Elastic Net Multinomial Logistic Regression for Fault Diagnostics of on-board Aeronautical Systems

F. Cannarile[1], M. Compare[1,2], P. Baraldi[2], G. Diodati[3], V. Quaranta[3], E. Zio[1,2,4,5]

[1]Aramis s.r.l., Milano, Italy

[2]Dipartimento di Energia, Politecnico di Milano, Italy

[3]CIRA, Italian Aerospace Research Centre, Capua, Italy

[4]MINES ParisTech, PSL Research University, CRC, Sophia Antipolis, France

[5]Eminent Scholar, Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Republic of Korea

Abstract

The objective of this work is the development of a fault diagnostic system for a shaker blower used in on-board aeronautical systems. Features extracted from condition monitoring signals and selected by the ELastic NET (ELNET) algorithm, which combines l^1-penalty with the squared l^2-penalty on model parameters, are used as inputs of a Multinomial Logistic regression (MLR) model. For validation, the developed approach is applied to experimental data acquired on a shaker blower system (as representative of aeronautical on-board systems) and on three additional experimental datasets of literature. The satisfactory diagnostic performances obtained show the potential of the method for developing sound diagnostic classifiers from a very large set of features, even when only few training data are available.

Keywords: Fault Diagnostics, Shaker blower, Aerospace industry, ELastic NET (ELNET), Discrete Fourier Transform (DFT), Multinomial Logistic Regression (MLR).

NOMENCLATURE

ELNET: ELastic NET

MLR: Multinomial Logistic Regression

PHM: Prognostics and Health Management

SMOS: Smart On-Board Systems

CIRA: Italian Aerospace Research Centre

ECS: Empirical Classification System

DFT: Discrete Fourier Transform

LASSO: Least Absolute Shrinkage Selection Operator

CM: Condition Monitoring

FCV: Fold Cross Validation

F: number of extracted features

N: number of available training examples

C: number of equipment type of defects

S: number of sensors

M: number of training units

f_s: sampling rate of sensor s

f: sampling rate common to all S sensors

$x_s^m(\tau)$: τ^{th} sample recorded from the s^{th} sensor on the m^{th} training unit

T_s^m: time of the last recorded sample from the s^{th} sensor on the m^{th} training unit

T^m: time of the last recorded sample common to all S sensors on the m^{th} training unit

$\boldsymbol{x}_s^m$: vector in $\mathbb{R}^{T^m}$containing all T^m samples from the s^{th} sensor on the m^{th} training unit

L: number of samples in a time window

N^m: number of time windows extracted from signal recorded from sensor s on the m^{th} training unit

$\lfloor\cdot\rfloor$: floor functiom

$\boldsymbol{x}_m^{i_m}$: i_m^{th} training example obtained from windowing the signals recorded from the m^{th} training unit

$\boldsymbol{x}_s^{i_m}$: i_m^{th} time window of length L extracted from signal recorded from sensor s on the m^{th} training unit

D: training data

$\boldsymbol{x}^i$: i^{th} training examples

y^i: class of the i^{th} training example

C: number of different classes

y^{test}: class of a new test example

$\boldsymbol{x}^{test}$: test example

$\boldsymbol{x}_s^{test}$: signal of length L recorded from sensor s on the test example

$\boldsymbol{x}_s^i$: signal of length L recorded from sensor s on the i^{th} training example

$X_s^i[k]$: k^{th} DFT coefficient

$Z_{s,k}^i$: absolute value of $X_s^i[k]$

$\tilde{L}$: number of different harmonics extracted from a real signal of length L

f^k: k^{th} frequency

Δf_{all}: frequency resolution

$\boldsymbol{z}^i$: vector of DFT amplitudes extracted from the i^{th} training example

$\boldsymbol{Z}_s^i$: vector of DFT amplitudes resulting by applying DFT to signal recorded from sensor s on the i^{th} training example

$\widetilde{D}$: training data after performing feature extraction

$\beta_{0,c}$: bias term entering in the definition of the probability of being in class c

$\boldsymbol{\beta}_c$: vector of model parameter of size F entering in definition of the probability of being in class c

r_{ic}: binary variable equal to 1 if the i^{th} training example belongs to class c, 0 otherwise

ω_i: weight for the i^{th} training example

N_{y_i}: number of training examples of ground truth class y_i

α: free parameter that determines the importance of the l_1-norm penalty and l_2-norm penalty

P_1: LASSO penalty

P_2: ridge penalty

t: regularization parameter

$\|\cdot\|_1$: l_1- norm of a vector

$\|\cdot\|_2$: l_2- norm of a vector

λ: regularization parameter

$\boldsymbol{z}^{test}$: vector of DFT amplitudes extracted from the test example

$\hat{y}^{test}$: predicted class of test example $\boldsymbol{z}^{test}$

α_a: a^{th} candidate value for free parameter α

A: number of candidate values for free parameter α

Λ: number of candidate values for regularization parameter λ

λ_{min}: smallest value considered for regularization parameter λ

λ_{max}: largest value considered for regularization parameter λ

ϵ: ration between λ_{max} and λ_{min}

G: number of folds

D_g: g^{th} fold

I_g: index set of examples belonging to the g^{th} fold

λ_u: u^{th} candidate value for regularization parameter λ_u

$\mathrm{CVD_g}(\lambda_\mathrm{u})$: cross-validated deviance computed on the g^{th} fold setting $\lambda = \lambda_u$

$\mathrm{MCVD}(\lambda_u)$: mean cross-validated deviance setting $\lambda = \lambda_u$

α_{sel}: selected value of free parameter α

λ_{sel}: selected value of regularization parameter λ

1. INTRODUCTION

Predictive maintenance is expected to reduce aircraft maintenance costs up to 30-40%, while increasing their availability up to 35%, with savings up to 700 million Euros per year in Europe [1], [2]. These bright expectations justify the increasing interest of the aerospace industry in Prognostics and Health Management (PHM) for on-board systems. In particular, the Italian PRO.R.A. program (PROgramma nazionale di Ricerche Aerospaziali) financed the SMOS (Smart On-Board Systems) project at the Italian Aerospace Research Centre (CIRA), which has among its goals the development of effective PHM technologies for on-board systems of aeronautical and aerospace vehicles.

This work is positioned at the early stages of the SMOS project, with the goal of investigating novel PHM approaches through the application to representative case studies. To do this, a blower used by

CIRA for environmental tests (i.e., an electromagnetic TIRA shaker) has been considered as first case study and an ad-hoc dataset has been built through experimental measurements, which serves as testbench to develop a fault diagnostic system for classifying the types of defects in blower components.

Generally speaking, fault diagnostics of industrial equipment is based on the development of an Empirical Classification System (ECS) [3]-[11] trained on condition monitoring signals to classify the types of defects.

With respect to diagnostics of rotating engines, such as the blowers and other aerospace capital systems, the current industrial practice generally uses Discrete Fourier Transform (DFT) to represent in the frequency domain the radial and axial vibration measurements provided by piezoresistive accelerometers, generally mounted on the bearings and/or on the stator casing (e.g., [12]-[13]). Other basis expansion methods such as discrete wavelet transform [14], Hilbert-Huang transform [15] are considered to this aim, although they are less frequently available than DFT in commercial software tools for vibration analysis.

The application of DFT to raw vibration signals, which are typically sampled at high sampling rate (e.g., in the magnitude order of kHz up to MHz), provides a very large set of amplitudes at different frequencies, which will be referred to as extracted features. Since many of them are irrelevant for fault diagnostics and can even negatively affect the ECS performance, DFT amplitude spectra undergo a feature selection step [16]-[19], which aims at finding the smallest subset of the extracted features yielding the largest classification accuracy.

In general, feature selection methods can be classified into three categories: filter, wrapper and embedding methods [18], [19]. Filter methods are used before the construction of the classification algorithm and rely on a numerical evaluation function, which estimates the expected classification performance of every extracted feature. Then, irrelevant and/or redundant features are discarded, whereas those with the largest expected performances are encoded in the classification model [18]. In wrapper approaches, the classification algorithm is used to map the possible subsets of extracted features onto a performance function including classification accuracy [17], [18]. The subset maximising this objective function is finally selected. Embedding methods directly learn the optimal subset of features during the training of the classification model itself. This is typically achieved by minimizing a loss function for classification, with a proper penalization/regularization term [18], [19].

The main disadvantage of filter methods lies in that they are generally less performing than wrapper and embedding methods [18], [19], as filters do not account for the effects of the selected feature

subset on the performance of the classification algorithm. Wrapper methods, however, are computationally less efficient than filters, since they require training a large number of classifiers. This becomes unfeasible for high dimensional classification problems (e.g., hundreds of features) [18], [19], in which wrapper methods can also lead to overfitting and high variance in the estimates of the ECS parameters [20]. Embedding methods provide a good compromise solution between filter and wrapper methods: they are less computationally intensive and less prone to overfitting than wrapper methods but, contrarily to filter methods, they interact with the classification model.

Among the available embedding feature selection methods ([18], [19]), the Least Absolute Shrinkage Selection Operator (LASSO [20]-[22]) has been used in a few applications related to condition monitoring. For example, a LASSO binary logistic regression model is developed in [23] to identify the influencing factors of outages in smart grids, i.e., covariates such as daily precipitation, minimum and maximum temperatures. In [24], a One-Vs-All (OVA) system of LASSO binary logistic regression models has been developed for diagnosing bearing faults. The main limitations of the LASSO in fault diagnostics applications are related to the degradation of its performances when *a*) there are many highly correlated features among those extracted from the signals and *b*) the number of extracted features, F, is much larger than that of available training examples N. With respect to *a*), it has been shown that if there is a group of variables among which pairwise correlations are very high, then the LASSO tends to select only one feature of the group [20]-[22]. With respect to *b*), when the number of extracted features, F, is much larger than that of available training examples N, the LASSO selects at most N features before it saturates [20]-[22]. This is a relevant issue for those fault diagnostics applications where few training data are available (e.g., safety-critical and high-values components of aerospace, nuclear and oil and gas industries characterized by very high reliability and unique or new designed components [25]). In these cases, the LASSO is not able to select the subset of relevant features maximizing the performance of the diagnostic ECS if its cardinality is larger than N.

For these reasons, in this work we resort to an embedding feature selection approach based on the ELastic NET (ELNET) regularization combined with a Multinomial Logistic Regression (MLR) classifier [20], [26]. The ELNET has been introduced in [26] to overcome the LASSO limitations discussed above and it has been proven to outperform the LASSO in terms of prediction accuracy on real case studies [20]-[22]. The ELNET combines the l_1-penalty of the LASSO with a squared l_2 penalty (also known as ridge penalty); this latter shrinks the coefficients of correlated features towards each other contrarily to the LASSO penalty which picks one of them and discard the others [26].

