TecnoLógicas
ISSN 0123-7799
ISSN-e 2256-5337
Vol. 20, No. 39
Mayo- agosto de 2017

Feature relevance estimation for vibration-based condition monitoring of an internal combustion engine

Estimación de características relevantes para el monitoreo de condición de motores de combustión interna a partir de señales de vibración

José Alberto Hernández-Muriel[1],
Andrés Marino Álvarez-Meza[2],
Julián David Echeverry-Correa[3],
Álvaro Ángel Orozco-Gutierrez[4]
and Mauricio Alexánder Álvarez-López[5]

Recibido: 03 de marzo de 2017,
Aceptado: 12 de mayo de 2017

Cómo citar / How to cite
J.A. Hernández-Muriel, A.M. Álvarez-Meza, J.D. Echeverry-Correa, A.A. Orozco-Gutiérrez and M.A. Álvarez-López, "A feature relevance estimation approach for vibration-based condition monitoring of internal combustion engine", *TecnoLógicas*, vol. 39, no. 20, 2017.

[1] Electronic Engineer, Electrical Engineering Department, Universidad Tecnológica de Pereira, Pereira-Colombia, j.hernandez12@utp.edu.co
[2] PhD in Engineering - Automatics, Electrical Engineering Department, Universidad Tecnológica de Pereira, Pereira-Colombia, andres.alvarez1@utp.edu.co
[3] PhD in Engineering - Electronic Systems, Electrical Engineering Department, Universidad Tecnológica de Pereira, Pereira-Colombia, jde@utp.edu.co
[4] PhD in Bioengineering, Electrical Engineering Department, Universidad Tecnológica de Pereira, Pereira-Colombia, aaog@utp.edu.co
[5] PhD in Computer Science, Department of Computer Science, University of Sheffield, Sheffield-United Kingdom, mauricio.alvarez@sheffield.ac.uk

Abstract

Condition monitoring of Internal Combustion Engines (ICE) benefits cost-effective operations in the modern industrial sector. Because of this, vibration signals are commonly monitored as part of a non-invasive approach to ICE analysis. However, vibration-based ICE monitoring poses a challenge due to the properties of this kind of signals. They are highly dynamic and non-stationary, let alone the diverse sources involved in the combustion process. In this paper, we propose a feature relevance estimation strategy for vibration-based ICE analysis. Our approach is divided into three main stages: signal decomposition using an Ensemble Empirical Mode Decomposition algorithm, multi-domain parameter estimation from time and frequency representations, and a supervised feature selection based on the *Relief-F* technique. Accordingly, we decomposed the vibration signals by using self-adaptive analysis to represent nonlinear and non-stationary time series. Afterwards, time and frequency-based parameters were calculated to code complex and/or non-stationary dynamics. Subsequently, we computed a relevance vector index to measure the contribution of each multi-domain feature to the discrimination of different fuel blend estimation/diagnosis categories for ICE. In particular, we worked with an ICE dataset collected from fuel blends under normal and fault scenarios at different engine speeds to test our approach. Our classification results presented nearly 98% of accuracy after using a k-Nearest Neighbors machine. They reveal the way our approach identifies a relevant subset of features for ICE condition monitoring. One of the benefits is the reduced number of parameters.

Keywords

Internal combustion engines, vibration signal, multi-domain features, relevance analysis, feature selection.

Resumen

El monitoreo de condición de motores de combustión interna (MCI) facilita que las operaciones del sector industrial moderno sean más rentables. En este sentido, las señales de vibración comúnmente son empleadas como un enfoque no invasivo para el análisis de MCI. Sin embargo, el monitoreo de MCI basado en vibraciones presenta un desafío relacionado con las propiedades de la señal, la cual es altamente dinámica y no-estacionaria, sin mencionar las diversas fuentes presentes durante el proceso de combustión. En este artículo, se propone una estrategia de análisis de relevancia orientada al monitoreo de MCI basado en vibraciones. Este enfoque incorpora tres etapas principales: descomposición de la señal utilizando un algoritmo de Ensemble Empirical Mode Decomposition, estimación de parámetros multi-dominio desde representaciones en tiempo y frecuencia, y una selección supervisada de características basada en Relief-F. Así, las señales de vibración se descomponen utilizando un análisis auto-adaptativo para representar la no-linealidad y no-estacionariedad de las series de tiempo. Luego, para codificar dinámicas complejas y/o no estacionarias, se calculan algunos parámetros en el dominio del tiempo y de la frecuencia. Posteriormente, se calcula un vector de índice de relevancia para cuantificar la contribución de cada una de las características multi-dominio para discriminar diferentes categorías de estimación de mezcla de combustible y diagnóstico de MCI. Los resultados de clasificación obtenidos (cercanos al 98% de acierto) en una base de datos de MCI, revelan como la propuesta planteada identifica un subconjunto de características relevantes en el monitorio de condición de MCI.

Palabras clave

Motores de combustión interna, señales de vibración, características multi-dominio, análisis de relevancia, selección de características.

1. INTRODUCTION

Nowadays, cost-effective operations are required in the modern industry because of the competitiveness in the sector, a higher demand for fuel, and the increase in power consumption. Therefore, machinery fault diagnosis represents an alternative to mitigate current problems in the industry [1], [2]. Certainly, the detection of abnormal states at their early stages helps to avoid greater or even severe faults. In practice, one of the most frequently used pieces of machinery is an Internal Combustion Engine (ICE). Indeed, motor vehicles and industrial engines have a significant effect on the whole world because they use gasoline as their energy source. Besides, the greater the engine fault the lower the efficiency. This is due to an increase in fuel consumption [3]. Hence, condition monitoring of ICE has been the focus of many research approaches [4]–[8].

In the literature, machinery condition monitoring has been addressed from three different perspectives: model-based, signal-based, and knowledge-based approaches. Model-based schemes aim to detect changes in machine behavior by using a mathematical model of the system [9]–[12]. This method was created to replace hardware redundancy by analytical redundancy [13], [14]. Such changes depend on deterministic models of industrial processes constructed by using either physical principles or system identification techniques. Nevertheless, in most cases, the models hold assumptions that do not correspond to the real performance of the machine. In turn, signal-based algorithms use measured signals rather than specific input-output models for fault diagnosis. Therefore, the device states are inferred from the measured data [15]. Generally, the pressure curve of the engine cylinder head is obtained for diagnosing the state of ICE [16], [17]. However, such signal involves intrusive measurement under controlled environments [18], [19]. Other types of variables have been explored to achieve signal-based condition monitoring, e.g. oil analysis, gas exhaust tests, fuel-air ratio, acoustic emission, and ignition time [20], [21]. Nonetheless, a special testbed is necessary to collect useful information. In contrast, non-invasive measurements comprise vibration transducers such as accelerometers, velocity pickups, and displacement probes. Remarkably, vibration signals provide an efficient way for monitoring the dynamic condition of a machine [22]. Although signal-based diagnosis provides a more realistic method to assess the monitoring than model-based ones, it depends on the analysis of a symptom compared to normal states. This process is mainly carried out by an expert in the field.

