Mechanical Systems and Signal Processing 60-61 (2015) 229–242

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Rolling element bearings diagnostics using the Symbolic Aggregate approXimation

George Georgoulas [a,*], Petros Karvelis [a], Theodoros Loutas [b], Chrysostomos D. Stylios [a]

[a] *Laboratory of Knowledge and Intelligent Computing, Department of Computer Engineering, Technological Educational Institute of Epirus, Arta, Greece*

[b] *Applied Mechanics Lab, Department of Mechanical Engineering and Aeronautics, University of Patras, Rio, Greece*

ARTICLE INFO

Article history:
Received 16 April 2014
Received in revised form
26 July 2014
Accepted 28 January 2015
Available online 23 February 2015

Keywords:
Symbolic Aggregate approXimation
Bearing fault diagnosis
Condition monitoring
Classification

ABSTRACT

Rolling element bearings are a very critical component in various engineering assets. Therefore it is of paramount importance the detection of possible faults, especially at an early stage, that may lead to unexpected interruptions of the production or worse, to severe accidents. This research work introduces a novel, in the field of bearing fault detection, method for the extraction of diagnostic representations of vibration recordings using the Symbolic Aggregate approXimation (SAX) framework and the related intelligent icons representation. SAX essentially transforms the original real valued time-series into a discrete one, which is then represented by a simple histogram form summarizing the occurrence of the chosen symbols/words. Vibration signals from healthy bearings and bearings with three different fault locations and with three different severity levels, as well as loading conditions, are analyzed. Considering the diagnostic problem as a classification one, the analyzed vibration signals and the resulting feature vectors feed simple classifiers achieving remarkably high classification accuracies. Moreover a sliding window scheme combined with a simple majority voting filter further increases the reliability and robustness of the diagnostic method. The results encourage the potential use of the proposed methodology for the diagnosis of bearing faults.

1. Introduction

Rolling element bearings are a very critical component in various machines, and account for 45–55% of asynchronous motor failures according to a report from the Electric Power Research Institute (ERPI) [1]. The last two decades, the interest for efficient and robust diagnostics of roller element bearings via condition monitoring approaches as well as advanced signal processing has seen an extremely high increase. Of paramount importance in the diagnostic task are two items: 1) the feature extraction of damage sensitive features from the recorded condition monitoring signals and 2) the pattern recognition approach that leads to fault identification. Regarding the 1st item, wavelet analysis [2–4], empirical mode decomposition (EMD) [5,6] envelope analysis [7], cepstrum analysis [8], Kurtogram [9], nonlinear feature extraction [10], and other tools are utilized [11]. Regarding the 2nd item, artificial neural networks (ANNs), support vector machines (SVMs) [4], hidden Markov models (HMMs) [3], and particle filters [7] are used for pattern recognition and fault identification but their success highly depends upon the diagnostic potential of the extracted features.

More specifically, in [2] the condition of an electric motor with two rolling bearings with one normal state and three faulty states each was studied. De-noising via the continuous wavelet transform (CWT) was conducted and support vector

* Corresponding author.
E-mail address: georgoul@gmail.com (G. Georgoulas).

http://dx.doi.org/10.1016/j.ymssp.2015.01.033

machines (SVMs) were used for the fault classification task with 100% accuracy. Ocak et al. [3], developed a new scheme based on wavelet packet (WP) decomposition and HMMs for the condition monitoring of bearing faults. In this scheme, vibration signals were decomposed into WPs and the node energies of the 3-level decomposition tree were used as features. An HMM was trained to model the normal bearing operating condition and its probabilities were then used to track the condition of the bearing. The normalized WPs quantifiers as a new feature set was introduced in [4] for the detection and diagnosis of localized bearing defect and contamination fault. Unlike the conventional feature extraction methods, which use the amplitude of wavelet coefficients, these new features were derived from probability distributions and are more robust for diagnostic applications.

In [5] EMD was utilized for the extraction of only four features both from the time and the frequency domain of two specific intrinsic mode functions (IMFs) which were then fed to an ensemble anomaly detector consisted of very simple individual detectors capable of detecting four different types of seeded faults in a benchmark data set [12]. Envelope analysis using Hilbert transform was used in [7] and the energy ratio of specific components extracted from the frequency spectra of the envelop signal were combined with static (KNN) and dynamic (particle filters) anomaly detectors with very promising results for bearings suffering with different fault severities and under different operating conditions. The minimum variance cepstrum (MVC) method for the detection of the signature periodic fault signals was investigated in [8] and automotive ball bearings and the results compared favorably to other well established techniques such as the wavelet transform and the complex envelop analysis. Wang and his co-workers proposed an enhanced Kurtogram whose kurtosis values are calculated based on the power spectrum of the envelope of the signals extracted from WP nodes at different depths [9]. The nodes corresponding to the highest kurtosis can then be considered for further analysis. The proposed method combined with autoregressive (AR) filtering performed well both in simulated signals corrupted by noise as well as real experiments involving bearings with outer race, inner race and rolling element faults. Nonlinear feature extraction based on recurrence quantification analysis (RQA) for feature extraction was proposed in [10]. Two of the extracted features manifested a monotonic behavior in relation to fault severity in bearing faults making them a viable candidate for severity assessment. In [11] Wiener entropy along with the WPT nodal energies were used as inputs to a support vector regression machine for the prediction of the remaining useful life (RUL) of bearings in an accelerated test framework with the results being in good agreement to the actual RUL curve for all the tested cases.

It is apparent that a plethora of feature extraction methods have been employed in the field of condition monitoring. Numerous possibilities exist and the number of features that one can extract from a diagnostic signal can be very large. At the same time, a lot of work has been carried out in the direction of transforming continuous valued time series data to a discrete representation, through a process known as discretization or symbolization. The reason for that trend is that discrete representations offer a much higher numerical computation efficiency (through dimensionality reduction) leading to faster execution, which is crucial for online implementations as well as for computations involving huge amounts of data [13,14]. Additionally, symbolization provides the the opportunity to apply methods from the mature fields of bioinformatics and text mining, etc. [13]. Especially in the field of condition monitoring and anomaly detection of electrical, mechanical and electromechanical systems a similar approach has been pursued by professor Ray and his coworkers using primarily D-Markov machines [15–17].

Symbolic time series representation is another concept from the data mining field not often encountered for bearing fault monitoring. In two characteristic works, Boutros and Liang [18] submit the vibration signals to bandpass filtering and a monitoring index vector is calculated which is then assigned a symbol using a codebook generated using k-means clustering. The discrete observation sequence is then used by an HMM for fault localization and severity assessment with very high reported accuracy rates. Sanjith et al. [19] utilize a symbolic representation of vibration data and the created dictionaries are used to form two indices that can be used to discriminate between healthy bearings and bearings with various seeded faults.

In this work, a novel approach is utilized for the analysis of vibration signals, transforming them into a discrete valued sequence and an “intelligent icons”—like approach for feature extraction [20,21]. The creation of the discrete value sequence is based on the Symbolic Aggregate approXimation (SAX), which was introduced in [13] as a method for indexing. However since its introduction SAX was further exploited for anomaly detection and classification using time series bitmaps [22] or intelligent icons [20,21]. These representations can be used as features in a standard classification framework, rather than mere visualization tools. Along this path, SAX and the intelligent icons schemes are used for non-conventional feature extraction from vibration recordings. The results of this work suggest that this new representation is very efficient for the condition monitoring and fault identification of bearings providing thus an extra tool for the practitioners in the field.