To the Authors' best knowledge, this is the first time that the ELNET-MLR embedding method is used for fault diagnostics, where the following characteristics makes it a breakthrough enhancement:

- The ELNET-MLR allows extracting a very large set of features while not requiring large computational times to train the classifier. This is fundamental in industrial applications where the available knowledge of the degradation and failure behaviour of the equipment under investigation is scarce and, thus, a widespread search within the extracted features is necessary to identify the relevant information for fault diagnostics.
- The results provided are easily understandable. This has a twofold justification: *i*) the ELNET-MLR is capable of drastically reducing the number of features; *ii*) the MLR classifier is linear, which allows directly interpreting the importance of a feature in the identification of a given defect type.
- The method can be used when few data are available, as the ELNET-MLR is less affected by overfitting than wrapper methods and does not suffer from the limitations of the LASSO.
- Although the focus of this work is on DFT basis expansion, the ELNET-MLR can be naturally extended to accommodate the use of other signal representations such as discrete wavelet transform and its variants [14], Hilbert-Huang transform [15], etc.

To be confident on the potential of the proposed approach for aerospace applications, the ELNET-MLR is validated also on three benchmark case studies concerning diagnostics on motor house bearings and gearboxes. Upon additional validation, the diagnostic methodologies developed for the shaker blower will be exploited to improve maintenance on other on-board systems of aerospace vehicles.

The remainder of the paper is organized as follows. Section 2 states the problem and describes the proposed diagnostic scheme. Section 3 provides details on the ELNET-MLR. Section 4 details the tuning of the parameters of the algorithms. In Section 5, we describe our approach to model assessment. The proposed methodology is validated on three different case studies of the literature in Section 6. In Section 7, the proposed method is applied to experimental data of a blower operated at CIRA plant. In Section 8, the methodology limitations and its possible developments are discussed. Finally, conclusions are drawn in Section 9.

2. PROBLEM STATEMENT

The problem of diagnosing the type of defect of an industrial equipment among C different possible types of defects is framed as a C-classes classification problem. We consider a component equipped

with S sensors collecting Condition Monitoring (CM) signals and we assume to have data recorded from M different components (training units) similar to that currently monitored.

Although in industrial applications each sensor is typically characterized by its own sampling rate $f_s, s = 1, \dots, S$, we assume for clarity of the notation that $f_s = f$ for all S sensors.

Let $x_s^m(\tau), \tau = 0, \dots, T_s^m$, be the τ^{th} sample recorded from the s^{th} sensor installed on the m^{th} training unit, where T_s^m denotes the index of the last recorded sample. For sake of clarity we also assume that $T_s^m = T^m$ for all S sensors.

Every signal $\boldsymbol{x}_s^m \in \mathbb{R}^{T^m}$ is partitioned into windows containing L samples. This way, we obtain $N^m = \lfloor T^m/L \rfloor$ training windows from the CM signal recorded from sensor s on component m, $s = 1, \dots, S$, $m = 1, \dots, M$.

Then, we can lump together the signals belonging to the same window from all the S sensors of a component $\boldsymbol{x}_m^{i_m} = \{\boldsymbol{x}_1^{i_m}, \dots, \boldsymbol{x}_s^{i_m}, \dots, \boldsymbol{x}_S^{i_m}\}$, $i_m = \{1, \dots, N^m\}$, where $\boldsymbol{x}_s^{i_m} = \left[x_s^{i_m}[0], \dots, x_s^{i_m}[L-1]\right] = \left[x_s^m[(i_m-1)\cdot L], \dots, x_s^m[(i_m-1)\cdot L + L - 1]\right] \in \mathbb{R}^L, s = 1, \dots, S, i_m = 1, \dots, N^m$.

Finally, all the $N = \sum_{m=1}^{M} N^m$ training examples are appended to form a unique training set of N input-output pairs $D = \left\{\boldsymbol{x}^i, y^i\right\}_{i=1,..N}$ where the input of the i-th example is

$$\boldsymbol{x}^i = \boldsymbol{x}_m^{i_m},\ m = \min\left\{j \middle| \sum_{k=1}^{j} N^k \geq i, j \leq M\right\}$$

$$i_m = \begin{cases} i - \sum_{k=1}^{m-1} N^k & m > 1 \\ i & \text{otherwise} \end{cases} \tag{1}$$

whereas the output $y^i \in \{1, \dots, C\}$ is the corresponding example class (i.e., a label of the type of defect). The final objective is the prediction of the class y^{test} of a new test example $\boldsymbol{x}^{test} = \{\boldsymbol{x}_1^{test}, \dots, \boldsymbol{x}_s^{test}, \dots, \boldsymbol{x}_S^{test}\}$, $\boldsymbol{x}_s^{test} \in \mathbb{R}^L, s = 1, \dots, L$, using the last L measurements collected from the test equipment.

3. PROPOSED METHODOLOGY

The first step in the development of the proposed diagnostic classifier is the extraction of relevant features from the raw signals $\boldsymbol{x}_s^i$. To this aim, according to the current industrial practice, we apply the Discrete Fourier Transform (DFT) to each signal $\boldsymbol{x}_s^i$, resulting in the following signal frequency-domain representation

$$X_s^i[k] = \frac{1}{L_s}\sum_{t=0}^{L-1} x_s^i[t] e^{-\frac{2\pi j t k}{L}} \quad k = 0, \dots, L-1 \tag{2}$$

Then, we consider the DFT amplitudes as candidate features to be given in input to the diagnostic classifier

$$Z_{s,k}^i = \left|X_s^i[k]\right| = \sqrt{\mathrm{Re}\left(X_s^i[k]\right)^2 + \mathrm{Im}\left(X_s^i[k]\right)^2} \quad k = 0, \dots, L-1 \tag{3}$$

where $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ denote the real and imaginary part of their complex number argument, respectively. Since the observed signals are real, i.e., $x_s^i[t] \in \mathbb{R}$, then $Z_{s,k}^i = Z_{s,L-k}^i$, whereby the DFT amplitude features $Z_{s,k}^i, k = 0, \dots, \tilde{L} = \frac{L}{2} - 1$, are considered, only. Notice that $Z_{s,k}^i$ is the amplitude corresponding to frequency $f^k = k \cdot \Delta f_{all}$, where $\Delta f_{all} = f_{all}/(2\tilde{L})$.

The matrix of signals $\boldsymbol{x}^i$, is therefore, transformed into a $F = S \cdot \tilde{L}$-dimensional vector of DFT amplitude features $\boldsymbol{z}^i = \left[\boldsymbol{Z}_1^i \cdots \boldsymbol{Z}_s^i \cdots \boldsymbol{Z}_S^i\right]$, obtained by sequentially appending the $\tilde{L}$-dimensional vectors $\boldsymbol{Z}_s^i = \left[Z_{s,0}^i, \dots, Z_{s,\tilde{L}-1}^i\right]$. Notice that in many industrial applications $F \gg N$, which entails that the number of training examples, N, is too small for an accurate estimate of the parameters of a classifier with in input all the extracted features. Further, it is known that many irrelevant features unnecessarily increase the complexity of the classification problem and can degrade the classification performance [7], [17]. Thus, the development of the diagnostic classifier is generally preceded by feature selection. To this aim, the ELNET shrinkage method is here combined with MLR. Figure 1 pictorially shows the proposed DFT and the ELNET-MLR based diagnostic system.

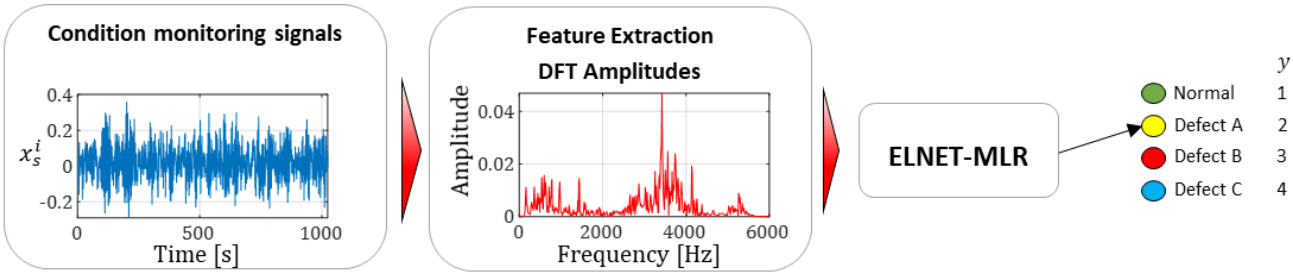

Figure 1. The proposed diagnostic system.

3.1 Elastic net multinomial logistic regression

In this Subsection, we briefly introduce the ELNET-MLR classification model. The interested reader is referred to [20]-[22] and [26] for further details.

Consider the training set $\widetilde{D} = \{\boldsymbol{z}^i, y^i\}_{i=1,\dots,N}$. The probability that observation $\boldsymbol{z}^i$ corresponds to class $c \in \{1, \dots, C\}$ is modelled as

$$\mathbb{P}(y^i = c|\boldsymbol{z}^i) = \frac{\exp(\beta_{0,c} + \boldsymbol{\beta}_c^T \boldsymbol{z}^i)}{\sum_{p=1}^{C} \exp(\beta_{0,p} + \boldsymbol{\beta}_p^T \boldsymbol{z}^i)} \tag{4}$$

where $\beta_{0,c}$ and $\boldsymbol{\beta}_c = [\beta_{c,1}, \dots, \beta_{c,F}] \in \mathbb{R}^F, c = 1, \dots, C,$ are unknown model parameters, whereas the denominator is for normalization. Notice that the overall number of parameters to be estimated is $C + C \cdot F$. The negative log-likelihood of the training data reads [20], [22],

$$l\left(\{\beta_{0,c}, \boldsymbol{\beta}_c\}_{c=1}^{C}\right) = -\frac{1}{N}\sum_{i=1}^{N} \omega_i \left[\sum_{c=1}^{C} r_{ic} \cdot (\beta_{0,c} + \boldsymbol{\beta}_c^T \boldsymbol{z}^i) - \log\left\{\sum_{c=1}^{C} \exp(\beta_{0,c} + \boldsymbol{\beta}_c^T \boldsymbol{z}^i)\right\}\right] \tag{5}$$

where r_{ic} is equal to 1 if $y^i = c$ and 0 otherwise, and a weight ω_i is associated to each example to account for an unbalanced training set (i.e., the number of training examples per class can be very different among different classes). In this work, we have resorted to a cost sensitive approach based on class weight balancing, where different weights are assigned to examples of different classes [27]. These have been set as ([27], [28])