More elaborate condition monitoring approaches comprise knowledge-based methods that address the problem of machine diagnosis from a pattern recognition perspective. Thus, a knowledge-based technique can be divided into three general stages: i) signal acquisition, ii) feature calculation from provided signals, and iii) condition assessment by unsupervised/supervised learning algorithms [22], [23]. Overall, the vibration signal is often selected as input data for knowledge-based approaches that aim to estimate relevant parameters for further learning stages. In this regard, vibration-based feature estimation comprises time, frequency, and time-frequency-based parameters [24]–[26]. Typical approaches to the time domain-based features have focused on statistical measures intended for stationary processes [27]. In general, they overlook early fault symptoms, which is more suitable to time-invariant processes. Frequency representations that employ Fast Fourier Transform (FFT) are the most common method for analyzing vibration signals. However, this technique is inadequate for identifying non-stationary events [28]. In recent years, time-frequency features based on Wavelet Transform (WT) have shown promising results in machine fault

diagnosis [25], [29], [30]. Still, energy leakage should occur in WT due to the limited length of the mother wavelet. Also, only signal features that match the input shape are likely to be detected. Conversely, all the other parameters will be masked or even completely ignored [31].

Different unsupervised/supervised clustering strategies conduct condition assessment in knowledge-based frameworks after the computation of the feature representation space from the acquired signals. They include Hidden Markov Models [32], Neural Networks [23], [33], [34], Genetic Algorithms [35], and Support Vector Machines [31], [36], [37]. However, the original dataset used for fault diagnosis often consists of a vast number of features, and the number of samples is limited because of the data acquisition workload. It is widely known that the features used to describe the patterns determine the search space to be explored during the learning phase. Thus, irrelevant and noisy features make the search space larger, increasing both time and complexity of the learning process while reducing the condition monitoring accuracy [38]. Feature extraction and selection techniques are often employed for dimensionality reduction. The aim of this type of reduction is to find a set of relevant features based on the input [39]. Other feature extraction methods have been explored as well, such as Principal Component Analysis (PCA) [40] and Kernel Principal Component Analysis (KPCA) [41]. Nevertheless, the new features (transformed by feature extraction methods) usually lose the original engineering meaning. Therefore, other dimensionality reduction methods known as feature selection are required. A self-weight algorithm and distance evaluation (as comparison method) are studied in [26] to select relevant parameters for bearing fault diagnosis. Yet, during ICE condition monitoring tasks, the selection of relevant parameters to support further learning stages remains an open issue.

In this work, we explore the application of a supervised feature selection strategy for time and frequency-based parameters from vibration signals as an alternative that benefits condition monitoring of ICE. Specifically, our feature relevance analysis approach includes a signal decomposition stage based in the Ensemble Empirical Mode Decomposition (EEMD) algorithm to extract nonlinear and nonstationary patterns. Then, statistical measures from time and frequency domains are estimated, including intra-band variations within the energy spectrum. Lastly, the relevance of each feature is computed by the supervised Relief-F algorithm. We tested our feature relevance approach on an ICE vibration dataset of combustion analysis. Our results reveal that our alternative achieves suitable classification accuracy with a low number of relevant features. The rest of the paper is organized as follows: In the Methodology section, we describe the theoretical background of the feature relevance approach we are introducing for vibration-based condition monitoring of ICE. Then, we describe the Experimental setup and discuss the Results we obtained. Finally, we outline the main ideas in the Conclusions.

2. METHODOLOGY

Let $\mathbf{X} \in \mathbb{R}^{N \times T}$ be a matrix holding N vibration segments $\boldsymbol{x} \in \mathbb{R}^{T}$ at T time intervals from a combustion engine. Our approach to feature relevance analysis reveals discriminative multi-domain parameters in fuel blend estimation/diagnosis of ICE. This method is divided into three stages: *i)* signal decomposition, *ii)* multi-domain feature computation, and *iii)* supervised feature selection with fuel blend estimation/diagnosis for ICE condition monitoring. Each step is described below.

2.1 Signal decomposition

The Ensemble Empirical Mode Decomposition (EEMD) technique is applied to each provided segment. This step works as a self-adaptive analysis to decompose nonlinear and non-stationary signals. In particular, we aim to overcome the mode mixing drawback of the traditional EMD algorithm by using EEMD noise-assisted data analysis [42]. Therefore, the EEMD algorithm decomposes a raw signal into a collection of J *True Intrinsic Mode Functions* (TIMFs), as follows:

$$\boldsymbol{x} = \sum_{j=1}^{J} \boldsymbol{z}_j + \boldsymbol{r}, \tag{1}$$

where $\boldsymbol{z}_j \in \mathbb{R}^{\boldsymbol{T}}$ is the j-th TIMF and $\boldsymbol{r} \in \mathbb{R}^{\boldsymbol{T}}$ is a residual. Each TIMF is calculated as the sample average over M ensemble trials after noise perturbation in the input signal. The above mentioned variables yield $z_j = \frac{1}{M}\sum_{m=1}^{M} \hat{z}_j^{(m)}$, being $\hat{\boldsymbol{z}}_{\boldsymbol{j}}^{(\boldsymbol{m})} \in \mathbb{R}^{\boldsymbol{T}}$ the j-th EMD-based IMF at the m-th trial $\boldsymbol{x}^{(\boldsymbol{m})} = \boldsymbol{x} + \boldsymbol{\varepsilon}^{(\boldsymbol{m})}$, with $\boldsymbol{\varepsilon}^{(\boldsymbol{m})} \in \mathbb{R}^{\boldsymbol{T}}$ following a white Gaussian distribution including a standard deviation $\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}} \in \mathbb{R}^{+}$. Thereby, the EEMD principle would populate the whole time–frequency space uniformly with the constituting components of different scales. As a result, the TIMFs are useful to extract relevant information regarding the combustion engine properties [24].

2.2 Multi-domain feature estimation

Since engine combustion analysis based on vibration signals requires coding complex and/or non-stationary dynamics both time and frequency-based parameters are calculated from the raw signal $\boldsymbol{x}$ and the obtained TIMFs $\boldsymbol{z}_{\boldsymbol{j}}(\boldsymbol{j} \in [\boldsymbol{1}, \boldsymbol{J}])$ [26], [43]. First, time-based parameters are computed as the statistical descriptors presented in Table 1. Broadly speaking, such parameters enable to code high-order moments for further discrimination stages. Indeed, skewness and kurtosis are widely used as indicators of main peaks in the signal and have been shown to be independent from load and speed variation [43].

Second, frequency-based parameters are estimated from vibration signals. Besides, the decompositions of such parameters are calculated by the well-known Fast Fourier Transform (FFT) [44]. Therefore, to reveal harmonic patterns in vibration signals, the frequency spectrum vector $s \in \mathbb{R}^K$ is computed: $s_k = |\sum_{t=1}^{T} x_t e^{-i2\pi kt/T}|$, in agreement with the frequency vector $\lambda \in \mathbb{R}^K$, where $\lambda_k = k\Lambda_s/2K$ and $\Lambda_s \in \mathbb{R}$ hold the sampling frequency $(k \in \{0,1,\cdots,K\})$. Afterwards, the statistical parameters are computed in accordance with the frequency domain (Table 2) [26]. Additionally, we estimated intra-band variations by dividing the frequency spectrum into B bands [29], [43]. Then, the statistical parameters in Table 1 were estimated for each obtained sub-band. Finally, a feature matrix $F \in \mathbb{R}^{N\times Q}$ holding N vibration segments with Q features was constructed after vector concatenation of the aforementioned time and frequency-based parameters.