The rest of the paper is structured as follows: Section 2 presents in brief the SAX approach and the intelligent icons algorithm, putting more emphasis on a modification of Piecewise Aggregate Approximation (PAA), which is part of the SAX framework. In Section 3 the involved data set and the conducted experiments are presented along with the achieved results. The paper is concluded with Section 4 with some discussion and some future directions.

2. Methodology

The proposed method for bearing fault detection/diagnosis consists of a series of steps that are depicted in Fig. 1. First, the signal acquired by an accelerometer is segmented to contain, as it will be explained later on, approximately 10 revolutions of data. Then each extracted segment undergoes SAX analysis i.e. z-score normalization to have zero mean and

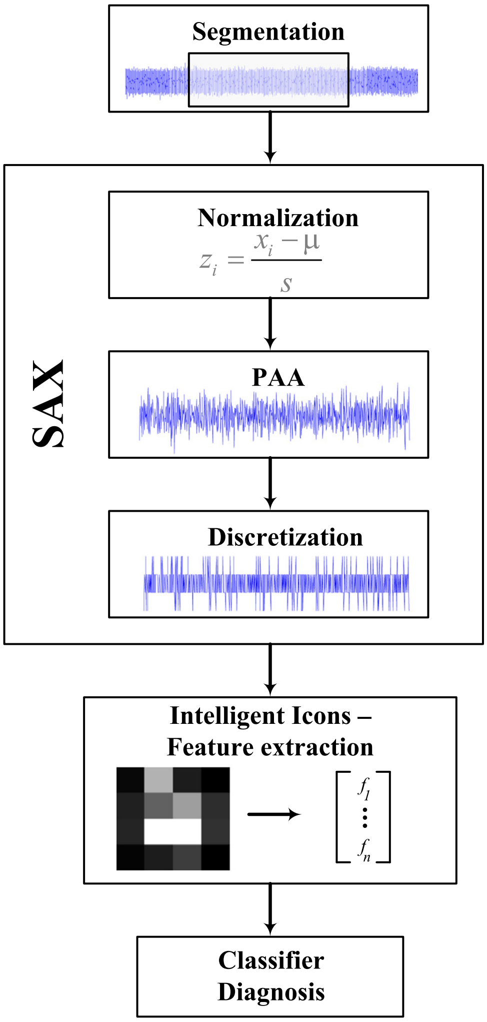

Fig. 1. A schematic of the proposed fault diagnosis method.

standard deviation equal to one, application of PAA and then discretization under the normality assumption, which creates a symbolic representation of the original signal. The symbolic representation is transformed into a feature vector through the application of the intelligent icons rationale and finally the feature vector is fed to a simple classifier that performs diagnosis. In the rest of this section each step will be presented in brief.

2.1. Feature extraction using the Symbolic Aggregate ApproXimation (SAX)

The main goal of the SAX algorithm is to convert a time series of arbitrary length N to a string of arbitrary length w, where $w < N$. The alphabet size is also an arbitrary integer $A > 2$. The SAX algorithm can be decomposed into two main steps. First, the signal is divided into equal sized sections. After that, the mean value of each section is calculated. By substituting each section with its mean, a reduced dimensionality is achieved. This process is known as Piecewise Aggregate Approximation (PAA) [23,24]. After the time-series has been transformed to its PAA representation, a discretization takes place in such a manner as to produce a word with approximately equiprobable symbols. As noted in the previous paragraph, prior to the computation of the PAA, the time-series is normalized in order to have zero mean and standard deviation of one.

2.1.1. Piecewise Aggregate Approximation (PAA)

PAA was independently introduced by Keogh et al. [23], and by Yi and Faloutsos [24] as a means to produce a low(er) dimensionality representation of a time series. PAA has already been used as a dimensionality reduction stage in a vibration based condition monitoring scheme [25].

According to the PAA framework, a time series $x=\{x[1],x[2],\ldots,x[N]\}$[1] of length N, can be represented in a w- dimensional space by a vector $x_{PAA}=[\overline{x}[1],\overline{x}[2],\ldots,\overline{x}[w]]$ where i-th element is given by the following equation:

$$\overline{x}[i]=\frac{w}{N}\sum_{j=N/w(i-1)+1}^{(N/w)i} x[j],\ for\ \ i=1,2,\ldots,w \tag{1}$$

The preceding equation states that in order to reduce the time series, or the discrete time signal, from an N dimension to a dimension w, one has to divide the series into equal sized windows taking the mean value of the data falling within a frame. This way a vector comprising of these values becomes the data-reduced representation. An example of PAA representation of a time series (vibration recording) is shown in Fig. 2.

The aforementioned representation applies for sequences whose length N can be divided exactly by w. In the following section a modification of the original PAA that can handle situations where N is not divided exactly by w is presented.

2.1.1.1. Modified Piecewise Aggregate Approximation. PAA was developed for discrete time signals. The process however is equivalent (or at least yields the same result) to calculating the height of a rectangle that occupies the same area with the area that is contained under the piecewise continues (time) signal $x_{ZOH}(t)$ that is created by applying a "zero-order hold" procedure to the original discrete time signal.

$$x_{ZOH}(t)=\begin{cases} x[1], & t\in[1,2) \\ x[2], & t\in[2,3) \\ \vdots & \vdots \\ x[N] & t\in[N,N+1) \end{cases} \tag{2}$$

The whole procedure is better explained through a simple example. At Fig. 3 on the left, the PAA procedure is depicted involving a discrete time signal with $N=12$ and $w=3$.

The discrete signal is $x=[1,\ 2,\ 4\ ,5,\ 2,\ 1,\ 1\ ,3,\ 4,\ 5,\ 4,\ 5]$ and the PAA representation x_{PAA} can be easily shown to be $x_{PAA}=[3,\ 1.75,\ 4.5]$ $(x_{PAA}[1]=(3/12)(1+2+4+5)=3$, etc) which is equal to:

$$\frac{\int_1^{(1+4)} x_{ZOH}(t)dt}{(4+1)-1}=\frac{1\times1+1\times2+1\times4+1\times5}{4}=x_{PAA}[1],$$

where the integral is over the limits of section A in the right part of Fig. 3 This procedure can be used also in the case of irregularly sampled signals (not constant sampling frequency) using this integral representation.

In general using the integral calculation Eq. (1) becomes

$$\hat{x}_{PAA}[i]=\frac{w}{N}\int_{(N/w)(i-1)+1}^{((N/w)i+1)} x_{ZOH}(t)dt\ \ \text{for}\ \ i=1,2,\ldots,w \tag{3}$$

where there is no need for (N) to be divided exactly by w.