$$\omega_i = \frac{N}{CN_{y_i}} \tag{6}$$

where N_{y_i} denotes the number of training examples of ground truth class $y_i \in \{1, \dots, C\}$. Notice that the i^{th} training example is given a weight inversely proportional to the number of instances belonging to its class and that $\sum_{i=1}^{N} \omega_i = N$. The parameters of the ELNET-MLR model [20], [22] are estimated by solving the following convex problem

$$\begin{gathered} \underset{\{\beta_{0,c}, \boldsymbol{\beta}_c\}_{c=1}^{C}}{\text{minimize}}\ l\left(\left\{\{\beta_{0,c}, \boldsymbol{\beta}_c\}_{c=1}^{C}\right\}\right) \\ \text{subject to } \alpha P_1\left(\left(\{\boldsymbol{\beta}_c\}_{c=1}^{C}\right)\right) + (1-\alpha) P_2\left(\left(\{\boldsymbol{\beta}_c\}_{c=1}^{C}\right)\right) \leq t \end{gathered} \tag{7}$$

where $P_1(\cdot)$ and $P_2(\cdot)$ denote the LASSO and the ridge penalties

$$\begin{gathered} P_1\left(\left(\{\boldsymbol{\beta}_c\}_{c=1}^{C}\right)\right) = \sum_{c=1}^{C} \|\boldsymbol{\beta}_c\|_1 \\ \|\boldsymbol{\beta}_c\|_1 = \sum_{e=1}^{F} |\beta_{c,e}| \end{gathered} \tag{8}$$

$$\begin{gathered} P_2\left(\left(\{\boldsymbol{\beta}_c\}_{c=1}^{C}\right)\right) = \sum_{c=1}^{C} \|\boldsymbol{\beta}_c\|_2^2 \\ \|\boldsymbol{\beta}_c\|_2^2 = \sum_{e=1}^{F} \left|\beta_{c,e}\right|^2 \end{gathered} \tag{9}$$

$t > 0$ is a free parameter that determines the amount of regularization and the parameter $\alpha \in [0,1]$ is a free parameter that determines the importance of the l_1-norm penalty P_1 (LASSO penalty) and l_2-norm penalty P_2 (ridge penalty).

By Lagrangian duality, there is a one-to-one correspondence between the constrained optimization problem in Eq. (7) and its Lagrangian form [20], [22]:

$$\underset{\{\beta_{0,c},\boldsymbol{\beta}_c\}_{c=1}^{C}}{\text{minimize}}\; l\left(\left\{\{\beta_{0,c},\boldsymbol{\beta}_c\}_{c=1}^{C}\right\}\right) + \lambda\left[\alpha P_1\left(\{\boldsymbol{\beta}_c\}_{c=1}^{C}\right) + (1-\alpha)P_2\left(\{\boldsymbol{\beta}_c\}_{c=1}^{C}\right)\right] \tag{10}$$

where $\lambda > 0$ is a user-specified regularization parameter: the smaller the value of λ, the larger the level of regularization imposed [20], [22]. From Eq. (10), the penalty applied to a single coefficient $\theta_{c,e}$ (without considering the regularization term $\lambda > 0$) is

$$\alpha\left|\beta_{c,e}\right| + (1-\alpha)\beta_{c,e}^2 \tag{11}$$

so that when $\alpha = 1$, it reduces to the l_1-norm or LASSO penalty, whereas when $\alpha = 0$, it reduces to the squared l_2-norm or ridge penalty. A geometrical interpretation of the ELNET penalty is provided in [20], [22]. To solve the regularization problem in Eq. (10), an efficient cyclical coordinate descent algorithm can be employed (see [20], [22], for further details).

Finally, the class (i.e., the type of defect) of a new test example $\boldsymbol{z}^{test}$ is predicted as

$$\hat{y}^{test} = \underset{c=1,..,C}{\text{argmax}}\; \frac{\exp\left(\widehat{\beta_{0,c}} + \widehat{\boldsymbol{\beta}_c^T}\boldsymbol{z}^{test}\right)}{\sum_{p=1}^{C}\exp\left(\widehat{\beta_{0,p}} + \widehat{\boldsymbol{\beta}_p^T}\boldsymbol{z}^{test}\right)} \tag{12}$$

where $\left\{\widehat{\beta_{0,c}},\widehat{\boldsymbol{\beta}_c}\right\}_{c=1}^{C}$ are the parameter estimates resulting from the minimization of the convex program in Eq. (10).

4. HYPERPARAMETER TUNING

The ELNET-MLR model encodes two hyperparameters, which must be properly tuned:

- the regularization parameter $\lambda > 0$ used in Eq. (10);
- the weights $\alpha \in [0,1]$ used in Eq (10).

For robust hyperparameter setting, we first consider a sequence α_a, $a = 1 \ldots, A$ of increasing values of α. For every α_a, we numerically solve Eq. (10) for a sequence of Λ values λ_u, uniformly spaced on the log-scale in the interval $\left[\lambda_{min},\lambda_{max}\right]$ [20], [22], where the value λ_{max} is set to be the largest

value for which the estimated vectors $\widehat{\boldsymbol{\beta}_c^{\lambda_{max}}} = \mathbf{0}, c = 1, .., C$, (i.e., such that the optimal solution is the all-zeroes vector) and $\lambda_{min} = \epsilon\lambda_{max}$, $\epsilon \in (0,1)$ [20], [22]. Then, we randomly partition the training set in G roughly equal-sized folds $D_g, g = 1, \dots, G$, with $\widetilde{D} = \bigcup_{g=1}^{G} D_g$. Let I_g be the index set of examples belonging to the g^{th} fold, i.e., $I_g = \{i: (\boldsymbol{x}^i, y^i) \in D_g\}$; we compute for each value of λ_u the Cross-Validated Deviance (CVD)

$$\mathrm{CVD}_g(\lambda_u) = -2 \sum_{i \in I_g} \log \hat{p}\,(y_i) \tag{13}$$

when fitting the model using the remaining $G - 1$ folds. In Eq. (13), $\hat{p}(y_i)$ is the estimated probability of example $i \in I_g$ being assigned to its ground truth class y_i. This step is repeated for $g = 1, \dots, G$ and the Mean Cross-Validated Deviance (MCVD) is computed [18], [20]:

$$\mathrm{MCVD}(\lambda_u) = \frac{1}{G} \sum_{g=1}^{G} \mathrm{CVD}_g(\lambda_u) \tag{14}$$

Then, we select the pair $(\alpha^{sel}, \lambda^{sel})$ such that

$$(\alpha^{sel}, \lambda^{sel}) = \underset{\substack{u=1,..,\Lambda \\ a=1,..,A}}{\operatorname{argmin}} \mathrm{MCVD}(\lambda_u) \tag{15}$$

5. MODEL ASSESSMENT

To assess the classification performance of the proposed method, we consider test data different from those used to develop the ELNET-MLR model and set the optimal hyperparameters $(\alpha^{sel}, \lambda^{sel})$ in a twice-nested cross validation procedure. The outer loop is used to compute the performance metrics, whereas the inner loop is used to set the optimal hyperparameters $(\alpha^{sel}, \lambda^{sel})$. The performance metrics considered are *i)* the mean classification accuracy (i.e., the mean percentage of examples correctly classified over the different test folds); *ii)* its standard deviation; *iii)* the minimum classification accuracy over the different test folds.

The twice-nested cross validation procedure has been applied in three different settings:

- *Setting 1* (S1): we randomly partition the available data into 5 folds; then, the outer loop is used to compute the overall classification accuracy (i.e., the ratio between the number of test examples correctly classified and the total number of test examples), whereas the inner loop is used to set the regularization hyperparameter (see Section 4). This is shown in Figure 2. At each iteration,

the training set contains the 80% all the available data. This is the setting typically used in the literature [20].

- *Setting* 2 (S2): we randomly partition the available data into 5 folds, containing approximatively 20% of all the available data each. Each single fold is used to train the classifier, which is tested on the remaining 4 folds (see Figure 3). In the inner loop, each training fold is further partitioned into 5 folds to set the regularization parameter (see Section 4).
- *Setting* 3 (S3): we randomly partition the available data into 20 folds, containing approximately 5% of all the available data each. Each single fold is used to train the classifier, which is tested on the remaining 19 folds (see Figure 3). In the inner loop, each training fold is further partitioned into 5 folds to set the regularization parameter (see Section 4).

Notice that in each setting the random sampling has been performed so that each (both inner and outer) fold approximately preserves the ratio among the number of examples in different classes

We explicitly highlight that differently from setting S1, where one fold is left out for testing, in settings S2 and S3 only one fold is used to train the classification model. Settings S2 and S3 are introduced to estimate the expected generalization error in the stressing case, where few data are available for model training.

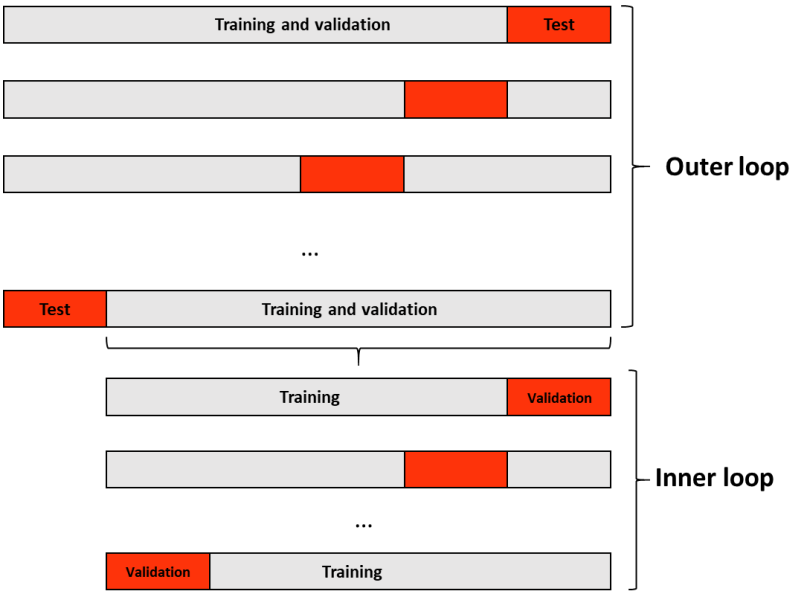

Figure 2: Twice nested 5- Fold Cross Validation (FCV): the inner loop is used to select the best regularization parameter λ, the outer loop is used to estimate classification accuracy (setting S1).

