offshore and Polar Engineers.
51. de la Hermosa Gonzalez, R.R.; Márquez, F.P.G.; Dimlaye, V.; Ruiz-Hernández, D. Pattern recognition by wavelet transforms using macro fibre composites transducers. *Mechanical Systems and Signal Processing* **2014**, *48*, 339-350.
52. de la Hermosa González, R.R.; Márquez, F.P.G.; Dimlaye, V. Maintenance management of wind turbines structures via mfcs and wavelet transforms. *Renewable and Sustainable Energy Reviews* **2015**, *48*, 472-482.
53. Márquez, F.P.G. A new method for maintenance management employing principal component analysis. *Structural Durability & Health Monitoring* **2010**, *6*, 89-99.
54. Zhang, L.; Li, K.; Bai, E.-W.; Irwin, G.W. Two-stage orthogonal least squares methods for neural network construction. *IEEE transactions on neural networks and learning systems* **2015**, *26*, 1608-1621.
55. Kramer, M.A. Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal* **1991**, *37*, 233-243.
56. Eybpoosh, M.; Berges, M.; Noh, H.Y. In *Nonlinear feature extraction methods for removing temperature effects in multi-mode guided-waves in pipes*, SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring, 2015; International Society for Optics and Photonics: pp 94371W-94371W-94316.
57. García Márquez, F.P.; Pliego Marugán, A.; Pinar Pérez, J.M.; Hillmanssen, S.; Papaelias, M. Optimal dynamic analysis of electrical/electronic components in wind turbines. *Energies* **2017**, *10*, 1111.
58. Jiménez, A.A.; Muñoz, C.Q.G.; Márquez, F.P.G. In *Machine learning and neural network for maintenance management*, International Conference on Management Science and Engineering Management, 2017; Springer, Cham: pp 1377-1388.
59. Scholz, M.; Vigário, R. In *Nonlinear pca: A new hierarchical approach*, ESANN, 2002; pp 439-444.
60. Hinton, G.; Goldberger, J.; Roweis, S.; Salakhutdinov, R. In *Neighborhood components analysis*, Proc. NIPS, 2004; pp 513-520.
61. Zhou, H.; Chen, J.; Dong, G.; Wang, H.; Yuan, H. Bearing fault recognition method based on neighbourhood component analysis and coupled hidden markov model. *Mechanical Systems and Signal Processing* **2016**, *66*, 568-581.
62. Loh, W.-Y.; Shih, Y.-S. Split selection methods for classification trees. *Statistica sinica* **1997**, 815-840.
63. Pal, M.; Mather, P.M. An assessment of the effectiveness of decision tree methods for land cover classification. *Remote sensing of environment* **2003**, *86*, 554-565.
64. Gaudenzi, P.; Nardi, D.; Chiappetta, I.; Atek, S.; Lampani, L.; Pasquali, M.; Sarasini, F.; Tirilló, J.; Valente, T. Sparse sensing detection of impact-induced delaminations in composite laminates. *Composite Structures* **2015**, *133*, 1209-1219.
65. Nardi, D.; Lampani, L.; Pasquali, M.; Gaudenzi, P. Detection of low-velocity impact-induced delaminations in composite laminates using auto-regressive models. *Composite Structures* **2016**, *151*, 108-113.
66. Boser, B.E.; Guyon, I.M.; Vapnik, V.N. In *A training algorithm for optimal margin classifiers*, Proceedings of the fifth annual workshop on Computational learning theory, 1992; ACM: pp 144-152.
67. Milgram, J.; Cheriet, M.; Sabourin, R. In *"One against one" or "one against all": Which one is better for handwriting recognition with svms?*, Tenth International Workshop on Frontiers in Handwriting Recognition, 2006; Suvisoft.
68. Michalski, R.S.; Carbonell, J.G.; Mitchell, T.M. *Machine learning: An artificial intelligence approach*. Springer Science & Business Media: 2013.
69. Dietterich, T.G. In *Ensemble methods in machine learning*, International workshop on multiple classifier systems, 2000; Springer: pp 1-15.
70. Yang, Y. An evaluation of statistical approaches to text categorization. *Information retrieval* **1999**, *1*, 69-90.
71. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research* **2006**, *7*, 1-30.
72. Garcia, S.; Herrera, F. An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. *Journal of Machine Learning Research* **2008**, *9*, 2677-2694.
73. Conover, W.J.; Iman, R.L. Rank transformations as a bridge between parametric and nonparametric statistics. *The American Statistician* **1981**, *35*, 124-129.

- 1 74. Blair, C.R.; Hochberg, Y. Improved bonferroni procedures for testing overall and pairwise
2 homogeneity hypotheses. *Journal of Statistical Computation and Simulation* **1995**, *51*, 281-289.
3 75. Holm, S. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*
4 **1979**, 65-70.

5