Assuming sequences with a normalized sampling period ($x=\{x[nT]\}, n=1,2,\ldots,N$ and $T=1$), equation can be written in a summation form as

$$\begin{aligned}\hat{x}_{PAA}[i]=&\frac{w}{N}\left(\left(1-\left(\left(\frac{N}{w}(i-1)+1\right)-\left\lfloor\frac{N}{w}(i-1)+1\right\rfloor\right)\right)\cdot x\left[\left\lfloor\frac{N}{w}(i-1)+1\right\rfloor\right]\right)\\ &+\frac{w}{N}\left(\sum_{j=\lfloor(N/w)(i-1)+1\rfloor+1}^{\lfloor(N/w)i+1\rfloor-1} x[j]+\left(\left(\left(\frac{N}{w}i+1\right)-\left\lfloor\frac{N}{w}i+1\right\rfloor\right)\cdot x\left[\left\lfloor\frac{N}{w}(i)+1\right\rfloor\right]\right)\right)\ \text{for}\ i=1,2,\ldots,w\end{aligned} \tag{4}$$

This procedure is also visually explained at Fig. 4, where the first and third term corresponds to the left side (A) and right side (B) is trimmed rectangles respectively. The procedure for regularly sampled sequences is equivalent to the one originally proposed by Dr Wei and Dr Keogh as part of the generalized SAX [26]

2.1.2. Discretization

After the PAA computation of the time series, a discretized representation is produced by defining a partitioning of the original continuous space. There are many approaches for selecting the partitioning [14]. The goal of the discretization procedure under the SAX framework is to produce symbols with equiprobability [13], which is related to the Maximum Entropy based Partitioning (MEP) [27]. However, having normalized the time series, the produced new time series will approximately follow a Gaussian distribution and so it is simply to determine the "breakpoints" that will produce equal-sized areas under a Gaussian curve [13].

In order to produce the symbols from the PAA coefficients, we employ the following procedure:

[1] The notation is slightly different from the one encountered in the time series literature, following more the signal processing rational.

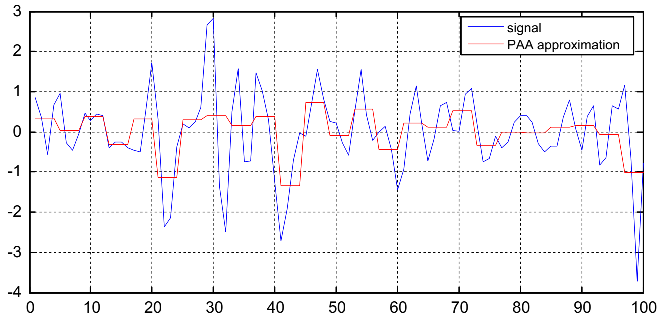

Fig. 2. A fraction of a vibration signal ($N=100$) and its PAA approximation ($w=25$).

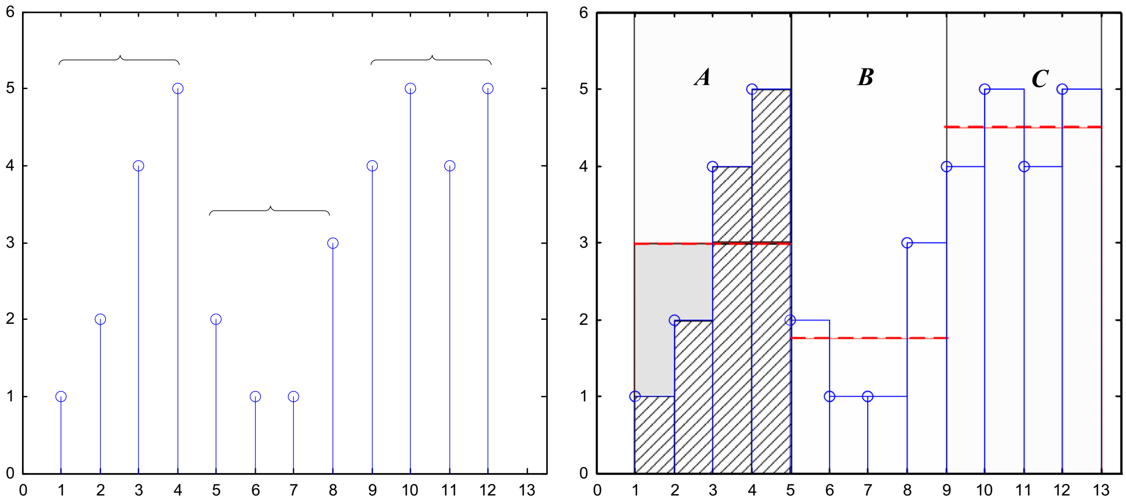

Fig. 3. An example of the "equivalence" between the discrete PAA calculation and the modified PAA through a geometric interpretation.

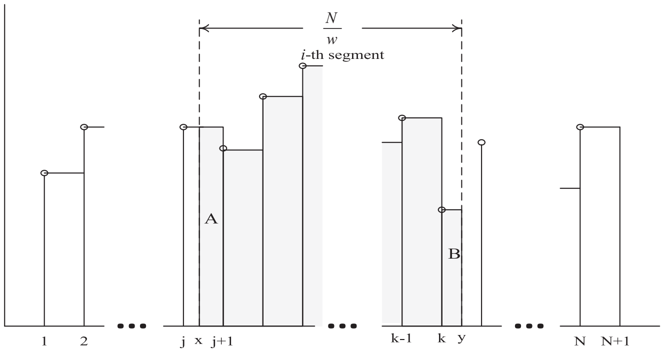

Fig. 4. Schematic explanation of the estimation of the i-th PAA value for an arbitrary segment. The following points are marked on the x axis: $x=((N/w)(i-1)+1)$, $j=\lfloor (N/w)(i-1)+1 \rfloor$, $y=((N/w)\,i+1)$and $k=\lfloor (N/w)\,i+1 \rfloor$.

- all PAA coefficients that are below the smallest breakpoint are mapped to the symbol "a",
- all coefficients greater than or equal to the smallest breakpoint and less than the second smallest breakpoint are mapped to the symbol "b", and the procedure continues up to the highest PAA coefficients value that is mapped to the last symbol of the chosen alphabet.

This procedure is visually displayed in Fig. 5.

2.1.3. Intelligent icons approach for feature extraction

SAX representation is primarily suitable for indexing and it was not originally meant for classification even though in [28], SAX representation is used directly for human action recognition. However the most common approach for the use of

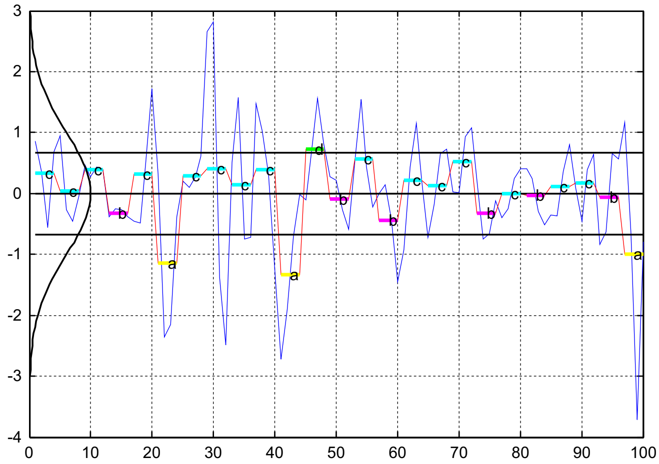

Fig. 5. The discretization of the time series depicted in Fig. 4 using an alphabet of size 4, $A = 4$.

a

a:3	b:6
c:15	d:1

b

aa:0	ab:0	ba:1	bb:0
ac:1	ad:1	bc:5	bd:0
ca:2	cb:4	da:0	db:1
cc:8	cd:0	dc:0	dd:0

Fig. 6. The icons of the string $S = cccbcaccccadbcbcccbcbccba$ (a) the frequencies and the subwords for $L = 1$ and (b) the same for $L = 2$.