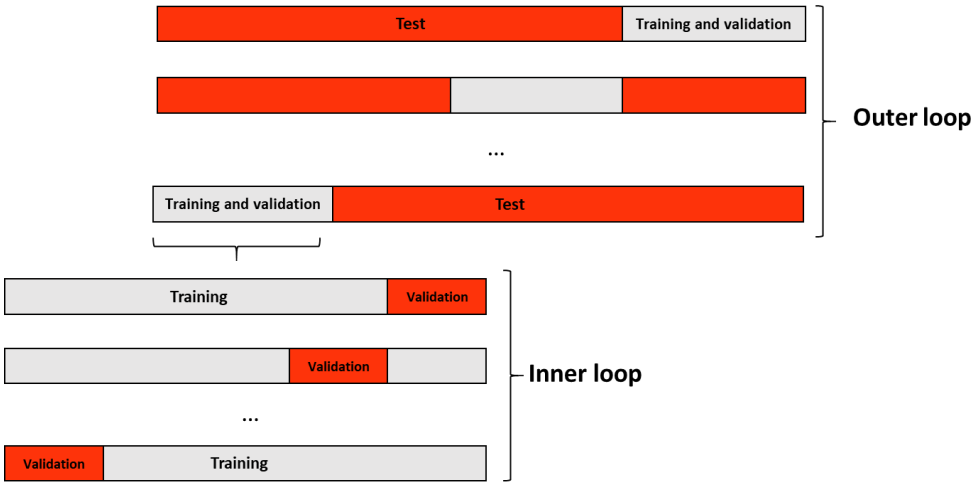

Figure 3: Twice-nested FCV, with single fold used for training on the outer loop (settings S2 and S3).

6. VALIDATION CASE STUDIES

The proposed ELNET-MLR diagnostic method has been verified with respect to data extracted from three benchmarks, before its application to the experimental data acquired on the CIRA blower. In all case studies, CM signals have been partitioned into non-overlapping windows of length L (whose value depends on the case study), then, DFT have been performed on each window resorting to Fast Fourier Transform (FFT) algorithm. This way, $\tilde{L}$ amplitude DFT amplitude features have been extracted from each signal window. Further, the extracted features have been scaled to have zero mean and unit variance (with respect to the training set). In cases where the sampling frequency changes among the different CM signals referring to the same physical quantity a down-sampling preprocessing step has been performed.

The setting of hyperparameters (α, λ), have been performed according to the procedure of Section 4 with $\alpha \in \{0.15, 0.25, 0.50, 0.75, 0.95, 1\}$, and λ assuming $\Lambda = 100$ different values λ_u uniformly spaced on log-scale starting from λ_{max}, which has been computed as discussed in [20], [22], to $\lambda_{min} = \epsilon \lambda_{max}$ with $\epsilon = 10^{-3}$.

All computations have been performed using an Intel Core i7-7770 CPU at 3.60 GHz processor with 8 GB RAM in MATLAB 2017b environment.

6.1. Bearing fault diagnostics: Case Western Reserve University bearing dataset

The Case Western Reserve University (CWRU) bearing dataset [29] contains vibrational signals collected during experimental tests on an engine with defective bearings. The test stand consists of a 2 hp motor, a torque transducer/encoder, a dynamometer, and control electronics. The test bearings support the motor shaft. Single point faults were introduced to the test bearings using electro-

discharge machining at four different motor speeds (1730, 1750, 1772, 1797 Round per Minute (RPM)). Table 1 reports the $C = 14$ types of considered defects. Although the dataset contains data collected with defects on both single-point drive end and fan end, we have only considered drive end bearing defects, as this allows comparing the proposed method with a larger number of works of literature. Vibration data were collected by $S = 2$ accelerometers, attached to the housing through magnetic bases. Accelerometers were placed at the 12 o'clock position on the drive end and fan end of the motor housing. Vibration signals were collected using a 16 channel DAT recorder at sampling frequency 12 KHz.

Bearing condition	**Fault Diameter**	**Label/Class**	**Number of Examples**
Normal	-	1	474
Inner Race Fault	7	2	476
Inner Race Fault	14	3	474
Inner Race Fault	21	4	475
Balls Fault	7	5	487
Balls Fault	14	6	474
Balls Fault	21	7	472
Outer race fault located at 6 o' clock (directly in the load zone)	7	8	474
Outer race fault located at 6 o' clock (directly in the load zone)	14	9	473
Outer race fault located at 6 o' clock (directly in the load zone)	21	10	475
Outer race fault located at 3 o'clock (orthogonal to the load zone)	7	11	475
Outer race fault located at 3 o'clock (orthogonal to the load zone)	21	12	475
Outer race fault located at 12 o'clock	7	13	474
Outer race fault located at 12 o'clock	21	14	476

Table 1: Label assigned to each bearing condition.

6.1.1 Feature Extraction

Each vibration signal is segmented using a fixed time window of approximately 0.086 seconds, which contains $L = 2^{10}$=1024 samples. Table 1, last column, reports the number of examples obtained for each class. The total number of available examples $(\boldsymbol{x}^i, y^i)$ is 6654. From each time window, $\tilde{L} = 512$, $s = 1, \ldots, 2$, DFT amplitude features have been extracted. Since these features are extracted from the signals measured by $S = 2$ different accelerometers, the total number of features extracted is $F = 1024$. Figure 4 shows the features (amplitude spectrum) extracted from an example in normal condition.

The number of MLR parameters to be estimated is $(F + 1) \times C = (1024 + 1) \times 14 = 14350$. This is larger than the number of training examples, which are 80% (S1 cross validation setting), 20% (S2

cross validation setting) and 5% (S3 cross validation setting) of the available data (i.e., roughly, 0.80*6654 ≈ 5323 (S1), 0.20*6654 ≈ 1330 (S2) and 0.05*6654 ≈ 332 (S3), respectively).

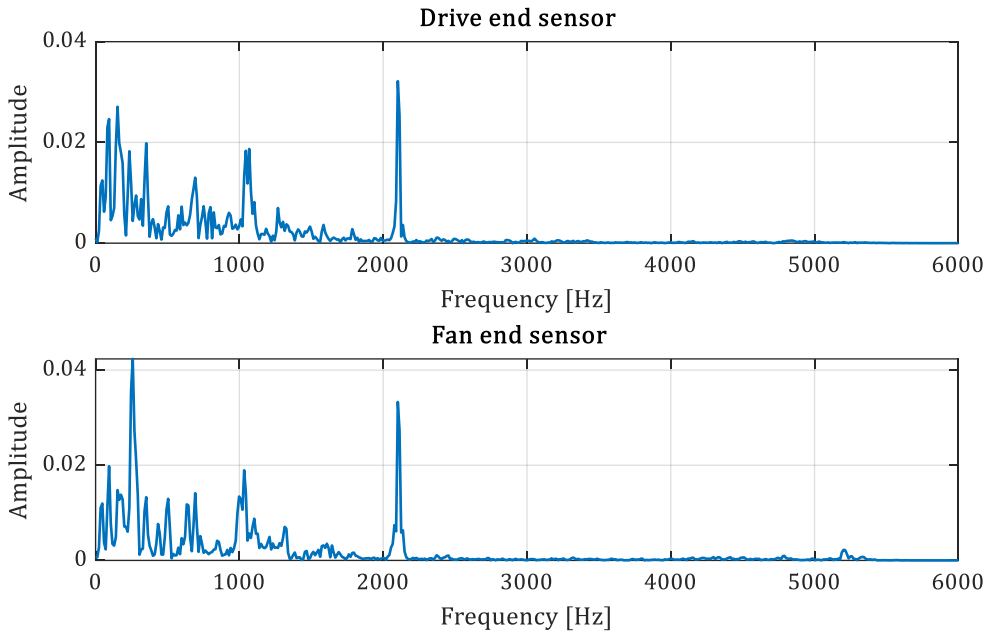

Figure 4: Extracted features from an example in normal condition (top, amplitude spectrum from drive end sensor; down, amplitude spectrum from fan end sensor)

6.1.2 Results

Table 2 compares the performance of the proposed method in settings S1, S2 and S3 (last three rows) with those obtained by other recent methods of the literature, which use supervised approaches with percentage of examples to train the classifier ranging in 20-75% of the total available set.

From Table 2, we can conclude that:

- Methods [30]-[32] build the ECS based on feature extraction and, then, feature dimensional reduction and/or feature selection. Comparing the results, we can see that the ELNET-MLR outperforms these methods, even when only 5% of the available data is used for classification model training.
- In [33]-[35], feature extraction feeds a Deep Neural Networks (DNN) architecture for building the ECS. In particular, [33]-[34] use DFT amplitude spectrum as in the present paper. The ELNET-MLR performs better than [33]-[34] even if we use only the 5% of the available data for classification model training.
- The methods in [36]-[37] do not require any feature extraction step; they aim at directly learning a good data representation from vibration signals. ELNET-MLR yields significantly

better performance than [36] and, particularly, than [37], where only 4 classes (instead of 14) are considered.

- The good performances of the ELNET-MLR in setting S3 highlight that this method is particularly suitable when few data are available. On the contrary, the effective training of deep architectures in the same setting is a very difficult task.

• Work	Method	Considered classes	Training examples (%)	Test accuracy (%)
[30]	Wavelet Kernel Local Fisher Discriminant Analysis (WKFDA)	1-2-3-4-5-6-7-8-9-10	40	98.80
[31]	Bi-spectrum based features combined with Support Vector Machines (SVM)	Normal, Inner Race Fault, Balls Fault, Outer Race fault at 3 o' clock (directly in the load zone) (i.e., classes 2-3-4, 5-6-7 and 8-9-10 are considered as a unique class)	50	96.98
[32]	Time, frequency, time-frequency domain features followed by Fisher Discriminant Analysis dimensional reduction and fuzzy ARTMAP neural network.	1-2-3-4-6-9	50	99.07
[33]	Deep Neural Networks fed by DFT amplitude spectrum	1-2-3-4-5-6-7-8-9-10	50	99.68 ± 0.22
[34]	Deep Output Kernel Learning (DOKL) machines fed by DFT amplitude spectrum	2-3-4-5-6-7-8-9	70	99.63
[35]	Hierarchical Deep Belief Networks fed by Wavelet Packet Transform (WPT) features.	1-2-3-4-5-6-7-8-9-10	50	99.03
[36]	Ensemble deep autoencoder	1-2-3-4-5-6-7-8-9-10-11-12	75	97.12 ± 0.11
[37]	Hierarchical Extreme Learning Machines (HELM)	Normal, Inner Race Fault, Balls Fault, Outer Race fault at 3 o' clock (directly in the load zone) (i.e., classes 2-3-4, 5-6-7 and 8-9-10 are considered as a unique class)	10	92.34
Proposed Method (S1)	ELNET-MLR fed by DFT amplitude features	1-2-3-4-5-6-7-8-9-10-11-12-13-14	80	**Mean:100.00 ± 0.00 Minimum: 100.00**
Proposed Method (S2)	ELNET-MLR fed by DFT amplitude features	1-2-3-4-5-6-7-8-9-10-11-12-13-14	20	**Mean: 99.98± 0.01 Minimum: 99.96**
Proposed Method (S3)	ELNET-MLR fed by DFT amplitude features	1-2-3-4-5-6-7-8-9-10-11-12-13-14	5	**Mean: 99.80± 0.18 Minimum: 99.12**

Table 2. Overall classification accuracy of different methods which use a supervised approach.