2.3 Supervised feature selection and fuel-blend estimation/diagnosis

In practice, the provided feature space matrix $\boldsymbol{F}$ reaches huge dimensions. This is crucial to identify the most discriminating features and to find a tradeoff between system complexity and accuracy [38]. To this end, we measure the contribution of the time and frequency-based parameters in terms of the supervised information. The latter is coded in a label vector $\varsigma \in \{1,2,\cdots,C\}^N$, where C is the number of engine combustion categories. In this work, we computed a relevance vector $\rho \in \mathbb{R}^Q$ based on the Relief-F algorithm, as follows [45]:

$$\rho_q = \frac{1}{N}\sum_{n=1}^{N}\left\{-\frac{1}{\vartheta}\sum_{f_{n'}\in\Omega_n^{\varsigma_n}} d(f_{np}, f_{n'p}) + \sum_{c\neq\varsigma_n}\frac{1}{\vartheta}\frac{p(\varsigma=c)}{1-p(\varsigma=\varsigma_n)}\sum_{f_{n'}\in\Omega_n^{c}} d(f_{np}, f_{n'p})\right\} \quad (2)$$

Table 1. Statistical parameters for a given vector $y \in \mathbb{R}^L$. Source: Authors.

Feature	Expression	Feature	Expression
Mean ($\boldsymbol{\mu_y}$)	$\sum_{l=1}^{L}\frac{y_l}{L}$	Skewness	$\sum_{l=1}^{L}\frac{y_l{}^3}{L\,R^3}$
Median	$\boldsymbol{median}_{y_l}(y_l)$	Max value	$\boldsymbol{max}_{y_l}(y_l)$
Standard deviation	$\left(\sum_{l=1}^{L}\frac{y_l-\mu_y}{L}\right)^{1/2}$	Min value	$\boldsymbol{min}_{y_l}(y_l)$
Root mean square ($\boldsymbol{R}$)	$\left(\sum_{l=1}^{L}\frac{y_l}{L}\right)^{1/2}$	Range	$\lvert\boldsymbol{max}_{y_l}(y_l)-\boldsymbol{min}_{y_l}(y_l)\rvert$
Peak ($\boldsymbol{\beta}$)	$\boldsymbol{max}_{y_l}(\lvert y_l\rvert)$	Interquartile range	$\boldsymbol{iqr}_{y_l}(y_l)$
Shape factor	$\frac{LR}{\sum_{l=1}^{L}\lvert y_l\rvert}$	Kurtosis ($\boldsymbol{\kappa}$)	$\sum_{l=1}^{L}\frac{y_l{}^4}{L\,R^4}$
Crest factor	$\frac{\beta}{R}$	Speed $\boldsymbol{\kappa}$	$\boldsymbol{\kappa}\{y'\}$
Impulse factor	$\frac{L\beta}{\sum_{l=1}^{L}\lvert y_l\rvert}$	Acceleration $\boldsymbol{\kappa}$	$\boldsymbol{\kappa}\{y''\}$
Clearance factor	$\frac{L^{1/2}\,\beta}{\sum_{l=1}^{L}\lvert y_l\rvert^{1/2}}$	Acceleration derivate $\boldsymbol{\kappa}$	$\boldsymbol{\kappa}\{y'''\}$

Table 2. Statistical parameters from the frequency spectrum ($\lambda, s \in \mathbb{R}^K$). Source: Authors.

Feature	Expression	Feature	Expression
Mean frequency ($\boldsymbol{\mu_\lambda}$)	$\sum_{k=1}^{K}\frac{s_k}{K}$	Standard deviation frequency	$\left(\sum_{k=1}^{K}\frac{(\lambda_k-\Lambda)^2 s_k}{K\mu_\lambda}\right)^{1/2}$
Central frequency ($\boldsymbol{\Lambda}$)	$\sum_{k=1}^{K}\frac{\lambda_k s_k}{K\mu_\lambda}$	Kurtosis	$\sum_{k=1}^{K}\frac{s_k{}^4}{K\mu_\lambda{}^2}$
Root mean square frequency	$\boldsymbol{\Lambda}^{1/2}$		

where $d: \mathbb{R}\times\mathbb{R}\longrightarrow\mathbb{R}$ is a given distance function, $\Omega_n^c = \{\boldsymbol{f}_{n'}: n' = 1,2,\cdots,\vartheta\}$ holds the ϑ-nearest neighbors of $\boldsymbol{f_n}$ according to d, and $p(\varsigma=c)\in[0,1]$ is the probability that a sample belongs to the c-th class ($c\in\{1,2,\cdots,C\}$; $q\in\{1,2,\cdots,Q\}$). In this sense, the higher ρ_q value the better the q-th feature for discriminating combustion categories. As a result, the calculated relevance vector $\boldsymbol{\rho}$ is employed to rank the multi-domain features. At the same time a classifier is trained. Fig. 1. shows the diagram of our Vibration-based Condition Monitoring of ICE using Feature Relevance Analysis.

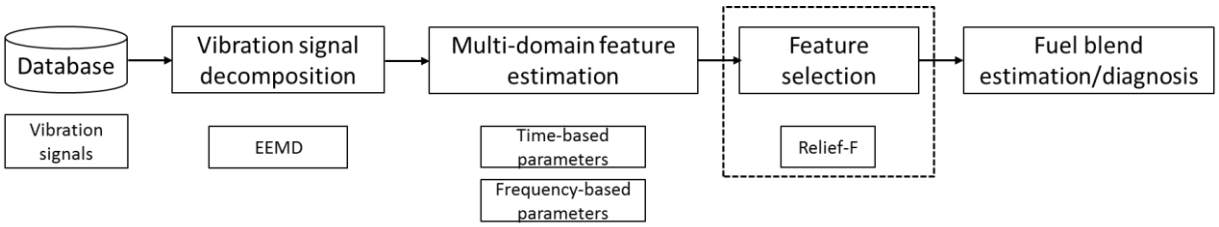

Fig. 1. Diagram of the proposed Vibration-based Condition Monitoring of ICE using Feature Selection. Source: Authors.