SAX for classification involves the creation of bitmaps or intelligent icons, which are used to display time series in a more compact form or more importantly to be used as feature vectors for classification/detection purposes [20,29].

Generally, in the digitalized world an icon stand for a file. It is a small image usually of size 32 × 32 and in the case that the file is an image then the icon is a sub-sample of the initial image. Thus, the user simply and quickly gets an idea of what image the file contains. An intelligent icon is a smart way of representing a long SAX string. Having acquired the SAX string one can construct a $(A/2) \times (A/2)$ (where $\mathrm{mod}(A, 2) = 0$) array simply by counting the frequencies of each symbol. We can also build an $A \times A$ array by simply counting the frequencies of subwords of the SAX string of size 2 (e.g. aa, ab, ac, etc.). In order to generalize this procedure we count the frequencies of specific subwords of length L. Firstly, we assign to each letter of the alphabet a unique value:

$$a = 0, b = 1, c = 2, d = 3, \ldots \tag{5}$$

Each word has an index for the location of each symbol in the table of the icon. For clarity, we may show them explicitly as subscripts. Then, in order to map a subword to the icon, we use the following equations to calculate its row and column values:

$$row = \sum_{n=0}^{L-1} (k_n div 2) 2^{L-n-1} \tag{6}$$

$$col = \sum_{n=0}^{L-1} (k_n 2^{L-n-1}) \mathrm{mod}\ 2^{L-n}. \tag{7}$$

As an example, the icon of the string $S = cccbcaccccadbcbcccbcbccba$ is displayed in Fig. 6.

In this work the intelligent icon representation (i.e. the frequency counts of subwords), is employed for the extraction of features rather than for the optical representation of the signal. The reason is that we are interested in an automated procedure rather than an alternative visual representation, which would require the familiarization of the end user with this specific representation. However as it can be seen in Fig. 7, there are distinct visual differences between the different classes.

But, for some combinations of word length and alphabet size there is no convenient way to create an intelligent icon representation. However, one can still count the frequency of occurrences of the different words and use them as feature vectors. This is exactly the proposed approach here: for each signal after the application of all the aforementioned steps the

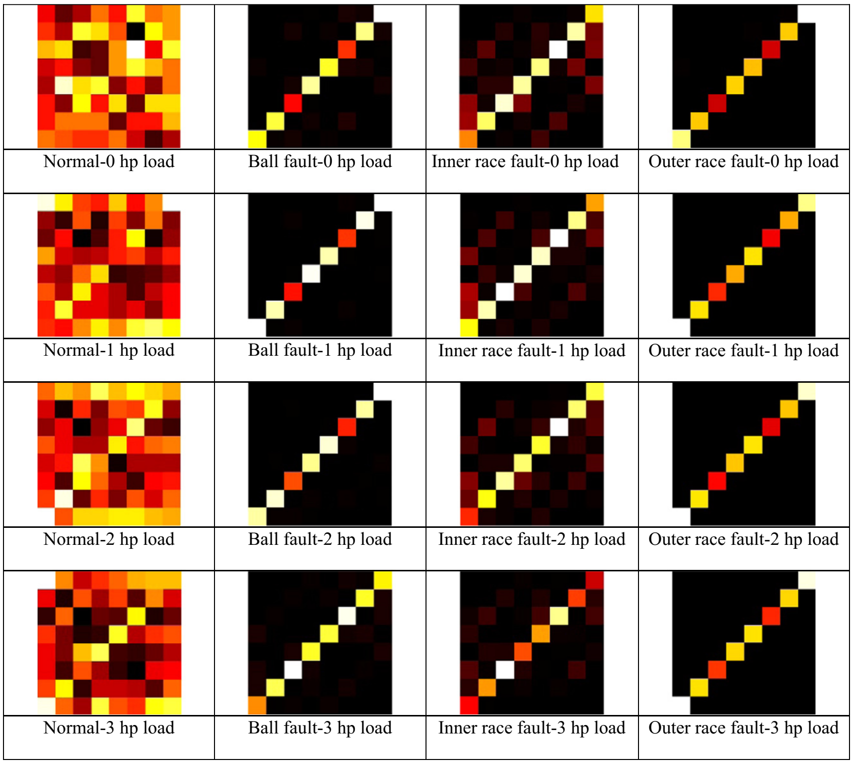

Fig. 7. The icons for different loading conditions and for different fault locations for seeded faults of size equal to 0.007 in. Notice the similarity between the icons belonging to the same column as compared to the icons belonging to other columns (conditions).

frequency of occurrences of each possible word are stored in a one dimensional vector, which consists the representation of the original signal into the feature space.

At this point, it is worth noting the similarity with the D-Markov modeling approach employed in the work of professor Ray and his co-workers [15–17]. In their approach the transition matrix Π which is created by counting transitions between states (where each state is just a word of length D) is also experimentally estimated by counting the occurrences of transitions and is then used as a feature vector for the diagnosis task. Even though different normalization takes place compared to the one adopted in the intelligent icons approach, the two approaches are very similar.

2.2. Classification and diagnosis

For the classification and diagnosis task two simple—though quite different—classifiers were employed: a) the nearest neighbor (NN) classifier, which is a nonlinear classifier and b) the minimum Mahalanobis distance classifier, which is a linear discriminant classifier (LDC). We choose to use two relatively simple classifiers, because in this work we basically focus on the use of the SAX based features for the diagnosis of bearing faults and also because as it has been reported in [30], sometimes in real life applications the performance of simple classification schemes is comparable to more complex and state of the art classification paradigms. Moreover these two classifiers have the very appealing property that they are parameter-free, leaving the tuning process entirely to the selection of the parameters involved during the application of SAX.

The nearest neighbor classifier as its name implies, given a labeled training set, assigns a new unseen example to the class of its nearest neighbor, where the “closeness” is assessed through an appropriately selected distance.

In the case of the minimum Mahalanobis distance classifier, each feature vector $\mathbf{x}$ is assigned to class i (normal, ball fault etc.) such that the value of the corresponding discriminant function is maximized:

$$i^* = \arg\max_i \left\{ 2\ln P(\omega_i) - (\mathbf{x}-\boldsymbol{\mu}_i)^T \mathbf{C}^{-1} (\mathbf{x}-\boldsymbol{\mu}_i) \right\}, \tag{8}$$

where $\boldsymbol{\mu}_i$is the mean of class i, $P(\omega_i)$ is the prior probability of class i, and $\mathbf{C}$ is the estimated covariance matrix, which is assumed constant across all classes (despite this assumption may not hold in many practical applications, this classification approach performs surprisingly well).

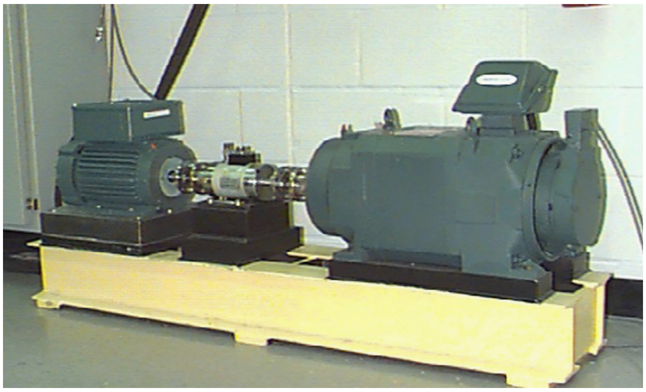

Fig. 8. The experimental test bench (courtesy of Professor K. Loparo).