6.2. Bearing fault diagnostics II: the MFPT bearing dataset

The MFPT bearing dataset [38] contains vibration signals gathered from $C = 3$ different bearing defect types from which N =913 examples are extracted. Data of bearing in normal conditions were collected for six seconds at a sampling rate of 97656 Hz, with 270 lbs of load. Seven inner race faults vibrational data were collected for three seconds at sampling frequency of 48848 Hz, with varying loads 0, 50, 100, 140, 200, 250 and 300 lbs. Three outer race faults vibrational data were collected

for six seconds at sampling rate of 97656 Hz and seven additional outer race faults vibrational data were collected for three seconds at sampling rate of 48848 Hz with loads 25, 50, 100, 150, 200, 250, and 300 lbs [38]. The data recorded at sampling frequency of 97656 Hz are down-sampled by a factor of 2 to match data recorded at 48848 Hz. Table 3 reports the number of available examples for each class of defect.

Bearing condition	Label/Class	Number of Examples
Normal	1	213
Inner Race Fault	2	245
Outer Race Fault	3	458

Table 3: Label assigned to each bearing condition.

6.2.1 Feature Extraction

Each vibration signal is partitioned using a fixed time window of approximately 0.042 seconds, which contains $L = 2^{12} = 4096$ samples. Table 3, last column, reports the number of examples obtained from each class. The total number is 1016. From each time window, $F = \tilde{L} = 2048$ DFT amplitude features have been extracted. Figure 5 shows the extracted features (amplitude spectrum) from examples belonging to classes 1, 2 and 3.

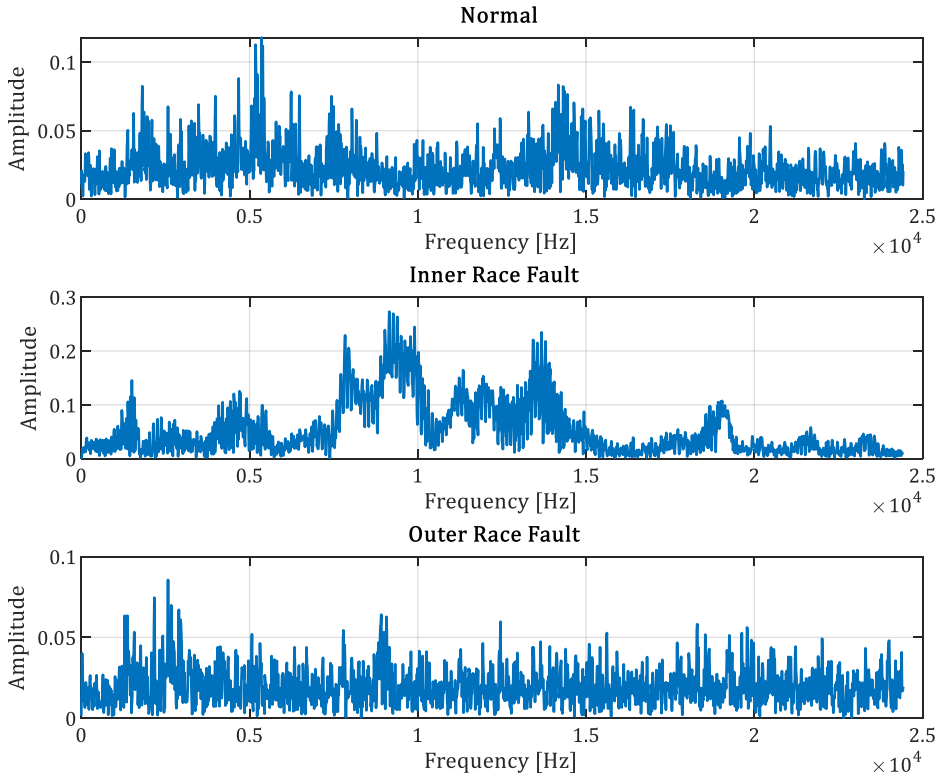

Figure 5: Extracted features from an example in class 1 (top), 2 (middle) and 3 (bottom).

The number of MLR parameters to be estimated is $(2048 + 1) \times 3 = 6147$. This is larger than the number of training examples, which are 80% (S1), 20% (S2) and 5% (S3) of the available data (i.e., roughly, 812, 203 and 50, respectively).

6.2.2. Results

Table 4 compares the performance of the proposed method in settings S1, S2 and S3 (last three rows) with those of the works available in the literature.

Work	Method	Training examples (%)	Test accuracy (%)
[39]	Convolution Neural Network (CNN) fed by Continuous Wavelet Transform (CWT) scalogram	Not declared	99.90
[40]	Convolutional Neural Network	Not declared	98.18
Proposed Method	ELNET-MLR fed by DFT amplitude features	80	**Mean: 100.00 ± 0.00 Minimum: 100.00**
Proposed Method	ELNET-MLR fed by DFT amplitude features	20	**Mean: 100.00 ± 0.00 Minimum: 100.00**
Proposed Method	ELNET-MLR fed by DFT amplitude features	5	**Mean: 99.07 ± 1.21 Minimum: 95.74**

Table 4. Overall classification accuracy of different recent methods which uses a supervised approach.

From Table 4, we can conclude that:

- when comparing the proposed method with [39], which is a supervised approach requiring a feature extraction step followed by the training of Convolutional Neural Network (CNN), our method performs slightly better if at least the 20% of the available data is used for classification model training. Notice that in [39] features are extracted from the Continuous Wavelet Transform (CWT) scalogram derived from vibrational signal. The considerations above about the simplicity and clarity of feature extraction in the ELNET apply to this case, as well.
- when comparing the proposed ELNET-MLR method with [40], which does not require any feature extraction step, our method provides superior performance.

Finally, notice that the classification accuracy is still very large when the percentage of the training size is 5% of the available data which corresponds to only roughly 50 examples available for training.

6.3. Gear-Box fault diagnostics

We consider the radial vibration signals provided in [41], which were acquired from a gearbox setup (see Figure 6) in three different gear conditions (Table 5): healthy, one chipped tooth, and three worn

teeth in helical gears. Table 5 reports the assigned labels to the gear box conditions. Acceleration signals were recorded for a duration of 10 seconds for each of the three gear box conditions, at sampling rate 10 KHz.

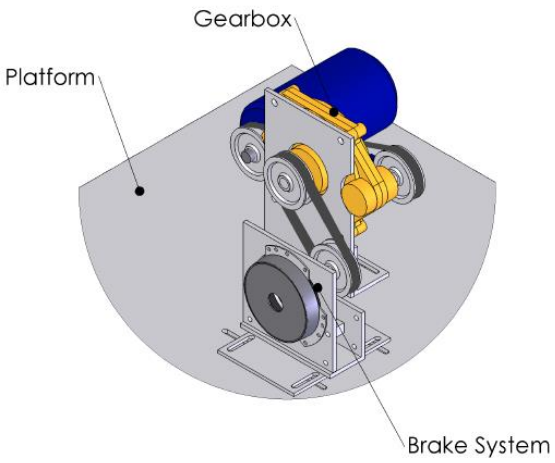

Figure 6: Schematic of the setup of the tested gear-box used in [37] (Image taken therein).

Gear-box condition	Label	Number of examples
Normal	1	97
One chipped tooth	2	97
Three worn teeth in helical gears.	3	97

Table 5: Label assigned to each gear-box condition.

6.3.1 Feature Extraction

Each vibration signal has been partitioned using a fixed time window of 0.1024 s, which corresponds to time windows of 2^{10} samples. Table 5, last column, reports the number of examples obtained for each class. The total number is 291. From each time window, $F = \tilde{L} = 512$ DFT amplitude features are extracted. Figure 7 shows the extracted features (amplitude spectrum) from an example belonging to classes 1, 2 and 3.

The number of MLR parameters to be estimated is $(512 + 1) \times 3 = 1539$. This is larger than the number of training examples, which are 80% (S1), 20% (S2) and 5% (S3) of the available data (i.e., roughly, 232, 58 and 15, respectively).

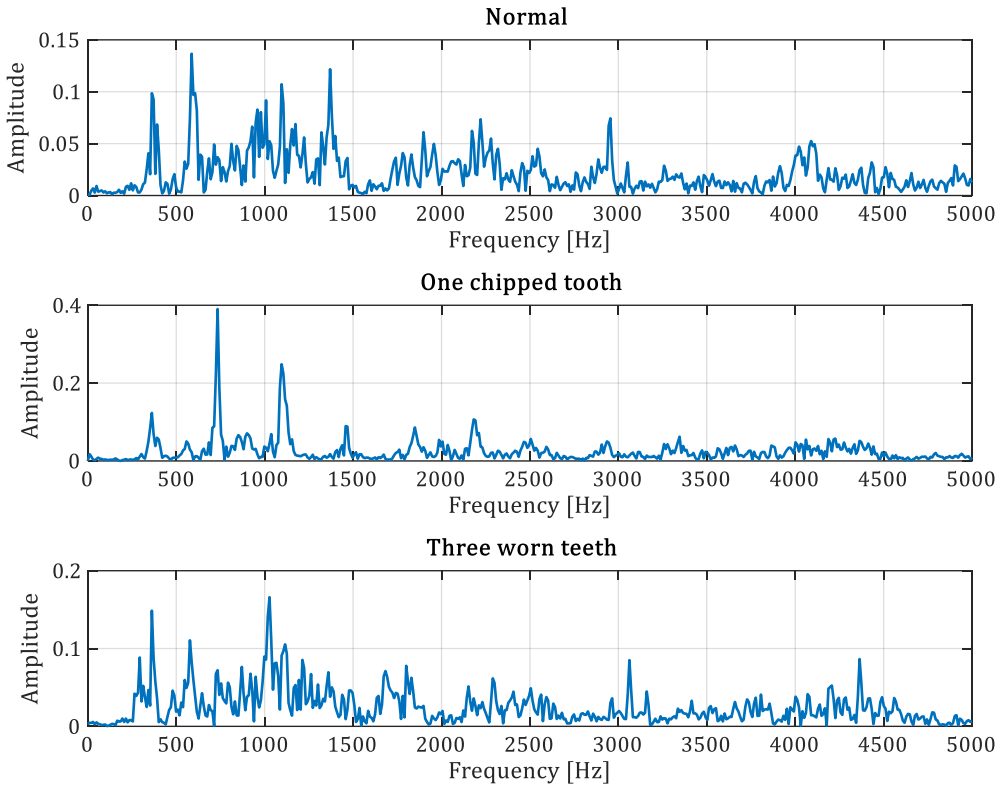

Figure 7: Extracted features from an example in class 1 (top), 2 (middle) and 3 (bottom).

