Use of Multiscale Permutation Entropy Feature Selection and Supervised Classifiers for Bearing Failures Diagnosis

Abstract—Entropy measurements are an accessible tool to perform irregularity and uncertainty measurements present in time series. In signal processing, the Multiscale Permutation Entropy is recently presented as a methodology of characterization capable of measuring randomness and non-linear dynamics existing in non-stationary time series, such as mechanical vibration signals. In this article, the Multiscale Permutation Entropy is combined with diverse feature selection techniques and multiple classifiers based on machine learning aiming to detect different operative states in an internal combustion engine. The best combination is selected from the evaluation of parameters like precision and computational time. Finally, the proposed methodology is established as an effective tool to diagnose failures in bearing systems with a high precision rate and a reduced calculation time.

Index Terms—Dynamics, Entropy, Machine, Multiscale, Permutation, Vibration.

I. INTRODUCTION

BEARING systems play an important role in rotary machines and in the modern manufacturing industry. Different methodologies have been developed for the detection and diagnosis of faults in its main components [1]. Generally, these diagnoses are made from the capture and processing of vibration signals, since they contain relevant information about the state of the machine [2]. However, these signals have a large number of non-stationary and non-linear characteristics, since their capture inevitably takes place with friction and impacts. To overcome this problem, a series of techniques have been developed for processing and classifying these signals. A widely used approach is based on the analysis of temporal and spectral characteristics of vibration signals [3] [4]. However, analysis in time, frequency and time-frequency domains are seriously affected by the signal length and sampling frequency of the capture [5]. To solve this problem another approach has been presented in recent years, which is based on entropy of different natures, such as Simple Entropy (ApEn) [6], Approximate Entropy (SampEn) [7], Multiscale Entropy (MSE) [8], Permutation Entropy (PE) [9][31] and Multiscale Permutation Entropy (MPE) [10]. All of the above mentioned have been used successfully for the characterization of signals of different nature. For instance, in [6] the ApEn is used for diagnosis and clinical monitoring in the area of physiology. In [11] PE is used for the classification of patients from EEG signals. Finally, in [1][8][32] is used the MPE and MSE for the identification and diagnosis of faults in bearing systems based on vibration signals. It should be noted that the MPE is an evolution of the MSE and the PE since it...
gives a much more complete measure of the nonlinear dynamic parameters of a system. MPE includes a combination of different scales and time delays, which identifies particularities that are not perceptible within other entropies [10][33]. All the previous methods of characterization allow to obtain a great quantity of information of the system that is being analyzed. When classifying, many of these characteristics can be redundant or irrelevant, which can create redundancy or over-training of the classifier, for which characteristics selection techniques are implemented, improving the quality and efficiency of the model. In the bearing fault diagnosis application, Variance based on Relevance Analysis (VRA) [1], Laplacian Score (LS) [12] and Relief (REL) [5] are normally used. After the selection of features, a classification process is performed with machine learning algorithms such as Multiple Vector Support Machines (SVMM) [1], [13], [14] and [15], Hidden Markov Models (HMM) [16] and Artificial Neural Networks (ANN) [17], [18] and [19]. However, these classifiers have a high degree of complexity, computation time and initial parameters that must be optimized. Few works have attempted to exploit the potential of less complex conventional classifiers, such as Nearest K-Neighbors (KNN) [20], [21], Decision Trees (TREE) [22] or Naïve Bayes (BAYES) [21]. This paper proposes a methodology for the diagnosis of bearing failures based on characteristics of the MPE. To make it more efficient and effective, it is combined with some feature selection techniques and multiple supervised classifiers. The verification of the advantage of the chosen parameters is done by comparison with different ones used in the state of the art.

The rest of the article is organized as follows: In section II.A is presented the mathematical formulation of the MPE and in section II.B is detailed the different feature selection techniques used. In section II.C the classifier used in the state of the art are exposed. Then, a novel methodology for the diagnosis of bearing failures based on MPE, feature selection techniques and supervised classifier is located in the section III. Finally, results are presented in section IV and the conclusions of this approach in section V.

II. METHODOLOGY

The proposed methodology combines a characterization method, a feature selection technique and a supervised classifier as show in Fig. 1. Each part of the methodology is described in the following sections. The proposed methodology begins with the characterization of vibration signals, then using an automatic classifier combined with a characteristic selection technique, the state of the system is estimated.