3. Results

3.1. Vibration recordings dataset

The data used in this research work come from a bearing installed in a motor driven mechanical system [12] at the drive end of the motor. The test bearings were deep groove ball bearings (SKF 6205 JEM) with single point faults. Three types of faults (outer race, inner race and ball faults) were introduced using electro-discharge machining (EDM) with three fault diameters: 0.007, 0.014 and 0.021 in. For the case of the outer race faults, experiments were conducted with faults located at 3 o'clock (directly in the load zone), at 6 o'clock (orthogonal to the load zone), and at 12 o'clock. However for this study the three sub-categories for the outer race faults were merged forming a broader class of outer race faults. Each bearing was tested under four different loads of 0, 1, 2, and 3 hp. For each test, vibration data were collected using accelerometers placed at the 12 o'clock position at the drive end at a sampling frequency of 12,000 samples/s. A more detailed description of the experimental set-up and the apparatus involved can be found in Case Western Reserve University's website [12], while Fig. 8 depicts the experimental test bench.

3.2. Experimental evaluation—diagnostic performance

In order to implement and evaluate the proposed SAX-based approach, a number of experiments conducted following mainly the scheme proposed in [31] with some minor modifications. More specifically, we addressed four different classification problems in batch mode (putting together all the extracted cases without considering their time dependencies —discarding in other words the time component). For the batch classification/diagnosis problems, during the segmentation phase, we extracted non overlapping time windows containing 10 revolutions. Table 1 summarizes the four classification problems with the respective dataset samples involved. Moreover we also conducted a pseudo-online version of the fourth classification problem (the cases were presented in time of occurrence keeping their chronological order) to show the way a postprocessing step that can cope with the spurious misclassifications that were observed during the first (batch) phase of the experiments.

In the first classification problem A (following the convention used in [31]) only incipient faults (faults with the smaller induced defect) were considered under all four load conditions. Therefore all data corresponding to normal conditions and the data corresponding to a fault defect size of 0.007 in. irrespective of the location were employed. A four class classification problem was formed (normal, ball fault, inner race fault and outer race fault). Unlike the approach in [31], where a hold-out procedure was used, in our case a 10-fold (stratified) cross validation procedure was repeated 10 times (10×10 CV) [32] in order to assess the performance of our approach. This resampling technique minimizes the chances of getting good (bad) results due to a lucky (unlucky) splitting of the available data [32], and also provides a means to assess the stability of the model. Moreover, for the tuning of the algorithm, i.e. for selecting the "optimal" SAX parameters (number of windows and alphabet size as well as the word length) a nested procedure involving a grid search was performed using only the training data at each fold. This way, the estimation of the performance was decoupled from the tuning process. The nested procedure is schematically described in Fig. 9.

In the second (B) classification problem, all faulty data corresponding to a fault defect size of 0.007 in. irrespective of the location and the applied load were used for training and all the faulty data corresponding to a fault defect size of 0.021 in. irrespective of the location and the applied load were used for testing. The normal data were used only for training purpose, during which again a nested procedure was followed for the selection of the parameters involved.

In the third classification problem C, the same procedure as in B was followed with the exception that the "severe" fault data along with the normal cases were employed for training and the "incipient" fault data for testing.

Table 1
Description of the four (batch) classification problems.

Classification setting	Number of samples training/testing	Defect size training/testing (inches)	Condition	Labeling convention
A	413/413	0/0	Normal	1A
	116/116	0.007/0.007	Ball	2A
	117/117	0.007/0.007	Inner race fault	3A
	348/348	0.007/0.007	Outer race fault	4A
B	413/–	0/–	Normal	1B
	116/116	0.007/0.021	Ball	2B
	117/116	0.007/0.021	Inner race fault	3B
	348/349	0.007/0.021	Outer race fault	4B
C	413/–	0/–	Normal	1C
	116/116	0.021/0.007	Ball	2C
	116/117	0.021/0.007	Inner race fault	3C
	349/348	0.021/0.007	Outer race fault	4C
D	413/413	0/0	Normal	1D
	116/116	0.007/0.007	Ball	2D
	116/116	0.014/0.014	Ball	3D
	116/116	0.021/0.021	Ball	4D
	117/117	0.007/0.007	Inner race fault	5D
	116/116	0.014/0.014	Inner race fault	6D
	116/116	0.021/0.021	Inner race fault	7D
	348/348	0.007/0.007	Outer race fault	8D
	116/116	0.014/0.014	Outer race fault	9D
	349/349	0.021/0.021	Outer race fault	10D

The fourth classification problem (D) employs 10 classes: normal, ball fault with defect size equal to 0.007 in., ball fault with defect size equal to 0.014 in. etc. The same procedure as in case A was employed for the estimation of the performance.

For comparison reasons, to have a baseline reference performance, we also performed the same set of experiments, using the aforementioned two classifiers, having as inputs a set comprised of the following five features that are commonly encountered in bearing health assessment:

- the standard deviation of the signal,
- the kurtosis of the signal,
- the crest factor of the signal,
- the zero crossing rate (ZCR) of the signal,
- and a parameter free measure of the signal's complexity [33]: $com = \sqrt{\sum_{i=1}^{N-1} (x[i] - x[i+1])^2}$.

The results of the proposed scheme are summarized in the Tables 2–6 in the form of confusion matrices for both classifiers. For problems A and D (Tables 2, 5 and 6) the aggregated matrices over the different repetitions are reported.

In Table 7 we summarize the overall classification accuracy of the proposed approach for the different diagnostic problems and the two different classifiers along with the results presented in [31] as well as the results of the "baseline" feature set. It must be noted that no strict (quantitative) comparison is applicable for the results of the present study and the results reported in [31] since the procedures followed are a bit different. Nevertheless it can be seen that the proposed approach is comparable to the results obtained in [31] where a much broader range of features were involved along with a more powerful inductive algorithm.

The diagnostic scheme that is based on SAX extracted features performs equally well as the diagnostic scheme that uses the "baseline" feature set for the classification tasks A when the NN classifier is employed. For the tasks B, C and D the SAX based scheme performs better proving that the novel proposed scheme can be a viable candidate of a condition monitoring system provided that representative data are available.

The success of the proposed approach can be intuitively assessed by looking into a lower embedding of the original feature space into three dimensions through multidimensional scaling (MDS), Fig. 10. As one can observe even in this low dimension embedding the feature vectors of most of the different classes create compact clusters. Especially the normal class seems well separated from the rest of the classes. Therefore with the proposed feature extraction method, the detection of an anomalous situation is almost certain to be detected. For the rest of the classes there is some overlapping between some of the classes due to some spurious data points near the borders of the classes. Also the scattering of the data provides an insight for the superiority of the NN over the LDC for this particular setting. The compact clusters with the irregular shapes fit better the NN classifier which only exploits local properties rather than the LDC who assumes Gaussian distributions.