6.3.2 Results

To the best of our knowledge, this case study has been used to validate the proposed diagnostic system only in [41]. The results therein are compared with those of the proposed method in Table 6.

Work	Method	Training examples (%)	Test accuracy (%)
[41]	Local Gaussian correlation of wavelet coefficients and linear Support Vector Machines (SVM)	37.50	99.78
Proposed Method	ELNET-MLR fed by DFT amplitude features	80	**Mean:100.00 $\pm$ 0.00 Minimum: 100.00**
Proposed Method	ELNET-MLR fed by DFT amplitude features	20	**Mean: 100.00 $\pm$ 0.00 Minimum: 100.00**
Proposed Method	ELNET-MLR fed by DFT amplitude features	5	**Mean: 99.91 $\pm$ 0.32 Minimum: 98.85**

Table 6. Overall classification accuracy of different methods which use a supervised approach.

From Table 6, we can conclude that the ELNET-MLR provides slightly better performance than [41] also in setting S3. One additional main advantage of the proposed method with respect to [41] is that after turning to the frequency domain, no further hand-crafted feature extraction step is required. On the contrary, in [41] an ad-hoc method based on local Gaussian correlation is developed to extract relevant features from the CWT scalograms.

6.4. Discussion about the results of the ECS based on ELNET-MLR

Results in Sections 6.1-6.3 show that the proposed ELNET-MLR based ECS fed by DFT amplitude features provides very accurate classification predictions. Further, the method has proven to be capable of drastically reducing the number of features: Tables 7, 8 and 9 report the average numbers of features selected for setting S1 in the three case studies. These are much smaller than the corresponding numbers of parameters to be estimated (Table 10). Finally, computational times are very encouraging (Table 10). These results motivate the use of the proposed ECS for the problem of diagnosing on-board aeronautical systems faults.

Class c	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Number of nonzero entries of $\{\beta_{0,c}, \beta_c\}_{c=1}^{C}$ (maximum = 1025)	4.60 (0.90)	13.80 (2.70)	13.00 (1.22)	15.20 (2.16)	22.00 (3.16)	15.60 (3.43)	25.40 (3.28)	26.60 (3.64)	22.00 (1.41)	24.40 (2.08)	12.80 (0.83)	14.80 (1.92)	18.80 (1.92)	17.60 (2.5)

Table 7: Mean number of nonzero entries of parameter vector $\{\beta_{0,c}, \boldsymbol{\beta}_c\}, c = 1, \dots, C$, and its standard deviation over the 5 CV folds in setting S1 for the case study discussed in Subsection 6.1 (CWRU bearing dataset [26]).

Class c	**1**	**2**	**3**
Number of nonzero entries of $\{\beta_{0,c}, \beta_c\}_{c=1}^{C}$ (maximum = 2049)	41.00 (4.35)	31.00 (4.31)	37.80 (3.76)

Table 8: Mean number of nonzero entries of parameter vector $\{\beta_{0,c}, \boldsymbol{\beta}_c\}, c = 1, \dots, 3$, and its standard deviation over the 5 CV folds in setting S1 for the case study discussed in Subsection 6.2 (MFPT bearing dataset [34]).

Class c	**1**	**2**	**3**
Number of nonzero entries of $\{\beta_{0,c}, \beta_c\}_{c=1}^{C}$ (maximum = 513)	7.80 (1.64)	8.20 (1.10)	5.60 (0.54)

Table 9: Mean number of nonzero entries of parameter vector $\{\beta_{0,c}, \boldsymbol{\beta}_c\}, c = 1, \dots, 3$, and its standard deviation over the 5 CV folds in setting S1 for case study discussed in Subsection 6.3 (gear-box dataset [36]).

Case Study	**Number of parameters to be estimated**	**Mean computational time (seconds)**
CWRU bearing dataset [29] (Subsection 6.1)	14350	847.9 s
MFPT bearing dataset [38] (Subsection 6.2)	6147	28.3 s
Gear-box dataset [41] (Subsection 6.3)	1539	2.44 s

Table 10: Mean computational time for model training over the 5 CV folds in setting S1.

7. SHAKER-BLOWER FAULT DIAGNOSTICS

In this Section, ELNET-MLR is applied to the signal dataset gathered through a test campaign performed at CIRA, where vibrational signals were acquired from a TIRA shaker blower. In this Section, ELNET-MLR is applied to the signal dataset gathered through a test campaign performed at CIRA, where vibrational signals were acquired from a TIRA shaker blower. The main performance characteristics of the radial high-pressure blower used for performing test experiment are reported in Table 11.

Volumetric flow rate [m^3/min]	32
Total pressure difference [Pa]	16000
Voltage [V]	400Δ
Frequency [Hz]	105
Current consumption [A]	25
Motor rating [kW]	11
Number of revolutions [min^{-1}]	6200
Blower speed [min^{-1}]	6200
Weight [kg]	92

Table 11. Main characteristics of the radial high-pressure blower.

More details on the blower geometry and dimensions can be found in [42].

In the experimental setting, the shaker was suspended by a flexible link to a crane, to avoid the influence of the dynamic response of the supporting structure. The shaker was connected to the engine by a flexible stinger attached to the case with a structural glue (Figure 8).

Figure 8: Experimental test setup.

Vibration load due to damages was injected by means of an electro-mechanical shaker applied to the case engine. We have experimentally reproduced $C = 3$ different types of defects (varying the defect type magnitudes):

- *Stator field asymmetry*: this type of defect is due to defects in both motor stator and rotor. They are mainly associated with eccentric rotor positioning, winding asymmetry, short circuits in the winding. They give rise to vibrations at a frequency of twice the electrical network frequency $f_e = 50$ Hz [43].

- *Defect of the bearing outer ring*: this type of defect leads to vibration features at specific frequencies, which depend on the bearing geometry and rotation speed. If we indicate by f_r the shaft rotation frequency, R_l the number of rolling bodies, v the diameter of the rolling elements, V the diameter of the average circumference of the rolling bearing and γ the contact angle of the rolling element with the guides, the main frequency generated by the rolling bearing defect reads:

$$f_{vo} = \frac{R_l}{2} f_r \left(1 - \frac{v}{V}\cos(\gamma)\right) \tag{14}$$

  An approximate formula is typically preferred to Eq. (14), which takes into account the effect of ball crawling [43]:

$$f_{vo} = 0.4 \cdot R_l \cdot f_r = \frac{0.4 \cdot R_l}{60}(\text{RPM}) \tag{15}$$

- *Defect of the bearing inner ring*: in this case, the characteristic frequency is [43]

$$f_{vi} = \frac{R_l}{2} f_r \left(1 + \frac{v}{V} \cos(\gamma)\right). \tag{16}$$

Also in this case, an approximated formula can be used [43]

$$f_{vi} = 0.6 \cdot R_l \cdot f_r = \frac{0.6 \cdot R_l}{60} (\text{RPM}) \tag{17}$$

Since the ball bearings support the rotor, any defect in the bearings will produce a radial displacement between rotor and stator. This movement generates stator currents at predictable frequencies [43]

$$f_p = |f_e \pm q f_v| \tag{18}$$

where $f_v = f_{vo}$ for defects on the outer ring and $f_v = f_{vi}$ for defects on the inner ring, whereas f_e is the electrical network frequency and q is an integer number.

Signals used to feed the electrodynamic shaker are characterized by the spectral content described above. Further, varying magnitudes of defect types have been experimentally simulated. Figures 9, 10 and 11 show examples of the excitation signals used for injecting the three types of defects used to develop the experimental dataset.

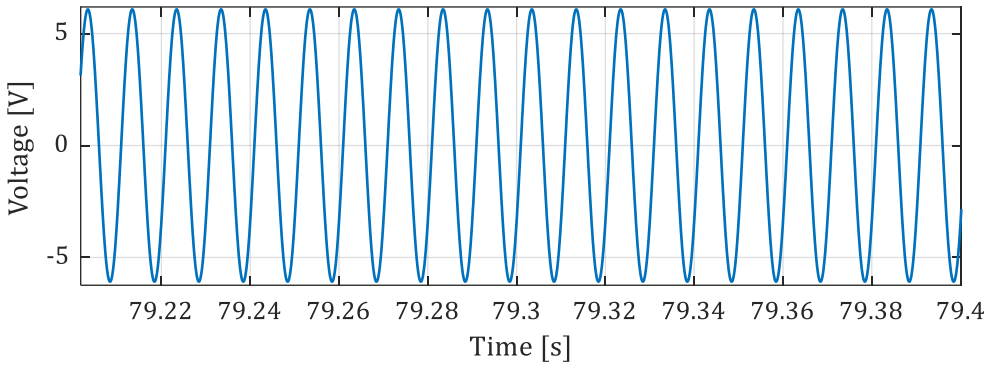

Figure 9: Shaker blower input signal: stator field asymmetry.

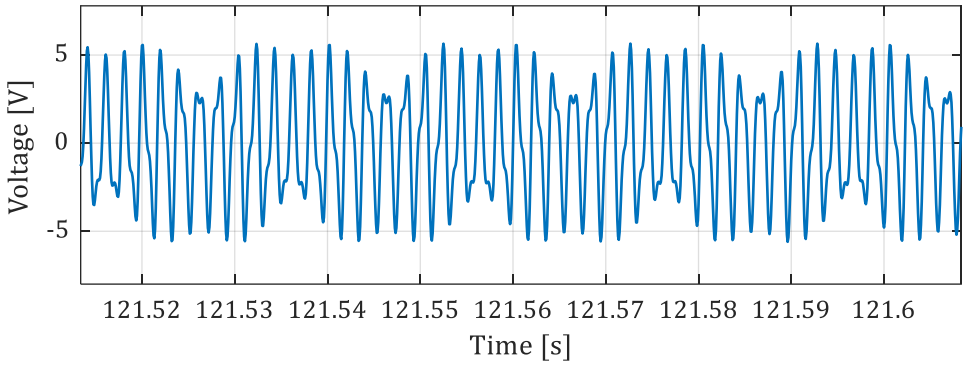

Figure 10: Shaker blower input signal: bearing outer ring defect.

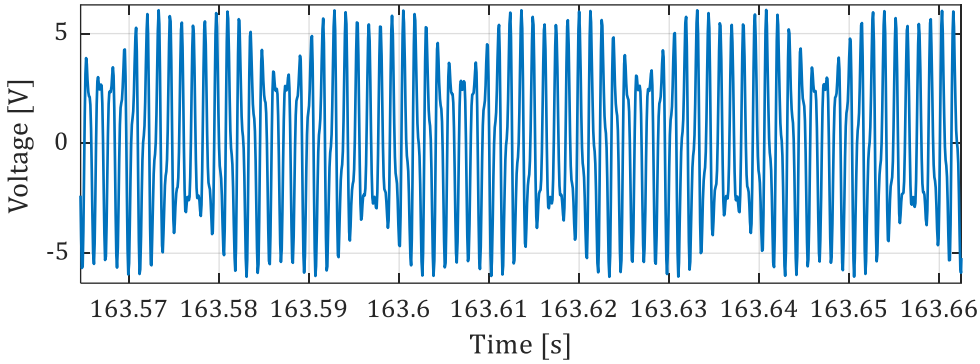

Figure 11: Shaker blower input signal: bearing inner ring defect.