3. EXPERIMENTAL SETUP

3.1 Database and preprocessing

To assess the convenience of the proposed methodology for condition monitoring, we used a vibration dataset of an ICE collected by "The Engine Laboratory" from Universidad Tecnológica de Pereira (Colombia). Fig. 2 shows the data acquisition strategy. It illustrates a four-cylinder ICE and three accelerometers attached to the engine block by magnetic bases on the vertical, longitudinal, and transversal axes (*VA*, *LA*, and *TA*) for piston slap vibration. Moreover, to facilitate the segmentation of the combustion cycle, a pressure and a proximity sensor were used to measure the pressure inside a combustion camera and the angle of the engine crankshaft, respectively. All instruments were fed data acquisition modules from National Instruments (references 9232, 9234, 9174) [46]. In order to evaluate the proposed fuel blend estimation/diagnosis strategy, the vibration, pressure, and angle signals were recorded simultaneously from the engine under the following operating conditions: three ethanol-gasoline fuel blends (E10, E20 and E30), two engine states (normal and simulated misfire), and three revolution operations (~1500 rpm, ~1700 rpm, and ~2000 rpm). Since quick pressure changes in a cylinder cause engine structure vibrations during combustion, one cycle is considered to be sufficient to code relevant information regarding the studied process [47]. Thereby, the data acquisition time is fixed to two seconds after stabilizing the engine without load, because a reasonable amount of combustion cycles are obtained within such time window length. As a consequence, 54 two-second long signals with a sample rate value of $\Lambda_s = 51.2\ kHz$ were acquired. Then, after combustion cycle segmentation, a matrix $\mathrm{X} \in \mathbb{R}^{N \times T}$ was constructed with $N = 1524$ and $T \in \{\sim 78.12, \sim 68.36, \sim 58.59\} ms.$ The latter depends on the revolution operation value. Three different fuel blend estimation/diagnosis scenarios were considered for engine condition monitoring by varying the label vector $\varsigma \in \{1,2,\cdots,C\}^N$: *i)* Class 3 problem ($C=3$), learning the fuel blend (E10, E20 and E30); *ii)* Class 6 problem ($C=6$), fuel mixture and engine state diagnosis (E10, E20, and E30 under normal or misfire state); and *iii)* Class 18 problem ($C=18$), estimation of the fuel blend plus engine state diagnosis and revolution operation (E10, E20, and E30 ethanol-gasoline fuel blend under normal or misfire state operating at ~1500 rpm, ~1700 rpm, and ~2000 rpm).

3.2 Training stage

Our proposal was tested on the above mentioned dataset. We decomposed each vibration segment using the EEMD technique. As suggested by authors in [26], $J = 8$ TIMFs are computed setting $\sigma_\varepsilon = 0.2$ and $\mathrm{M} = 200$. Afterwards, 4527 multi-domain features were calculated: 162 and 4365 for time (T) and frequency (F) domain, respectively. Regarding the frequency-based features, the spectrum vector size was set at $K = 2048$ to compute the parameters in Table 2. We called such features *Frequency set-1* (*F1*). Besides, we

fixed the number of bands at $B = 80$ and $B = 20$ for the raw and the TIMFs signals, respectively. A Hamming-based windowing with 50% of overlap was used for band analysis. The window size was calculated: $w_s = \lfloor (2 * K/(B+1)) \rfloor$. At this point, we set $K = 8192$ to compute the parameters in Table 1 (*Frequency set-2, F2*) and meet suitable statistical estimations. Then, we constructed the feature matrix $F \in \mathbb{R}^{N \times Q}$ holding $N =$ 1517 vibration segments with Q=4527 features. This matrix was employed to rank the relevance of each given multi-domain parameter based on the *Relief-F* technique. We set two parameters to compute the relevance vector in Eq. (2): neighborhood size value $\vartheta = 1$ and distance function $d(f_{np}, f_{n'p}) = |f_{np} - f_{n'p}| / (max_n(f_{np}) - min_n(f_{np}))$ [45]. The validation is assessed by estimating the classification performance. A k-Nearest Neighbors classifier was applied to all the considered fuel blend estimation/diagnosis scenarios. The number of nearest neighbors in the classifier was the one providing the best accuracy within the following testing range {3, 5, 7, 9}. Namely, we calculated the accuracy of the performance by using a nested 10-fold cross-validation scheme and adding, one by one, the features ranked by the amplitude of $\boldsymbol{\rho}$. Then, based on each given ranking of the relevant features, the original data was randomly partitioned into 10 equal sized subgroups. Then, a single subset was earmarked for testing the model and the remaining nine were used as training. A 10-fold partition is carried out again on such training set to learn the classifier's free parameter (number of neighbors). Finally, the computed parameter is employed to calculate the system accuracy on the testing set. This process is repeated depending on the number of folds to achieve an average accuracy and its standard deviation.

3.3 Method comparison

For comparison purposes, we considered the following representative unsupervised and supervised feature selection methods: *i) The Variance-based Relevance Analysis* (*VRA*) ranks the input features based on a variability criterion [48]. *ii)* The *Self-weight ranking* codes the feature relevance in terms of a self-similarity measure [26]. *iii)* The *Laplacian Score approach* computes a Laplacian Graph from input samples to highlight the importance of each provided characteristic from graph edges [49]. iv) The *Distance-weight* method quantifies the relevance of the distance between samples from different clusters by using supervised information [50]. The required free parameters were fixed following the descriptions in the literature of each approach.

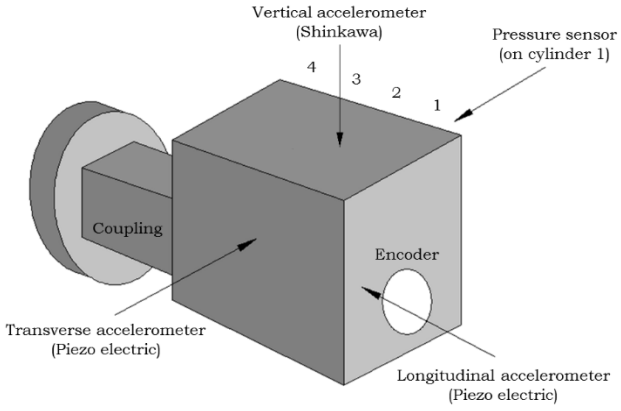

Fig. 2. Illustration of the vibration signal recording on a four-cylinder ICE. Source: Authors

3. RESULTS AND DISCUSSION

First, we introduced an illustrative example of the time and frequency parameters concerning the studied classes. Fig. 3 shows the time-domain waveform of some vibration segments presenting the *C=6* problem along with their corresponding frequency spectrums. As it can be seen, the frequency spectrum varies in every state. Particularly, the greatest amount of information is contained until 10 kHz. Besides, Fig. 4 presents the first 8 TIMFs obtained with the EEMD applied to some LA vibration segments (*C=6* problem) in time and frequency domains. We notice that the TIMF magnitudes decrease progressively. Thus, the first 4 TIMFs code the main energy variations of the raw vibration signal. Subsequently, we established the performed accuracy after feature ranking for the *C=18* problem using the LA sensor (Fig. 5). Also, the most significant result, in terms of achieved accuracy vs. the number of relevant features required, was found for each relevance approach. As shown, the lowest results were obtained for *VRA* {89.08±2.84} %, *Self-weight* {84.98±2.94} %, and *Laplacian-score* {85.65±1.35} %. Therefore, it can be assumed that unsupervised-based approaches are not able to highlight relevant features from complex ICE categories. In turn, the supervised algorithms, *Distance-weight* and *Relief-F*, achieved acceptable accuracies: {96.31±1.67} % and {95.60±1.18} %, respectively. However, *Relief-F* requires only 301 features to accurately classify ICE categories. Conversely, the *Distance-weight* approach employs 1501 features. Thus, the suggested *Relief-F* finds the lowest number of relevant parameters providing a highly accurate classification for ICE analysis.

Furthermore, to highlight the relevance of the multi-domain parameters, Fig. 6 shows the percentage of *Relief-F*-based selected features from *T, F1,* and *F2* sets in the *C=18* problem (*VA, LA* and *TA* sensors are considered). Remarkably, nearly **60%** of the relevant features belong to the *F2* set. In fact, from a mechanical point of view, vibration signals normally consist of a combination of the fundamental frequency, a narrowband component, and the harmonics. For this reason, when a mechanical element undergoes an abnormal state, discriminative patterns are reflected on the frequency spectrum [26]. To illustrate the latter, a 2D projected space is computed from the *Relief-F*-based ranking of the F2 set. For this study, we employed the well-known *t-Student Stochastic Neighbor Embedding* (*t-SNE*) algorithm that finds a low-dimensional space using a Kullback-Leibler similarity preservation cost function [51]. As seen in Fig. 7, the samples were divided into three subgroups that represent the different operating speeds. Additionally, each subgroup shows a suitable separability between fuel mixtures and engine states.