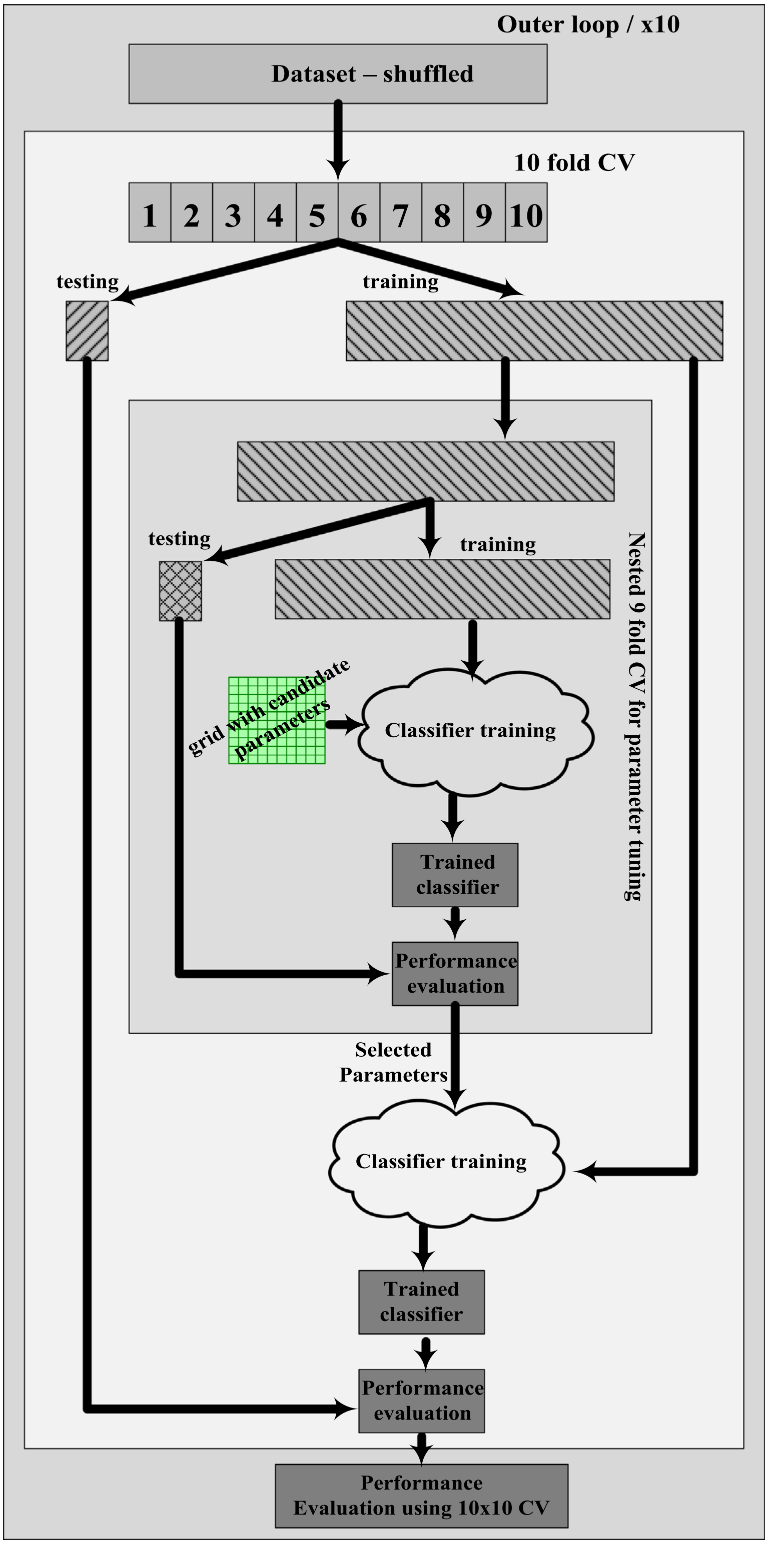

Fig. 9. Schematic representation of the nested procedure for the selection of parameters and the performance estimation.

To further validate the proposed approach a pseudo-online experiment was conducted. For the pseudo-online case no resampling procedure was employed. Instead the 2/3 of each signal was used for training and the rest 1/3 of all signals were concatenated. Since the NN algorithm seems to outperform the LDC method, for this experiment only a NN classifier was tested with the SAX generated features. As it can be seen in Fig. 11(b), very few samples are misclassified. This can be treated, in most cases using a hierarchical approach that can exploit the temporal aspect of the phenomenon. Such a scheme is shown in Fig. 12 where the decision of five consecutive windows is aggregated via a majority voting procedure. The result of such a procedure is depicted in Fig. 11(c). As it can be seen the spurious misclassifications are removed. The longer the “filter” (the more individual detectors), the less probable is to have misclassifications. On the other hand with more

Table 2
Aggregated confusion matrix for the classification problem A for the two classifiers.

LDC/NN		Estimated class			
		1A	2A	3A	4A
True class	1A	**4130/4130**	0/0	0/0	0/0
	2A	0/0	**1160/1160**	0/0	0/0
	3A	0/0	0/0	**1170/1170**	0/0
	4A	0/0	0/0	0/0	**3480/3480**

Table 3
Confusion matrix for the classification problem B for the two classifiers.

LDC/NN		Estimated class			
		1B	2B	3B	4B
True class	1B	–	–	–	–
	2B	**2/0**	**99/95**	**4/2**	**11/19**
	3B	0/0	0/0	**65/102**	**51/14**
	4B	0/0	**42/0**	**17/6**	**290/343**

Table 4
Confusion matrix for the classification problem C for the two classifiers.

LDC/NN		Estimated class			
		1C	2C	3C	4C
True class	1C	–	–	–	–
	2C	0/0	**77/116**	**39/0**	0/0
	3C	0/0	**7/33**	**110/83**	**0/1**
	4C	0/0	**0/125**	**83/20**	**265/203**

Table 5
Aggregated confusion matrix for the classification problem D for the nearest neighbor classifier (the gray blocks correspond to faults occurring at the same location but for different severities).

NN		Estimated class									
		1D	2D	3D	4D	5D	6D	7D	8D	9D	10D
True class	1D	**4130**	0	0	0	0	0	0	0	0	0
	2D	0	**1149**	0	**11**	0	0	0	0	0	0
	3D	0	0	**1113**	**20**	**7**	**20**	0	0	0	0
	4D	0	**31**	0	**1114**	0	0	0	**15**	0	0
	5D	0	0	0	0	**1170**	0	0	0	0	0
	6D	0	0	0	0	0	**1160**	0	0	0	0
	7D	0	0	0	0	0	0	**1160**	0	0	0
	8D	0	0	0	**6**	0	0	0	**3474**	0	0
	9D	0	0	0	0	0	0	0	0	**1160**	0
	10D	0	0	0	0	0	0	0	0	0	**3490**

Table 6
Aggregated confusion matrix for the classification problem D for the LDC (the gray blocks correspond to faults occurring at the same location but for different severities).

LDC		Estimated class									
		1D	2D	3D	4D	5D	6D	7D	8D	9D	10D
True class	1D	**4130**	0	0	0	0	0	0	0	0	0
	2D	0	**1146**	0	**14**	0	0	0	0	0	0
	3D	0	0	**1074**	**20**	0	**56**	0	0	0	**10**
	4D	0	**24**	0	**1056**	0	**7**	**4**	**69**	0	0
	5D	0	0	0	0	**1170**	0	0	0	0	0
	6D	0	0	0	0	0	**1160**	0	0	0	0
	7D	0	0	0	0	0	0	**1160**	0	0	0
	8D	0	0	0	0	0	0	0	**3480**	0	0
	9D	0	0	0	0	0	0	0	0	**1160**	0
	10D	0	0	0	0	0	0	0	0	0	**3490**

Table 7
Overall accuracy of the different classification problems for the two classifiers along with the results reported in [31]. For cases A and D also the standard error is reported in parenthesis.