A. Multiscale Permutation Entropy and Acronyms

Multiscale Permutation Entropy (MPE) is used in this paper as a signal characterization method, since it is a measure that allows to detect dynamic changes in the time series. It is based on the comparison in neighboring values without taking into account the size of the values and, therefore, has a calculation simple and fast [23]. The above, allows to position the MPE as a particularly useful and robust tool in the presence of dynamic noise [1]. In order to describe the MPE measure, it is important to review the entropy proposed by Shannon [14], described as follows: Considering a time series $\{x_i\}_{i=1}^T$ in a space of representation of characteristics, where $T$ is the length of the time series, the entropy is represented as in (1):

$$H(X) = -\sum_{i=1}^T p(x_i) \ln p(x_i)$$  \hspace{1cm} (1)

Where $x_i \in \mathbb{R}$ and $p(x_i)$ is the marginal probability. The time series can be represented with a delay of time and dimension given by (2):

$$X_i^{m,\tau} = (x_i, x_{i+\tau}, ..., x_{i+(m-1)\tau})$$  \hspace{1cm} (2)

Where $i = 1, 2, ..., T - (m-1)\tau$, $m$ is the dimension and $\tau$ the delay. To perform the computation of MPE, the signal must be truncated in $N = T - (m - 1)\tau$ subvectors. Then, for each subvector is calculated the entropy mapped in a space of $m$ different symbols denoted as $\sigma_i^{m,\tau}$ by (3):

$$H(m,\tau) = -\sum_{i,\sigma_i^{m,\tau}} p(\sigma_i^{m,\tau}) \ln p(\sigma_i^{m,\tau})$$  \hspace{1cm} (3)

The probability $p(\sigma_i^{m,\tau})$ is calculated by (4):

$$p(\sigma_i^{m,\tau}) = \sum_{j \leq N} 1_{u_{\text{type}(u)=\tau_i} \{X_j^{m,\tau}\}} \sum_{j} 1_{u_{\text{type}(u)=\tau_i} \{X_j^{m,\tau}\}}$$  \hspace{1cm} (4)

Where the judgment type denotes the map from pattern
space to symbol space. Also, \( 1_A(u) = 1 \) if \( u \in A \) and \( 1_A(u) = 0 \) if \( u \notin A \). The MPE can take values between the ranges \([0, \ln(m!)]\) and it is invariant under nonlinear monotonic transformations. The values of \( m \) and \( \tau \) vary from 1 to 8 and the values, which are used for calculating \( \text{MSE} \) [8] and \( \text{PE} \) [9]. In Fig 2 is plotted the behavior of the MPE, when are varying \( m \) and \( \tau \) for a vibration signal.

When choosing the delay and dimension parameters, the nature of the signals must be taken into account. If the parameters are too small, the nonlinear dynamic of the characteristics from signals will not be analyzed effectively. If the parameters are too large, useful information will be deleted in consequence, which will result in a wicked analysis.

B. Featured Selection Technique

For the comparison of the characteristic selection technique, the ones most used in the literature were implemented.

- **Relief (REL).** Unsupervised method to generate a ranking based on the predictors give neighbors of the same class or different class.
- **Laplacian Score (LS).** Unsupervised method to generate a ranking based on the input characteristics based on a variability criterion [1].
- **Variance-based Relevance Analysis (VRA).** Unsupervised method, it generates a ranking with the input features based on a variability criterion [12].
- **Non-negative matrix factorization (NF).** Unsupervised method to analyze the relevance of the characteristics from a reduction in dimensionality by non-negative factorization techniques [20].
- **Self-Weight Ranking (SW).** Unsupervised method, technique, it codes the feature relevance in terms of a self-similarity measures [5].
- **Distance-Weight Ranking (DW).** Supervised method, this technique quantifies the relevance of the distance between samples from different clusters by using supervised information.

C. Supervised Classifiers

Characteristic selection techniques are combined with different supervised classifiers to find the best combination that suits the nature of the signal.

- **K-Nearest Neighbors (KNN).** Supervised and non-parametric classification method that estimates the posterior probability that an element belongs to the class from a set of information provided [24].
- **Decision Trees (TREE).** Supervised classifier based on prediction systems based on rules, which serve to represent and categorize a series of conditions that occur successively, for the resolution of a problem [22].
- **Naive bayes (BAYES).** Supervised classifier for multiclass learning. This classifier is based on estimation of prediction and re-substitution [21].
- **Multiple Support Vector Machines (SVMM).** Supervised classifier based on a hyperplane or set of hyperplanes in a space of very high (or even infinite) dimensionality that can be used in problems of classification or regression [1], [13], [15] and [14].

III. EXPERIMENTAL DATA

The validation of the proposed methodology is carried out by evaluation performed when classifying bearing fault signals obtained from the Case Western Reserve database [25]. In this database, signals were collected for the normal bearings (Nor), faults in the internal train (IR1), external train (IR2) and ball (BE). Faults are also found in order of severity, 0.007 inches in diameter to 0.040 inches in diameter and at variable engine speeds of 1720 to 1797 RPM. Each experiment was repeated three times and the data was collected at 12 kHz for 5 seconds. Each signal was divided into 10 sub-signals in order to have more samples per class and imitating the experimental framework established in the literature [26]. A sub-signal of each of the faults can be seen in Fig.3.