Moreover, Tables 3, 4, and 5 show the achieved classification results regarding the vibration sensor (*LA, VA,* and *TA*) in all the provided ICE analysis scenarios (*C=3, C=6,* and *C=18*). In general terms, our feature relevance analysis strategy outperforms state-of-the-art algorithms in classification accuracy and the number of selected features. This is because our approach addresses different ICE analysis scenarios from various sensor positions. Lastly, it is important to note that the *F2* set with a *Relief-F*-based selection provides an excellent alternative for ICE fuel blend estimation/diagnosis.

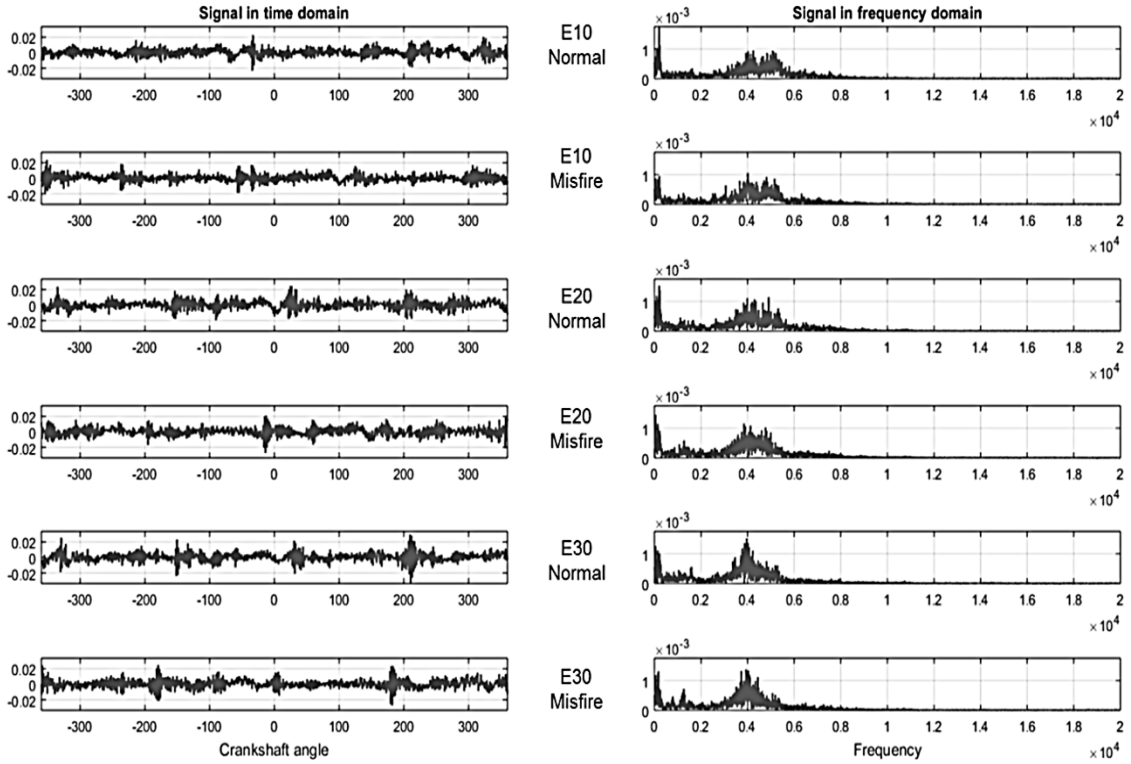

Fig. 3. Samples of vibration signals (*LA*) and their corresponding spectrums.
Fuel blend and engine state labels are shown. Source: Authors

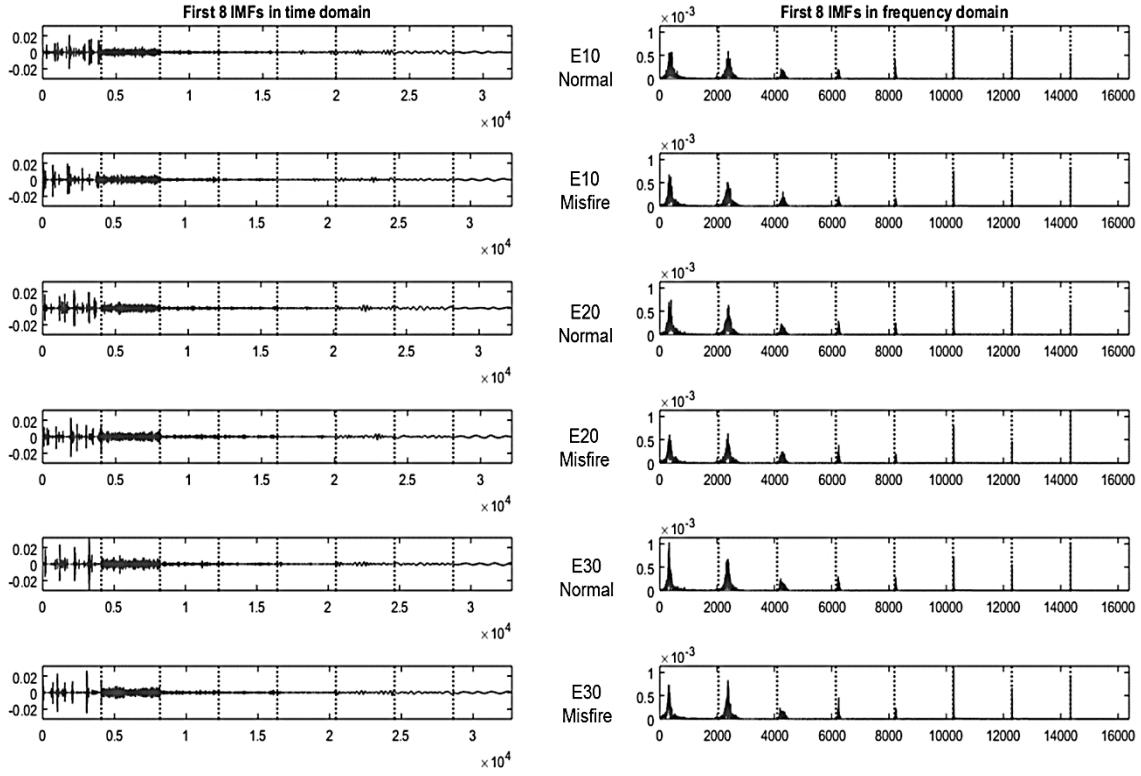

Fig. 4. Samples of the achieved TIMF from vibration signals (*LA*) and their spectrums.
Fuel blend and engine state labels are shown. Source: Authors

