This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2015.2463680, IEEE Transactions on Industrial Informatics

A Symbolic Representation Approach for the Diagnosis of Broken Rotor Bars in Induction Motors

Petros Karvelis, George Georgoulas, Ioannis P. Tsoumas, *Member IEEE*, Jose Alfonso Antonino-Daviu, *Senior Member IEEE*, Vicente Climente-Alarcón, *Member IEEE,* and Chrysostomos D. Stylios, *Senior Member IEEE*

***Abstract*— One of the most common deficiencies of currently existing induction motor fault diagnosis techniques is their lack of automatization. Many of them rely on the qualitative interpretation of the results, a fact that requires a significant user expertness and that makes their implementation in portable condition monitoring devices difficult. In this paper, we present an automated method for the detection of the number of broken bars of an induction motor. The method is based on the transient analysis of the start-up current using wavelet approximation signal that isolates a characteristic component that emerges once a rotor bar is broken. After the isolation of this component, a number of stages are applied that transform the continuous-valued signal into a discrete one. Subsequently an intelligent icons like approach is applied for condensing the relative information into a representation that can be easily manipulated by a nearest neighbor classifier. The approach is tested using simulation as well as experimental data achieving high classification accuracy.**

***Index Terms*— Discrete Wavelet Transform, Rotor Faults, Intelligent Icons, Symbolic Representation, Piecewise Aggregate Approximation.**

NOMENCLATURE

f_{LSH}	Frequency of the Left Sideband Harmonic {Hz}
s	Slip {-}
f_s	Power supply frequency {Hz}
f_{sampl}	Sampling frequency {Hz}
$x[\cdot]$	Discrete time signal {-}
$x(\cdot)$	Continuous time signal {-}
$a_n(t)$	Approximation signal at level n {-}
$d_j(t)$	Detail signal at level j {-}
a_n^i	Approximation coefficients at level n {-}
d_j^i	Detail coefficients at level n {-}
φ^n	Scaling function at level n {-}
ψ^j	Wavelet function at level j {-}
x_{PAA}	PAA representation of a signal/time series {-}
$\bar{x}[n]$	n th element of the PAA representation {-}
Σ	Alphabet
$\lvert\Sigma\rvert$	Cardinality of the Alphabet

I. INTRODUCTION

INDUCTION motors from a fraction of kW up to several MWs constitute the driving force of the industry [1]-[3]. Although they have lower efficiency and more volume and weight than permanent magnet synchronous motors, their low cost, simple construction and robustness have led to their prevalence in industry applications, where the latter issues play a more crucial role. Furthermore, in such applications, the efficiency of the complete drive system including the working machine is more important than the efficiency of the motor; the improvement in the latter can only be marginal.

Regarding the low voltage drives (400V/690V) with power ratings from several hundreds of kWs up to 1MW, ca. 2/3 of the applications have to do with pumps, fans and compressors [1]. The motors of the drive system in these applications are squirrel-cage induction motors with aluminum bars and rings for the rotor cage. No matter how robust the squirrel-cage motors are, they can still suffer from electrical and mechanical faults that can lead to failures and production shutdowns [4], [5]. Therefore, it is of paramount importance to detect a fault as precisely as possible and as early as possible; this will lead to a faster repair and a shorter downtime.

Motor Current Signature Analysis (MCSA) is currently the most prominent approach for rotor fault diagnosis in asynchronous machines, [6]-[9], as well as to other electric equipment [10], primarily due to its non-invasive nature [11], [12].

Manuscript received July 22, 2014, revised March 1, and June 10, 2015.

P. Karvelis, G. Georgoulas and C. D. Stylios are with the Laboratory of Knowledge and Intelligent Computing, Dept. of Computer Engineering, TEI of Epirus, 47100 Artas, Kostakioi, Greece (emails: pkarvelis@kic.teiep.gr, georgoul@kic.teiep.gr, stylios@teiep.gr).

I. P. Tsoumas is with the Siemens Industry Sector, Large Drives R&D, Nuremberg, Germany (e-mail: ytsoumas@ieee.org).

J. A. Antonino-Daviu is with the Instituto de Ingeniería Energética, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain (e-mail: joanda@die.upv.es).