	NN+SAX based features	LDC+SAX based features	ANFIS+6 feature sets from [30]	NN+“baseline” features	LDC+“baseline” features
A	**100 (0)**	**100 (0)**	**100**	**100 (0)**	99.84 (0.01)
B	**92.94**	78.14	92.5	64.54	32.36
C	69.19	77.80	**90.83**	29.78	39.76
D	**99.43 (0.12)**	98.94 (0.21)	91.53	98.58 (0.03)	87.99 (0.06)

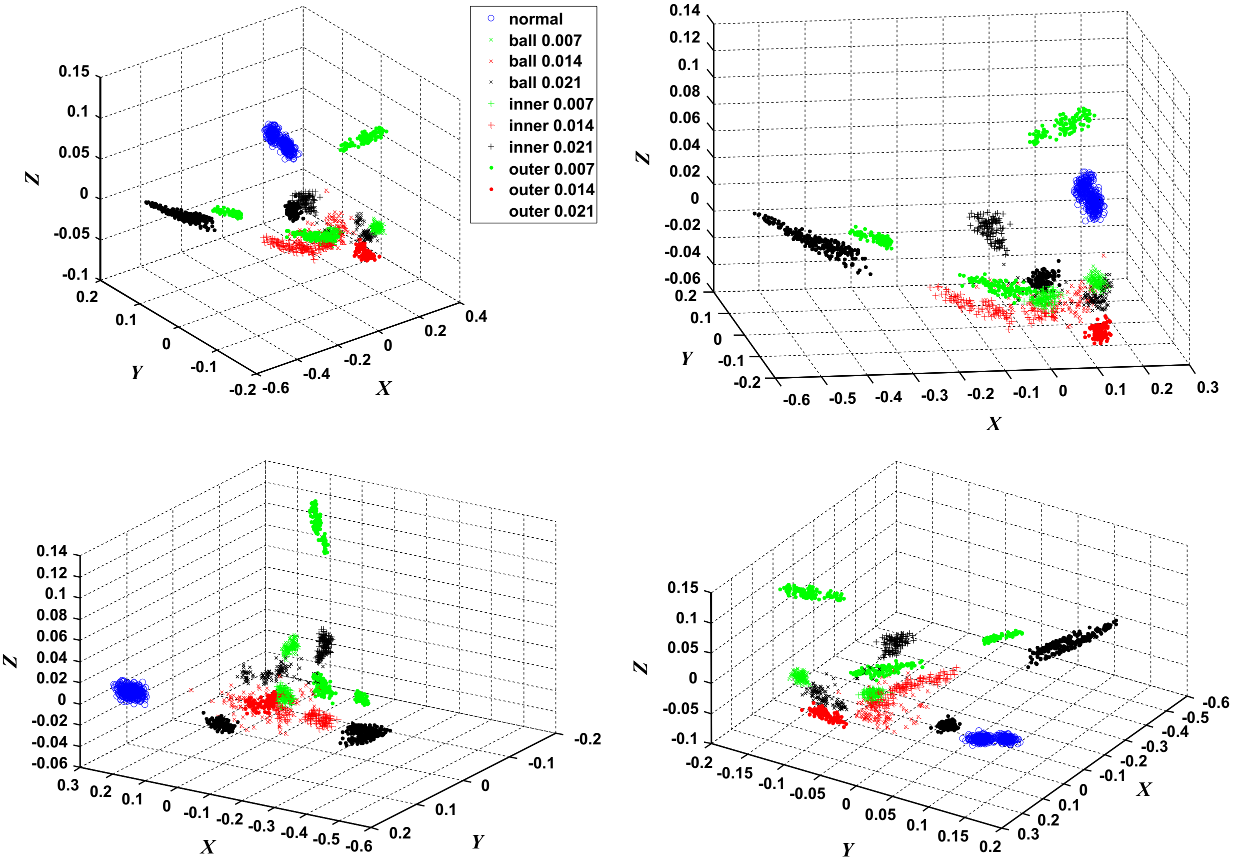

Fig. 10. Four different views of the result of MDS on the 10 class data after their projection into a three dimensional space.

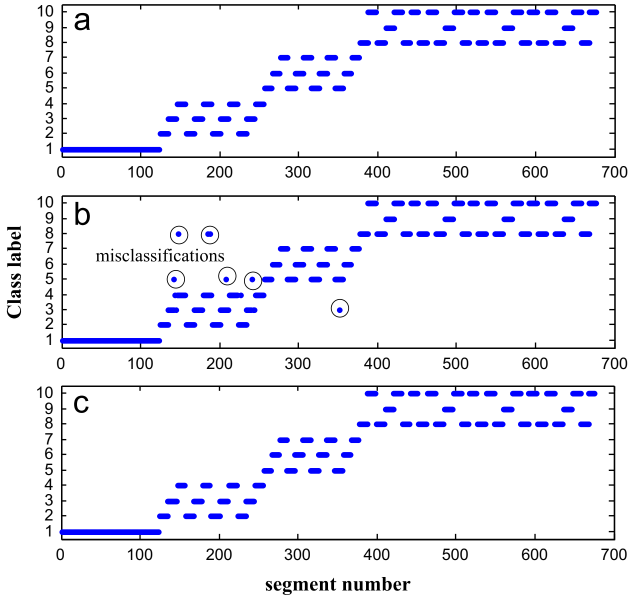

Fig. 11. The results of the application of the pseudoonline experiment: a) actual labels, b) the output of an individual NN detector and c) the results of the diagnosis scheme after the hierarchical majority voting approach.

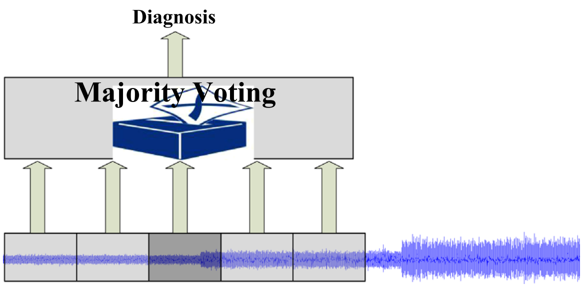

Fig. 12. The hierarchical diagnosis scheme.

detectors, the response to a change may take longer to appear. However since such changes do not appear very fast a delay in the response of the detector is not of such importance.

4. Discussion and conclusions

This research work proposes the use of a symbolic representation of vibration signals for the diagnosis of bearing faults. More specifically SAX is employed for the transformation of a real valued vibration signal into a sequence of symbols. The sequence of symbols is condensed into a much lower representation based on the frequency of appearances of all potential words given a pre-specified word-length. This representation comprises the feature vector that is used for representing the original signal and subsequently for classification using standard pattern recognition approaches. In fact the representation is so effective that even simple classifiers can achieve remarkable classification accuracies. The proposed scheme seems competitive regarding both more involved classification schemes [31] as well as diagnostic schemes that employ well established features for bearing condition assessment for this specific seeded fault setting. Further validation is needed using also naturally degraded equipment and different operational settings for establishing the connection between different faults and their “icon-like” representation in order for this approach to become a standard tool for bearing condition monitoring.