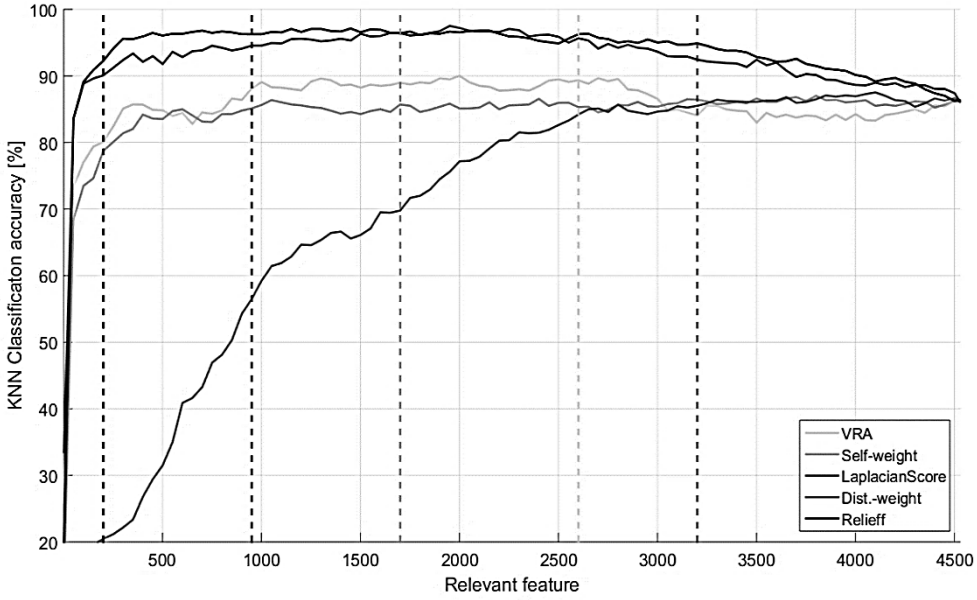

Fig. 5. Performed accuracy after feature ranking for the *C=18* problem using the *LA*. Significant results are marked by dashed lines. Source: Authors

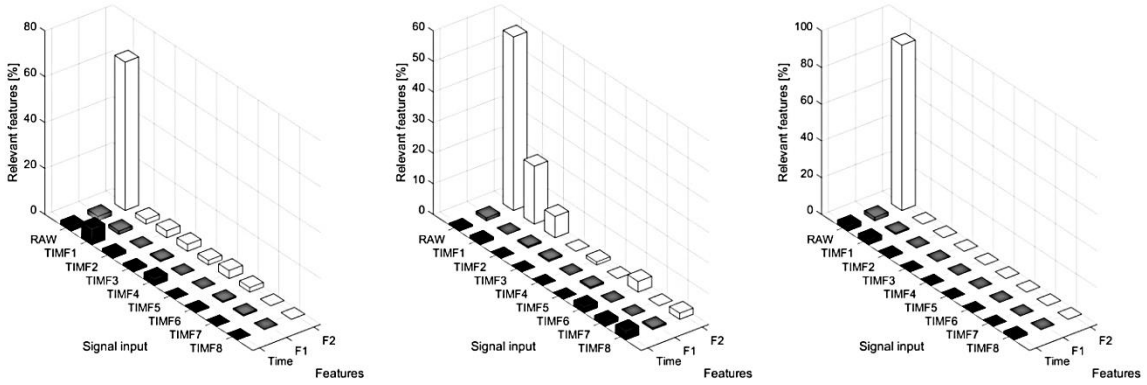

Fig. 6. Percentage of selected features within the relevant subset estimated with Relief-F for the *C=18* scenario. Left: *VA*, Middle: *LA*, Right: *TA*. Source: Authors

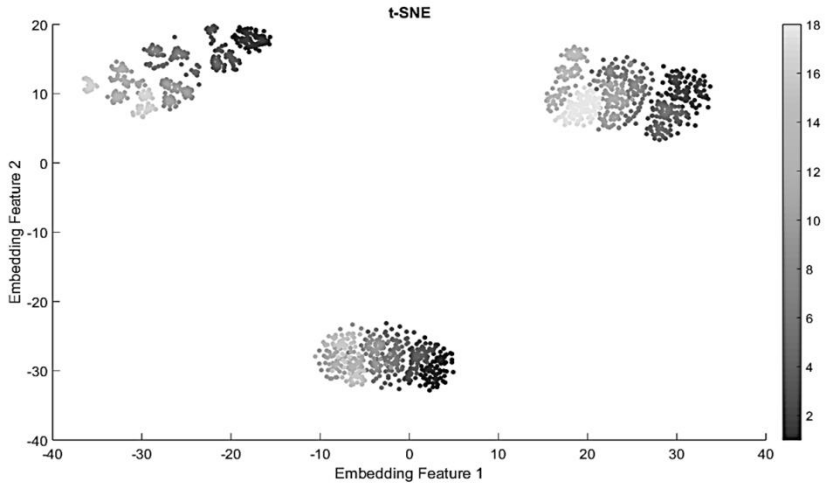

Fig. 7. Projected space using the *t-SNE* technique applied to the *F2* set after *Relief-F*-based ranking. The *C=18* problem is presented for *LA*. Source: Authors.

Table 3. Obtained classification results regarding the required number of features. The best results of performed accuracy (ACC) and feature ranking are shown. for *LA*. N. Feat stands for number of required features. Source: Authors.

Method	C=3		C=6		C=18	
	ACC [%]	N. Feat.	ACC [%]	N. Feat.	ACC [%]	N. Feat.
All features	92.99±1.89	4527	86.09±3.27	4527	86.24±2.54	4527
VRA [49]	95.08±1.07	951	89.02±3.65	1001	89.08±2.84	1001
Self-weight [26]	91.99±3.04	451	85.44±3.11	551	84.98±2.92	601
Laplacian score [50]	93.04±1.91	3001	85.42±3.19	2701	85.69±1.35	2801
Distance weight [51]	98.10±1.53	401	94.16±1.55	951	96.31±1.67	1501
Relief-F	**98.82±0.51**	**201**	**95.34±1.63**	**301**	**95.60±1.18**	**301**
F2 + Relief-F	97.75±0.57	119	91.66±1.09	184	92.27±0.88	184

Table 4. Obtained classification results regarding the required number of features. The best results of performed accuracy (ACC) and feature ranking are shown. for *VA*. N. Feat. stands for number of required features. Font: Authors.

Method	C=3		C=6		C=18	
	ACC [%]	N. Feat.	ACC [%]	N. Feat.	ACC [%]	N. Feat.
All features	93.88±2.79	4527	91.60±2.01	4527	91.56±2.14	4527
VRA [49]	99.34±0.69	251	98.30±1.02	251	98.29±0.70	251
Self-weight [26]	92.78±2.25	851	91.08±2.22	1251	90.54±2.16	1451
Laplacian score [50]	93.17±2.67	3751	90.16±2.76	3451	90.56±1.12	3751
Distance weight [51]	98.29±0.83	901	96.52±1.23	951	97.77±1.28	451
Relief-F	**99.08±0.83**	**201**	**98.49±0.88**	**151**	**98.62±0.66**	**151**
F2 + Relief-F	96.72±0.71	120	96.37±0.70	94	95.97±0.74	95

Table 5. Obtained classification results regarding the required number of features. The best results of performed accuracy (ACC) and feature ranking are shown for *TA*. N. Feat stands for number of required features. Font: Authors.