Stator field asymmetry has been simulated through a pure tone at 100 Hz. Bearing outer and inner ring defects have been simulated each one as a superposition of four harmonics. The harmonic content of the bearing outer ring defect signal (f_{vo}= 546Hz) is 496 Hz, 596Hz, 1042Hz and 1142Hz, with relative amplitudes 1, 0.1, 0.4 and 0.06. The harmonic content of the bearing inner ring defect signal (f_{vi} = 819Hz) is 769Hz, 869Hz, 1588Hz and 1688Hz, with relative amplitudes 1, 0.1, 0.4 and 0.06.

To simulate the damage growth over time, the amplitudes of the three signals have been varied in seven steps between about 0.05 V and 0.7 V

The test instrumentation used to drive the shaker and acquire the accelerometers signals was Siemens/LMS Scadas III front-end, equipped with a generation module and multiple acquisition modules.

The generated and acquired signals were sampled at frequency 12.8 kHz and measured during the normal running condition of the blower. Eleven tests were performed. In every test, a signal of about 140s fed the shaker, which contains seven replicas of 5s window for each of the three damages. Between the replicas, there was a 1s of zero signal to switch between faults and fault levels.

Figure 12 shows the input signal for the shaker, whereas Figure 13 shows a typical signal acquired by the accelerometers.

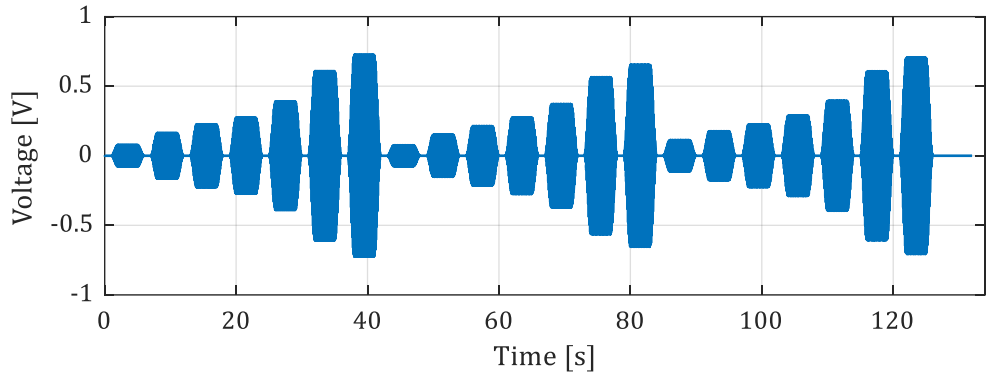

Figure 12: Time history of the driven signal

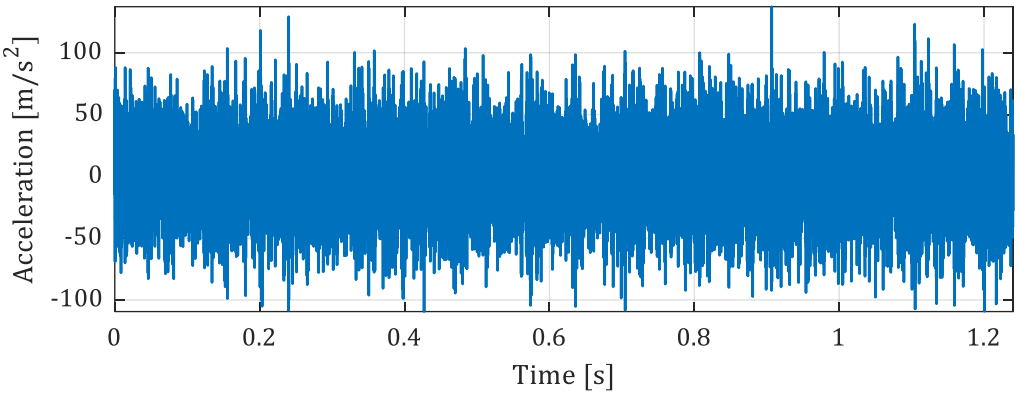

Figure 13: Typical acceleration signal.

Table 12 reports the labels assigned to the blower conditions and the number of experimentally generated examples.

Shaker-blower condition	Label	Number of examples
Asymmetry of the stator magnetic field	1	77
Defect of the bearing outer ring	2	77
Defect of the bearing inner ring	3	77

Table 12: Label assigned to each shaker blower condition.

The vibration measurements were performed by three tri-axial accelerometers PCB type 356A16, with a sensitivity of about 100 mV/g (Figure 14). Accelerometers $E1, E2$ and $E3$ were positioned at the connection between the electrical motor and the ground, on the input power case (the same place where the electrodynamical shaker applies its force), and on the pump cover, respectively. In this

setting, $S = 9$ acceleration signals have been acquired, i.e., 3 acceleration signals on the x, y and z directions from each of the three sensors.

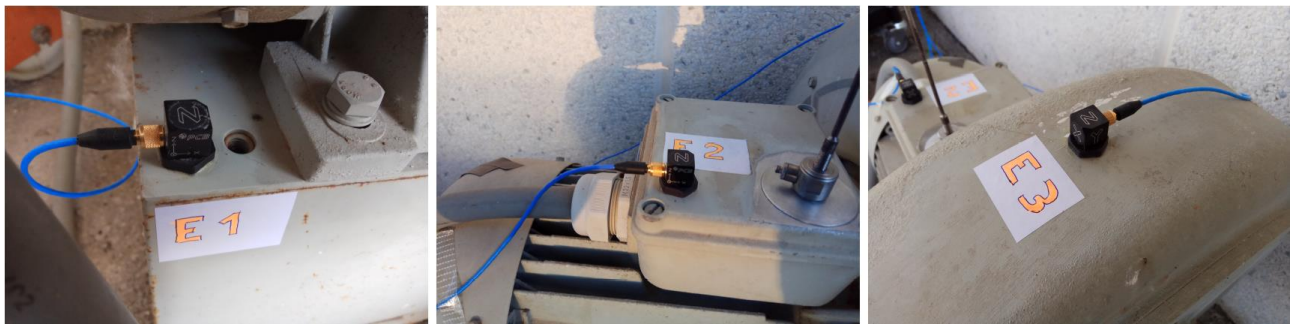

Figure 14: Acceleration transducers location and excitation device (close to E2).

7.1 Feature Extraction

Each vibration signal is partitioned using a fixed time window of approximately 0.64 seconds, which corresponds to time windows of 2^{14} samples. Table 13 reports the number of examples obtained for each class.

From each time window, $\widetilde{L_s} = 8192, s = 1, \dots, 9$, DFT amplitude features are extracted from $S = 9$ vibrational signals. Then, the total number of features extracted is $F = 73728$. Figure 15 shows the extracted features (amplitude spectrum) from sensor $s = 1$ on an example belonging to classes 1, 2 and 3.

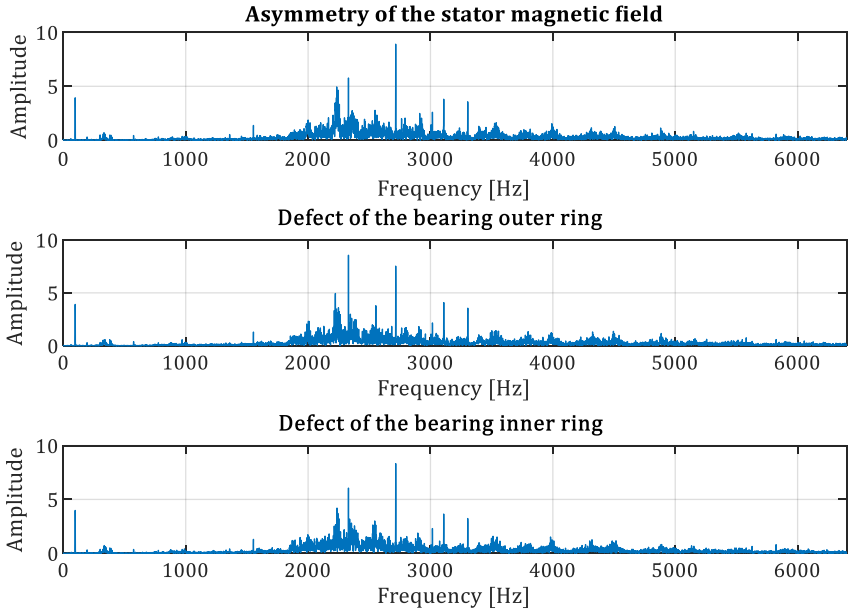

Figure 15: Extracted features from sensor 1 from an example in class 1 (top), 2 (middle) and 3 (bottom).

Class	1	2	3
Number of Examples	77	77	77

Table 13: Number of available examples for each class

7.2 Results

The performance of the proposed method is evaluated through a twice-nested 5- FCV (i.e, setting S1). Table 14 reports the mean percentage classification accuracy and its standard deviation, whereas Table 15 reports the mean confusion matrix over the 5 cross-fold-validation, which is a $3x3$ matrix whose entry $(c, \tilde{c}), c, \tilde{c} = 1, \dots, 3$, represents the percentage of times that an example being in class c is labelled as $\tilde{c}$.

Classification accuracy (%)
95.25 (0.18)

Table 14: Classification accuracy.

		Predicted class		
Ground truth class		**1**	**2**	**3**
	1	94.75	2.58	2.67
	2	2.58	97.42	0.00
	3	6.42	0.00	93.58

Table 15: Mean confusion matrix over 5 cross-fold validation.

From Table 15, we can conclude that:

- examples whose ground truth class is 1 are correctly classified 94.75% of times. When the ECS misclassifies an example of class 1, this is assigned to class 2 in 2.58% of times or 3 in 2.67% of times;
- examples whose ground truth class is 2 are correctly classified 97.42% of times. When the ECS misclassifies an example of class 2, this is assigned to class 1 in 2.58% of the cases;
- examples whose ground truth class is 3 are correctly classified 93.58% of times. When the ECS misclassifies an example of class 3, this is assigned to class 1 in 6.42% of the cases.

Table 16 reports the mean number of selected features and corresponding standard deviations for each class: the proposed approach allows drastically reducing the feature set.