To deal with spurious misclassifications, a hierarchical approach exploiting the temporal association of the different conditions was proposed which further enhances the performance accuracy of the diagnosis scheme. Furthermore, as part of the developed technique, a modified/extended PAA representation was presented, which can be used as a general dimensionality reduction tool.

On the other hand the proposed method has the intrinsic barrier as all data driven techniques; it requires the existence of representative data in order to build the diagnosis scheme. Nevertheless our study revealed that in case of sufficient data the proposed representation can provide a competitive alternative to feature extraction, especially for online applications. In future work we will further exploit the use of symbolic representations for faults other than seeded ones as well as for the case of multiple faults which have not been investigated in this study.

Acknowledgments

The authors would like to thank Professor Kenneth Loparo for granting permission to use Fig. 8 and also for its general contribution to the field of bearing fault detection by making available the seeded fault bearing data set.

References

[1] C. Lanham, Understanding the Tests That are Recommended for Electric Motor Predictive Maintenance, Baker Instrument Company, Fort Collins, CO, 2006.
[2] S. Abbasion, A. Rafsanjani, A. Farshidianfar, N. Irani, Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine, Mech. Syst. Signal Process. 21 (7) (2007) 2933–2945.
[3] H. Ocak, K.A. Loparo, F.M. Discenzo, Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics, J. Sound Vib. 302 (4–5) (2007) 951–961.
[4] Y. Feng, F. Schlindwein, Normalized wavelet packets quantifiers for condition monitoring, Mech. Syst. Signal Process. 23 (3) (2009) 712–723.
[5] G. Georgoulas, T. Loutas, C.D. Stylios, V. Kostopoulos, Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition, Mech. Syst. Signal Process. 41 (1) (2013) 510–525.
[6] Y. Lei, J Lin, Z. He, M.J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mech. Syst. Signal Process. 35 (1) (2013) 108–126.
[7] B. Zhang, G. Georgoulas, M. Orchard, A. Saxena, D. Brown, G. Vachtsevanos, S. Liang, Rolling element bearing feature extraction and anomaly detection based on vibration monitoring, in: Proceedings of the 16th Mediterranean Conference on Control and Automation, Greece, June 2008, pp. 1792–1797.
[8] C.S. Park, Y.C. Choi, Y.H. Kim, Early fault detection in automotive ball bearings using the minimum variance cepstrum, Mech. Syst. Signal Process. 38 (2) (2013) 534–548.
[9] D. Wang, P.W. Tse, K.L. Tsui, An enhanced kurtogram method for fault diagnosis of rolling element bearings, Mech. Syst. Signal Process. 35 (1) (2013) 176–199.
[10] Y. Qian, R. Yan, M. Shan, Damage assessment of mechanical systems based on recurrence quantification analysis, in: Proceedings of the IEEE Conference on Prognostics and System Health Management (PHM), 2012, pp. 1–5.
[11] T.H. Loutas, D. Roulias, G. Georgoulas, Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression, IEEE Trans. Reliab. 62 (4) (2013) 821–832.
[12] K.A. Loparo, Bearing Vibration Dataset, Case Western Reserve University", 2014 ⟨http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website⟩ (last accessed 13.01.14).
[13] J. Lin, E. Keogh, S. Lonardi, B. Chiu, A symbolic representation of time series, with implications for streaming algorithms, in: Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego, CA, 2003.
[14] C.S. Daw, C.E.A. Finney, E.R. Tracy, A review of symbolic analysis of experimental data, Rev. Sci. Instrum. 74 (2) (2003) 915–930.
[15] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Signal Process. 84 (7) (2004) 1115–1130.
[16] V. Rajagopalan, A. Ray, Symbolic time series analysis via wavelet-based partitioning, Signal Process. 86 (11) (2006) 3309–3320.
[17] S. Sonti, E. Keller, J. Horn, A. Ray, Stability monitoring of rotorcraft systems: a dynamic data-driven approach, J. Dyn. Syst. Meas. Control 136 (2) (2014) 024505.1–024505.6.
[18] T. Boutros, M. Liang, Detection and diagnosis of bearing and cutting tool faults using hidden Markov models, Mech. Syst. Signal Process. 25 (6) (2011) 2102–2124.
[19] B.M.M.A. Sanjith, C. Sujatha, T. Jayakumar, Symbolic dynamics based bearing fault detection, in: Proccedings of the 5th India International Conference on Power Electronics (IICPE), December 2012, pp. 1–5.
[20] E. Keogh, L. Wei, X. Xi, S. Lonardi, J. Shieh, S. Sirowy, Intelligent icons: integrating lite-weight data mining and visualization into gui operating systems. in: Proceedings of the ICDM, 2006.
[21] J. Lin, E. Keogh, L. Wei, S. Lonardi, Experiencing SAX: a novel symbolic representation of time series, Data Min. Knowl. Discov. 15 (2) (2007) 107–144.
[22] N. Kumar, N. Lolla, E. Keogh, S. Lonardi, C.A. Ratanamahatana, Time-series bitmaps: a practical visualization tool for working with large time series databases, in: Proceedings of the SDM, 2005.
[23] E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra, Dimensionality reduction for fast similarity search in large time series databases, J. Knowl. Inf. Syst. 3 (3) (2000) 263–286.
[24] K. Yi, C. Faloutsos, Fast time sequence indexing for arbitrary Lp norms, in: Proceedings of the 26th International Conference on Very Large Databases, Cairo, Egypt, 2000.
[25] W. Caesarendra, B. Kosasih, A.K. Tieu, C.A. Moodie, Circular domain features based condition monitoring for low speed slewing bearing, Mech. Syst. Signal Process. 45 (1) (2014) 114–138.
[26] SAX Homepage. ⟨http://www.cs.ucr.edu/~eamonn/SAX.htm⟩ (last accessed 15.06.2014).
[27] V. Rajagopalan, A. Ray, R. Samsi, J. Mayer, Pattern identification in dynamical systems via symbolic time series analysis, Pattern Recognit. 40 (11) (2007) 2897–2907.
[28] I.N. Junejo, Z.A. Aghbari, Using SAX representation for human action recognition, J. Vis. Commun. Image Represent. 23 (6) (2012) 853–861.
[29] P. Karvelis, G. Georgoulas, I. Tsoumas, C.D. Stylios, J. Antonino-Daviu, V. Climente-Alarcon, An intelligent icon approach for rotor bar fault detection, to be presented, in: Proceedings of the Annual Conference of the IEEE Industrial Electronics Society (IECON), Vienna-Austria, 10–13 November 2013.
[30] D.J. Hand, Classifier technology and the illusion of progress (with discussion), Stat. Sci. 21 (2006) 1–34.
[31] Y. Lei, Z. He, Y. Zi, Q. Hu, Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs, Mech. Syst. Signal Process. 21 (5) (2007) 2280–2294.
[32] N. Japkowicz, M. Shah, Evaluating Learning Algorithms: A Classification Perspective, Cambridge University Press, New York, NY, 2011.
[33] G.E. Batista, X. Wang, E.J. Keogh, A complexity-invariant distance measure for time series, in: Proceedings of the SDM, 2011.