Method	C=3		C=6		C=18	
	ACC [%]	N. Feat.	ACC [%]	N. Feat.	Acc.	N. Feat.
All features	89.04±2.25	4527	84.72±4.51	4527	85.03±1.56	4527
VRA [49]	91.73±2.78	2751	87.81±3.86	2601	87.21±2.30	2151
Self-weight [26]	90.16±2.18	2251	85.77±4.05	1701	83.62±2.52	1151
Laplacian score [50]	90.75±1.83	3251	88.00±2.50	3201	88.45±3.24	3201
Distance weight [51]	96.59±1.36	651	96.78±1.13	951	96.31±1.22	851
Relief-F	**99.21±0.75**	**201**	**97.97±0.84**	**201**	**97.96±0.99**	**151**
F2 + Relief-F	97.05±0.87	181	95.27±1.19	184	95.43±0.92	136

4. CONCLUSIONS

We presented a feature relevance estimation strategy to enhance vibration-based condition monitoring of ICE. In this sense, our approach incorporated a signal decomposition stage based on the EEMD algorithm to reveal nonlinear and non-stationary patterns. Moreover, multi-domain parameters were computed from both the raw vibration signal and the EEMD-based decompositions. The aim was to code high-order moments from time domain and energy variations in the frequency spectrum. Furthermore, we computed a relevance index vector using a *Relief-F*-based function to measure the contribution of each input feature to the discrimination of different fuel blend estimation/diagnosis categories. Therefore, we assessed the selected feature set that meets a given stopping criteria in terms of classification accuracy vs. number of relevant features. We tested our strategy on a vibration dataset of an ICE collected by "The Engine Laboratory" from Universidad Tecnológica de Pereira (Colombia). The set contains different combustion analysis scenarios, namely: three fuel blends, two engine states (normal and misfire), and three operation speeds. Our strategy highlights intra-band variations in frequency domain as the most relevant features. As a result, is outperforms the compared approaches that carry out unsupervised and supervised feature selection and achieve suitable classification accuracy with the lowest number of relevant features.

As future work, the authors plan to test the introduced feature relevance analysis on different vibration-based fault diagnosis tasks, e.g., bearing and gearbox fault diagnosis [52]. Moreover, improving the feature relevance index by using information theory and nonlinear mapping functions would be an interesting task [53]. Finally, our process for feature relevance evaluation can be complemented by redundancy analysis, based on linear and nonlinear correlation methods of unsupervised learning. After that, it would be able to deal with complex cluster distributions and lighten the computational burden [26].

5. ACKNOWLEDGEMENTS

This research was supported by COLCIENCIAS. Project name: "Diseño y desarrollo de un sistema prototipo en línea para el diagnóstico de motores de combustión interna". Project code: 1110-669-46074.

6. REFERENCES

[1] J. Flett and G. M. Bone, "Fault detection and diagnosis of diesel engine valve trains," *Mech. Syst. Signal Process.*, vol. 72–73, pp. 316–327, May 2016.

[2] D. Martínez-Rego, O. Fontenla-Romero, A. Alonso-Betanzos, and J. C. Principe, "Fault detection via recurrence time statistics and one-class classification," *Pattern Recognit. Lett.*, vol. 84, pp. 8–14, Dec. 2016.

[3] N. D. Liyanagedera, A. Ratnaweera, and D. I. B. Randeniya, "Vibration signal analysis for fault detection of combustion engine using neural network," in *2013 IEEE 8th International Conference on Industrial and Information Systems*, 2013, pp. 427–432.

[4] B. Samimy and G. Rizzoni, "Mechanical signature analysis using time-frequency signal processing: application to internal combustion engine knock detection," *Proc. IEEE*, vol. 84, no. 9, pp. 1330–1343, 1996.

[5] J.-D. Wu and C.-Q. Chuang, "Fault diagnosis of internal combustion engines using visual dot patterns of acoustic and vibration signals," *NDT E Int.*, vol. 38, no. 8, pp. 605–614, Dec. 2005.

[6] L. Barelli, G. Bidini, C. Buratti, and R. Mariani, "Diagnosis of internal combustion engine through vibration and acoustic pressure non-intrusive measurements," *Appl. Therm. Eng.*, vol. 29, no. 8–9, pp. 1707–1713, Jun. 2009.

[7] F. Payri, J. M. Luján, J. Martín, and A. Abbad, "Digital signal processing of in-cylinder pressure for combustion diagnosis of internal combustion engines," *Mech. Syst. Signal Process.*, vol. 24, no. 6, pp. 1767–1784, Aug. 2010.

[8] J. Chen, R. B. Randall, and B. Peeters,

"Advanced diagnostic system for piston slap faults in IC engines, based on the non-stationary characteristics of the vibration signals," *Mech. Syst. Signal Process.*, vol. 75, pp. 434–454, Jun. 2016.

[9] L. Xu and H. E. Tseng, "Robust model-based fault detection for a roll stability system," *IEEE Trans. Control Syst. Technol.*, vol. 15, no. 3, pp. 519–528, May 2007.

[10] Xuewu Dai, Zhiwei Gao, T. Breikin, and Hong Wang, "Disturbance Attenuation in Fault Detection of Gas Turbine Engines: A Discrete Robust Observer Design," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 39, no. 2, pp. 234–239, Mar. 2009.

[11] H. R. Karimi, M. Zapateiro, and N. Luo, "A linear matrix inequality approach to robust fault detection filter design of linear systems with mixed time-varying delays and nonlinear perturbations," *J. Franklin Inst.*, vol. 347, no. 6, pp. 957–973, Aug. 2010.

[12] Y. Zhu and Z. Gao, "Robust observer-based fault detection via evolutionary optimization with applications to wind turbine systems," in *2014 9th IEEE Conference on Industrial Electronics and Applications*, 2014, pp. 1627–1632.

[13] S. Ding, *Model-based fault diagnosis techniques: design schemes, algorithms, and tools*. Springer Science & Business Media, 2008.

[14] J. Chen and R. J. Patton, *Robust model-based fault diagnosis for dynamic systems*, vol. 3. Springer Science & Business Media, 2012.

[15] Z. Gao, C. Cecati, and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques-Part I: Fault Diagnosis With Model-Based and Signal-Based Approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.

[16] G. O. Chandroth, A. J. C. Sharkey, and N. E. Sharkey, "Cylinder pressures and vibration in internal combustion engine condition monitoring," in *Proceedings of Comadem*, 1999, vol. 99, pp. 294–297.

[17] J. Antoni, J. Daniere, F. Guillet, and R. B. Randall, "Effective vibration analysis of IC engines using cyclostationarity. Part II-new results on the reconstruction of the cylinder pressures," *J. Sound Vib.*, vol. 257, no. 5, pp. 839–856, Nov. 2002.

[18] S. A. Ali and S. Saraswati, "Reconstruction of cylinder pressure using crankshaft speed fluctuations," in *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, 2015, pp. 456–461.

[19] X. Zhao, Y. Cheng, and S. Ji, "Combustion parameters identification and correction in diesel engine via vibration acceleration signal," *Appl. Acoust.*, vol. 116, pp. 205–215, Jan. 2017.

[20] T. Denton, *Advance Automotive Fault Diagnosis: Auto-motive Technology: Vehicle Maintenance and Repair*. Routledge, 2016.

[21] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006.