V. Climente-Alarcón is with the Department of Electrical Engineering and Automation, Aalto University, P.O. Box 13000, 00076 Espoo, Finland (e-mail: viclial@ieee.org).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2015.2463680, IEEE Transactions on Industrial Informatics

It is known that a breakage of a bar in the rotor cage leads to a distortion in the air-gap magnetic field. More specifically the effect of the breakage is typically represented by a 'fault field' that is superposed to the normal field that is present under healthy conditions. The spatial waveform of the air-gap flux density caused by the fault field is a stepped bipolar wave, whose amplitude and spectral composition cyclically change over time. Rigorous characterization of this field can be found in [13]. This 'fault field' induces multiple current harmonics in the stator windings. The identification of these harmonics in the Fourier spectrum of the steady-state current is the basis of the MCSA method [13]. In this regard, the evaluation of the amplitude of the left sideband harmonic (LSH) has drawn the most attention since this amplitude is closely related to the fault severity [2]. The frequency of the LSH is determined by the equation [2], [14]:

$$f_{LSH} = |1-2s| \cdot f_s , \quad (1)$$

where f_s is the power supply frequency and s is the slip.

MCSA is predominantly used in many industrial sites as well as by most of the few available condition monitoring devices to assess the rotor condition [2]. Traditional MCSA has however, important drawbacks as extensively reported by several authors [14], [15], such as the incorrect diagnostics of this tool (either false negatives or false positives) and its inability to cope with time varying conditions, that can have huge economic repercussions [16]-[17].

These problems of MCSA have justified the attempts to promote the industrial penetration of new advanced techniques. Indeed, some of them can avoid part of the MCSA constraints, increasing the reliability of the diagnostic procedure, as reported in several works. In this context, the analysis of start-up current using advanced signal processing tools has been proven to be especially suitable; the fault-associated patterns appearing in the resulting time-frequency maps are very unlikely to be caused by a different phenomenon or reason, a fact that justifies its use, especially in controversial cases in which application of traditional MCSA is not suitable [16]-[18].

In spite of the advances in this area, most of these techniques still rely on a qualitative interpretation of the resulting patterns that must be carried out by an expert user [19]. In other words, the implementation of these techniques in unsupervised systems that do not require continuous human intervention is still in an experimental phase.

This work proposes a new, computationally efficient method to automatize the detection process of rotor faults and produce an alarm in case of a detected fault. The proposed method is based on the transformation of the discrete time, continuous valued wavelet approximation signal [13] of the start-up current, to a discrete time, discrete valued signal. The process known as discretization or symbolization allows for the application of highly efficient algorithms successfully applied at other fields such as bioinformatics or text mining [20], which may come handy in embedded system implementation or online monitoring. Moreover symbolic data representation is usually less sensitive to measurement noise, a situation which often occurs in industrial environments [21].

In the field of electric equipment condition monitoring, symbolic representations have been primarily used for anomaly detection [22] using statistical properties of the generated symbolic sequences. Symbolic dynamics were involved in the detection of broken bars both for inverter-fed [23] and line-fed motors [24], and for the detection of stator voltage imbalance in induction motors [25]. In [26] a variety of fault conditions in an induction motor (voltage imbalance, bearing faults, broken bars) were detected using symbolic filtering. In [27] the demagnetization of permanent magnet machines was estimated while in [28] the state of charge of a battery was estimated using symbolic time series analysis.

All the aforementioned symbolic dynamic approaches involve steady state operation. In [29] the Symbolic Aggregate approXimation (SAX) method was employed for the detection of broken bars during start-up. In [30] an intelligent icons representation capable of discriminating between healthy and faulty situations is presented without however being able to quantify the severity of the fault. In [31] the information extracted from the second complex Intrinsic Mode Function (IMF) of the start-up current was condensed into a symbolic time series which was then classified using a scheme based on discrete Hidden Markov Models (HMMs).

In this work an approach that builds upon the basic blocks of SAX but uses a different partitioning for the discretization process is proposed. The achieved results using both simulated and experimental data are promising, indicating the potential use of the method for the automatic identification of broken bars in induction motors.

The rest of the paper is structured as follows: Section II provides the necessary background of the different components of the proposed procedure along with the description of the method. Section III presents the testing procedure and the achieved results both for the simulation and the experimental test case. Finally, Section IV concludes the paper providing also some hints for future work.

II. Proposed Procedure

The proposed diagnostic procedure is depicted in Fig. 1 and as can be seen it involves a number of stages. In the rest of the section the technical background of each of the stages is presented.