The signals are characterized with the MPE and then dividing it into training and testing. A feature selection technique is applied to the training group to obtain a ranking of relevance and it is applied to the test group. Then, with the reorganized training group, a classifier is trained and the test group is evaluated. The evaluation process is explained in Fig. 4.
IV. RESULTS AND DISCUSSION

This paper exposes a methodology for classify bearing fault in vibration signal. The classification carried out through a cross validation of \( k = 5 \) and repeated by changing the number of training characteristics. After the validation, the best results were chosen through the shortest distance to the ideal point \([100\% \ (Acc); \ 0 \ (Std); \ 0 \ (Ca)]\), where Acc is the accuracy, Std is the standard deviation of the accuracy of the cross validation and Ca is the number of characteristics. In tables II and II is exposed the accuracy with a 95% confidence interval for the combinations of classifiers and feature selection techniques for 4 (Nor, IR1, IR2 and BE) and 10 (Nor, IR1, IR2 and BE, combined with different speeds) classes are appreciated.

The best results are obtained with the KNN classifier, regardless of the feature selection technique. Specifically, the best accuracy was obtained with the KNN classifier combined with the Relief feature selection technique. The quantity of characteristics used for the classification were 9 and 10 for 4 and 10 classes respectively. It should be noted that the results are obtained thanks to the characterization made with the MPE, which achieves a high level of separability of the classes that allows the classifiers to adapt and solve the proposed application. Finally, a summary of the best classifications can be seen in the table III.

| TABLE III. COMPARISON OF THE BEST CLASSIFICATION RESULTS. |
|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| Author                          | Number Classes | Character. | Feat. Sel. | Classifier      | Number Features | Acc. % |
| Zhang et al.[14]                | 3               | PE+EMD     | -          | SVM              | 12              | 97.75  |
| Yuwono et al.[17]               | 3               | WPT        | -          | HMM              | 12              | 95.8   |
| Ben et al.[19]                  | 7               | TP+FR+EMD  | -          | ANN              | 10              | 93     |
| Zhu et al.[28]                  | 10              | HE+SE+MSE  | -          | SVM+PSO          | 9               | 97.75  |
| Han et al.[16]                  | 14              | SE+LDM     | -          | SVM              | -               | 100    |
| Liu et al.[29]                  | 4               | TP-FR      | -          | WPT+SVMM+PSO     | 81              | 97.5   |
| Tiwari et al.[2]                | 4               | MPE        | -          | ANFC             | 16              | 92.15  |
| William et al.[20]              | 4               | ZC         | -          | ANN              | 10              | 97.13  |
| Ocak et al.[30]                 | 3               | LPM        | -          | HMM              | 30              | 99.6   |
| Wei et al.[5]                   | 6               | FR+WPT     | Relief     | AP               | 18              | 96     |
| Shao et al.[31]                 | 16              | DAE+CAE    | LPP        | Softmax          | 19              | 96     |
| Zheng et al.[1]                 | 6               | GCMP+CA     | LS         | SVM+PSO          | 2               | 98.81  |
| Liang et al.[21]                | 4               | TP-FR      | NMF        | KNN              | 3               | 92.86  |
| Muru et al.[18]                 | 4               | SSA        | EMD        | ANN              | 10              | 95.14  |
| This work                       | 4               | MPE        | REL        | KNN              | 9               | 99.72  |
| This work                       | 10              | MPE        | REL        | KNN              | 10              | 99.55  |
The proposed experimental methodology achieved a mean accuracy of 99.72% for classification of 4 different bearing system failures. Table IV shows the confusion matrix for this classifier.

<table>
<thead>
<tr>
<th>Classified</th>
<th>Nor</th>
<th>IR1</th>
<th>IR2</th>
<th>BE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nor</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IR1</td>
<td>0</td>
<td>70</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>IR2</td>
<td>0</td>
<td>0</td>
<td>70</td>
<td>0</td>
</tr>
<tr>
<td>BE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>71</td>
</tr>
</tbody>
</table>

The proposed methodology is capable of classifying the failures with a high success rate, to the point that only one sample avoided obtaining 100% accuracy. The effectiveness of the methodology can be seen with also achieved a mean accuracy of 99.55% for the classification of 10 different bearing system failures.

V. CONCLUSIONS

This article presents a methodology for the diagnosis of bearing failures based on the Multiscale Permutation Entropy (MPE) technique. The MPE proves to be a highly effective characterization methodology to find information that allows to separate classes. Specifically, in the mechanical vibration signals that have a high non-stationary, the MPE manages to find characteristics that would not be detected by other methodologies. The MPE measures the non-linear dynamics existing in non-stationary time series and when combined with Reliess as a feature selection technique, a robust tool for classification applications is obtained. For the classification a method of K-Neighbors Nearest (KNN) was used, which manages to adapt to the nature of the characteristics. The results confirm a classification accuracy of more than 99.9% with a computation time of 16:37 seconds, which exceeds the results established in the literature.

ACKNOWLEDGMENT

The authors thank the Master in Electrical and Mechanical Engineering of the Technological University of Pereira for their support throughout the investigation. In addition, we would like to thank COLCIENCIAS for supporting the project entitled: “Desarrollo de un sistema de monitoreo para el análisis energético y de condición de emisiones en motores de combustión interna diésel con base en técnicas no destructivas” with code 1110-776-57801, through which the research described in this article was developed.

REFERENCES