[22] F. Al-Badour, M. Sunar, and L. Cheded, "Vibration analysis of rotating machinery using time–frequency analysis and wavelet techniques," *Mech. Syst. Signal Process.*, vol. 25, no. 6, pp. 2083–2101, Aug. 2011.

[23] Y. Shatnawi and M. Al-khassaweneh, "Fault Diagnosis in Internal Combustion Engines Using Extension Neural Network," *IEEE Trans. Ind. Electron.*, vol. 61, no. 3, pp. 1434–1443, Mar. 2014.

[24] Y. Lei, Z. He, and Y. Zi, "Application of the EEMD method to rotor fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 23, no. 4, pp. 1327–1338, May 2009.

[25] M. Buzzoni, E. Mucchi, and G. Dalpiaz, "A CWT-based methodology for piston slap experimental characterization," *Mech. Syst. Signal Process.*, vol. 86, pp. 16–28, Mar. 2017.

[26] Z. Wei, Y. Wang, S. He, and J. Bao, "A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection," *Knowledge-Based Syst.*, vol. 116, pp. 1–12, Jan. 2017.

[27] S. Ericsson, N. Grip, E. Johansson, L.-E. Persson, R. Sjöberg, and J.-O. Strömberg, "Towards automatic detection of local bearing defects in rotating machines," *Mech. Syst. Signal Process.*, vol. 19, no. 3, pp. 509–535, May 2005.

[28] A. Taghizadeh-Alisaraei, B. Ghobadian, T. Tavakoli-Hashjin, S. S. Mohtasebi, A. Rezaei-asl, and M. Azadbakht, "Characterization of engine's combustion-vibration using diesel and biodiesel fuel blends by time-frequency methods: A case study," *Renew. Energy*, vol. 95, pp. 422–432, Sep. 2016.

[29] J. Da Wu and J. C. Chen, "Continuous wavelet transform technique for fault signal diagnosis of internal combustion engines," *NDT E Int.*, vol. 39, no. 4, pp. 304–311, Jun. 2006.

[30] A. Moosavian, G. Najafi, B. Ghobadian, M. Mirsalim, S. M. Jafari, and P. Sharghi, "Piston scuffing fault and its identification in an IC engine by vibration analysis," *Appl. Acoust.*, vol. 102, pp. 40–48, Jan. 2016.

[31] Z. Liu, X. Chen, Z. He, and Z. Shen, "LMD Method and Multi-Class RWSVM of Fault Diagnosis for Rotating Machinery Using Condition Monitoring Information," *Sensors*, vol. 13, no. 7, pp. 8679–8694, Jul. 2013.

[32] S. S. H. Zaidi, S. Aviyente, M. Salman, K.-K. Shin, and E. G. Strangas, "Prognosis of Gear Failures in DC Starter Motors Using Hidden Markov Models," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1695–1706, May 2011.

[33] J. Da Wu and C. H. Liu, "An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network," *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4278–4286, Apr. 2009.

[34] M. A. Rizvi, A. I. Bhatti, and Q. R. Butt, "Hybrid Model of the Gasoline Engine for Misfire Detection," *IEEE Trans. Ind. Electron.*, vol. 58, no. 8, pp. 3680–3692, Aug. 2011.

[35] F. Fiippetti and P. Vas, "Recent developments of induction motor drives fault diagnosis using AI techniques," in *IECON '98. Proceedings of the 24th Annual Conference of the IEEE Industrial Electronics Society (Cat. No.98CH36200)*, 2000, vol. 4, no. 5, pp. 1966–1973.

[36] Q. Hu, Z. He, Z. Zhang, and Y. Zi, "Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 688–705, Feb. 2007.

[37] Y. S. Wang, Q. H. Ma, Q. Zhu, X. T. Liu, and L. H. Zhao, "An intelligent approach for engine fault diagnosis based on Hilbert–Huang transform and support vector machine," *Appl. Acoust.*, vol. 75, pp. 1–9, Jan. 2014.

[38] L. Liang, F. Liu, M. Li, K. He, and G. Xu, "Feature selection for machine fault diagnosis using clustering of non-negation matrix factorization," *Measurement*, vol. 94, pp. 295–305, Dec. 2016.

[39] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing Fault Detection by a Novel Condition-Monitoring Scheme Based on Statistical-Time Features and Neural Networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.

[40] A. Malhi and R. X. Gao, "PCA-Based Feature Selection Scheme for Machine Defect Classification," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004.

[41] R. Shao, W. Hu, Y. Wang, and X. Qi, "The fault feature extraction and classification of gear using principal component analysis and kernel principal component analysis based on the wavelet packet transform," *Measurement*, vol. 54, pp. 118–132, Aug. 2014.

[42] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, Jan. 2009.

[43] C. Li, J. Valente de Oliveira, M. Cerrada, F. Pacheco, D. Cabrera, V. Sanchez, and G. Zurita, "Observer-biased bearing condition monitoring: From fault detection to multi-fault classification," *Eng. Appl. Artif. Intell.*, vol. 50, pp. 287–301, Apr. 2016.

[44] Y. Chen, X. Pei, S. Nie, and Y. Kang, "Monitoring and Diagnosis for the DC-DC Converter Using the Magnetic Near Field Waveform," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1634–1647, May 2011.

[45] M. Robnik-Šikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Mach. Learn.*, vol. 53, no. 1/2, pp. 23–69, 2003.

[46] J. A. Grajales, H. F. Quintero, C. A. Romero, E. Henao, J. F. López, and D. Torres, "Combustion pressure estimation method of a spark ignited combustion engine based on vibration signal processing," *J. Vibroengineering*, vol. 18, no. 7, pp. 4237–4247, Nov. 2016.

[47] R. Johnsson, "Cylinder pressure reconstruction based on complex radial basis function networks from vibration and speed signals," *Mech. Syst. Signal Process.*, vol. 20, no. 8, pp. 1923–1940, Nov. 2006.

[48] G. Daza-Santacoloma, J. D. Arias-Londono, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ru'\iz, and G. Castellanos-Dominguez, "Dynamic feature extraction: an application to voice pathology detection," *Intell. Autom. Soft Comput.*, vol. 15, no. 4, pp. 667–682, 2009.

[49] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Neural Information Processing Systems, NIPS 2005*, 2005, vol. 18, p. 189.

[50] B. S. Yang, T. Han, and J. L. An, "ART–KOHONEN neural network for fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 18, no. 3, pp. 645–657, May 2004.

[51] J. A. Lee, E. Renard, G. Bernard, P. Dupont, and M. Verleysen, "Type 1 and 2 mixtures of Kullback–Leibler divergences as cost functions in dimensionality reduction based on similarity preservation," *Neurocomputing*, vol. 112, pp. 92–108, Jul. 2013.

[52] C. Verucchi, G. Bossio, J. Bossio, and G. Acosta, "Fault detection in gear box with induction motors: an experimental study," *IEEE Lat. Am. Trans.*, vol. 14, no. 6, pp. 2726–2731, Jun. 2016.

[53] A. M. Álvarez-Meza, J. A. Lee, M. Verleysen, and G. Castellanos-Domínguez, "Kernel-based dimensionality reduction using Renyi's α-entropy measures of similarity," *Neurocomputing*, vol. 222, pp. 36–46, Jan. 2017.