The mean computational time needed to train the ELNET-MLR is 305 seconds, which is a reasonable computational time if we consider that $(73728 + 1) \cdot 3 = 221187$ parameters needed to be estimated.

Class c	**1**	**2**	**3**
Number of nonzero entries of $\{\beta_{0,c}, \boldsymbol{\beta}_c\}_{c=1}^{C}$ (maximum = 73729)	67.00 (5.95)	30.60 (3.36)	28.80 (4.08)

Table 16: Mean number and its standard deviation of nonzero entries of $\{\beta_{0,c}, \boldsymbol{\beta}_c\}_{c=1}^{C}$.

8. DISCUSSION AND OUTLOOKS

In this work, we developed a diagnostic classification system, which benefits from the computationally lean and effective feature selection provided by ELNET. This is a very relevant result, since no cumbersome direct development and identification of hand-crafted features from frequency domain signal representation is required after moving to the frequency domain. Moreover, the proposed classification scheme is based on a linear combination of the extracted DFT amplitude features (see Eq. (4)); although this allows easily interpreting the selected features, nonetheless it can be a limitation in terms of classification accuracy in some applications. One possible solution is to resort to nonlinear models such as deep learning models which allow learning relevant data representation for the classification task directly from the data, without requiring any prior knowledge.

To conclude, it seems fair to say that when the direct use of DFT amplitude features do not provide good results, we can use deep architectures to try to achieve better performance. From this, it emerges that although the ELNET-MLR has very good performance, we cannot conclude that it should be always preferred to other methods.

Finally, we have not considered that aerospace equipment typically works in changing operational conditions. This challenges the feature extraction and selection steps. This topic will be investigated in future research work.

9. CONCLUSIONS

In this work, we have developed a novel diagnostic ECS based ELNET-MLR method for identifying the type of defect of shaker blower used in the aerospace industry, which is based on the extraction of DFT amplitude features. The proposed solution allows selecting in reasonable computational time a limited number of features from a very large set of possible alternatives. The proposed approach has shown to provide state-of-the-art classification performance when applied to three case studies of

literature. In particular, the proposed approach has proven to be also useful when few training data are available. This situation is typical of high reliability and unique or new designed industrial equipment, for which condition monitoring data up to the occurrence of a failure may be not available.

REFERENCES

[1] https://www.aviasg.com/en/press-release/can-predictive-maintenance-become-the-future-of-aircraft-mro.html (accessed on 25 January 2019).

[2] https://h2020-remap.eu/ (accessed on 25 January 2019).

[3] Suo, M., Zhu, B., An, R., Sun, H., Xu, S., Yu, Z., "Data-driven fault diagnosis of satellite power system using fuzzy Bayes risk and SVM", Aerospace Science and Technology, 84, pp. 1092-1105, 2019.

[4] Xi, P.-P., Zhao, Y.-P., Wang, P.-X., Li, Z.-Q., Pan, Y.-T., Song, F.-Q., "Least squares support vector machine for class imbalance learning and their applications to fault detection of aircraft engine", Aerospace Science and Technology, 84, pp. 56-74, 2019.

[5] Lu, F., Jiang, J., Huang, J., Qiu, X., "Dual reduced kernel extreme learning machine for aero-engine fault diagnosis", Aerospace Science and Technology, 71, pp. 742-750, 2017.

[6] Cannarile, F., Compare, M., Baraldi, P., Di Maio, F., Zio, E., "Homogeneous continuous-time, finite-state hidden semi-markov modeling for enhancing empirical classification system diagnostics of industrial components", Machines, 6(3), art. no. 34, 2018.

[7] Hu, Y., Baraldi, P., Di Maio, F., Zio, E., "A Systematic Semi-Supervised Self-adaptable Fault Diagnostics approach in an evolving environment", Mechanical Systems and Signal Processing, 88, pp. 413-427, 2017.

[8] Ruiz, M., Mujica, L.E., Alférez, S., Acho, L., Tutivén, C., Vidal, Y., Rodellar, J., Pozo, F., "Wind turbine fault detection and classification by means of image texture analysis", Mechanical Systems and Signal Processing, 107, pp. 149-167, 2018.

[9] Baraldi, P., Cannarile, F., Di Maio, F., Zio, E., "Hierarchical k-nearest neighbors classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions", Engineering Applications of Artificial Intelligence, 56, pp. 1-13, 2016.

[10] Cannarile, F., Compare, M., Zio, E., "A fault diagnostic tool based on a first principle model simulator", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 10437 LNCS, pp. 179-193, 2017.

[11] Shen, C., Qi, Y., Wang, J., Cai, G., Zhu, Z., "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder", Engineering Applications of Artificial Intelligence, 76, pp. 170-184, 2018.

[12] Yang, B.S.; Widodo, A. "Introduction of Intelligent Machine Fault Diagnosis and Prognosis", Nova Science Publishers: New York, NY, USA, 2009.

[13] Caesarendra, W., Tjahjowidodo, T., "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing", Machines, 5(4), art. no. 21, 2017.

[14] Mallat, S., "A Wavelet Tour of Signal Processing", Academic Press, 1999.

[15] Huang, N.E., "Introduction to the Hilbert-Huang Transform and its related mathematical problems", Interdisciplinary Mathematical Sciences Hilbert-Huang Transform and Its Applications: 1–26, 2014.

[16] Iguyon, I., Elisseeff, A., "An introduction to variable and feature selection, Journal of Machine Learning Research", 3, pp. 1157-1182, 2003.

[17] Kohavi, R., John, G.H., "Wrappers for feature subset selection", Artificial Intelligence, 97(1-2), pp. 273-324, 1997.

[18] Salem Alelyani, Jiliang Tang, and Huan Liu, “Feature selection for clustering: A review”. In Data Clustering: Algorithms and Applications, pages 29–60, CRC Press, 2013.
[19] Lu, M., “Embedded feature selection accounting for unknown data heterogeneity”, Expert Systems with Applications, 119, pp. 350-361, 2019.
[20] Hastie, T., Tibshirani, R., Wainwright, M., “Statistical learning with sparsity: the LASSO and generalizations”, Chapman and Hall/CRC, 2015.
[21] Tibshirani, R., “Regression Shrinkage and Selection via the LASSO”, Journal of the Royal Statistical Society. Series B (Methodological), 58(1), pp. 267-288, 1996.
[22] Friedman, J., Hastie, T., Tibshirani, R., “Regularization paths for generalized linear models via coordinate descent”, Journal of Statistical Software, 33(1), pp. 1-22, 2010.
[23] Cai, Y., Chow, M.-Y., Lu, W., Li, L., “Statistical feature selection from massive data in distribution fault diagnosis”, IEEE Transactions on Power Systems, 25 (2), art. no. 5405081, pp. 642-648, 2010.
[24] Duque-Perez, O., Del Pozo-Gallego, C., Morinigo-Sotelo, D., Fontes Godoy, W., “Bearing fault diagnosis based on LASSO regularization method”, Proceedings of the 2017 IEEE 11th International Symposium on Diagnostics for Electrical Machines, Power Electronics and Drives, SDEMPED 2017, 2017-January, pp. 331-337, 2017.
[25] Cannarile, F., Baraldi, P., Zio, E., “An evidential similarity-based regression method for the prediction of equipment remaining useful life in presence of incomplete degradation trajectories”, Fuzzy Sets and Systems, https://doi.org/10.1016/j.fss.2018.10.008, 2018.
[26] Zou, H., Hastie, T., “Regularization and variable selection via the elastic net”, Journal of the Royal Statistical Society. Series B: Statistical Methodology, 67 (2), pp. 301-320, 2005.
[27] He, H., Garcia, E.A., “Learning from imbalanced data”, IEEE Transactions on Knowledge and Data Engineering, 21 (9), pp. 1263-1284, 2009.
[28] Picek, S., Heuser, A., Jovic, A., Bhasin S., Regazzoni, F., “The curse of class imbalance and conflicting metrics with machine learning for side-channel evaluations”, IACR Transactions on Cryptographic Hardware and Embedded Systems, 1, pp. 209–237, 2019.
[29] Case Western Reserve University Bearing Data Center Website. Available online: http://csegroups.case.edu/ bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website (accessed on 25 June 2018).
[30] Van, M., Kang, H.-J.,“Bearing Defect Classification Based on Individual Wavelet Local Fisher Discriminant Analysis with Particle Swarm Optimization”, IEEE Transactions on Industrial Informatics, 12(1), art. no. 7327187, pp. 124-135, 2016.
[31] Saidi, L., Ben Ali, J., Fnaiech, F., “Application of higher order spectral features and support vector machines for bearing faults classification”, ISA Transactions, 54, pp. 193-206, 2016.
[32] Xu, Z., Li, Y., Wang, Z., Xuan, J., “A selective fuzzy ARTMAP ensemble and its application to the fault diagnosis of rolling element bearing”, Neurocomputing, 182, pp. 25-35, 2016.
[33] Jia, F., Lei, Y., Lin, J., Zhou, X., Lu, N., “Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data”, Mechanical Systems and Signal Processing, 72-73, pp. 303-315, 2016.
[34] Mao, W., Feng, W., Liang, X., “A novel deep output kernel learning method for bearing fault structural diagnosis”, Mechanical Systems and Signal Processing, 117, pp. 293-318, 2019.
[35] Gan, M., Wang, C., Zhu, C., “Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings”, Mechanical Systems and Signal Processing, 72-73, pp. 92-104, 2016.
[36] Shao, H., Jiang, H., Lin, Y., Li, X., “A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders, Mechanical Systems and Signal Processing, 102, pp. 278-297, 2018.
[37] Lin, Y., Li, X., Hu, Y., “Deep diagnostics and prognostics: An integrated hierarchical learning framework in PHM applications”, Applied Soft Computing Journal, 72, pp. 555-564, 2018.

[38] MFPT bearing data set. Available online: http://www.mfpt.org/FaultData/FaultData.htm. (accessed on 1 July 2018).
[39] Verstraete, D., Ferrada, A., Droguett, E.L., Meruane, V., Modarres, M., "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings", Shock and Vibration, 2017.
[40] D. Lee, V. Siu, R. Cruz, and C. Yetman, "Convolutional neural net and bearing fault analysis," in Proceedings of the International Conference on Data Mining series (ICDM) Barcelona, pp. 194–200, San Diego, CA, USA, 2016.
[41] Zamanian, A.H., Ohadi, A., "Gear fault diagnosis based on Gaussian correlation of vibrations signals and wavelet coefficients", Applied Soft Computing Journal, 11(8), pp. 4807-4819, 2011.
[42] https://www.elektror.de/en/product/HRD7FU (Accessed on 1 July, 2019)
[43] https://moodle2.units.it/pluginfile.php/75130/mod_resource/content/2/diagnosi%20macchinario.pdf (Accessed on 1 July, 2018).