A. Discrete Wavelet Transform for the extraction of the fault related approximation signal

Discrete Wavelet Transform (DWT) decomposes a signal $x(t)$ into an approximation signal and a number of detail signals, each one representing a particular "coarseness" of the signal:

$$\begin{aligned} x(t) &= \sum_i a_i^n \varphi_i^n(t) + \sum_i \sum_j^n d_i^j \psi_i^j(t) \\ &= a_n + \sum_{j=1}^n d_j \end{aligned} , \quad (2)$$

where a_n and d_j are the approximation signal at level *n* and

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2015.2463680, IEEE Transactions on Industrial Informatics

the detail signal at level j respectively, which in turn are constructed by the scaling a_i^n and wavelet coefficients d_i^j using the scaling φ^n and wavelet functions ψ^j.

Signals containing a mixture of features that reside at different time and frequency resolutions are well-suited for DWT analysis [32]. The detail signal $d_j(t)$ represents the frequency content of the signal $x(t)$ residing in $\left[f_{samp}/2^{j+1}, f_{samp}/2^j\right]$, where f_{samp} is the sampling frequency. The lowest frequency content $\left[0, f_{samp}/2^{n+1}\right]$ of the original signal $x(t)$ is represented by the approximation signal $a_n(t)$.

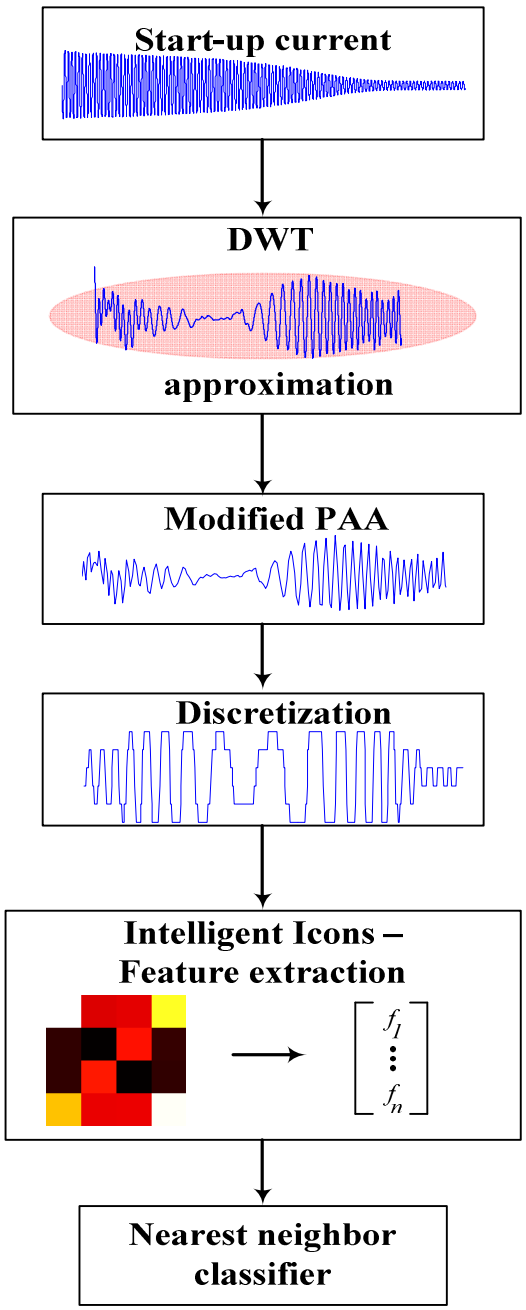

Fig. 1. The proposed diagnostic approach.

As it has been shown in [13], [16], [18], [33] in the case of broken rotor bars, the faulty component can be "isolated" by the approximation signal $a_n(t)$, provided that a suitable decomposition level is selected. For example for the case of $f_{samp} = 5$ kHz and $f_s = 50$ Hz, the appropriate level is 6, with the approximation signal containing frequencies $[0, 39.06]$ kHz [18]. This specific interval has been proven quite effective in isolating the evolution of the LSH [18] (even though more elaborate uses of the DWT have been proposed [34]). When a different f_{samp} is used that does not lead to that specific frequency content, sampling rate conversion can be achieved using polyphase filters [35].

Different mother wavelets have been used for the extraction of the approximation signal [13], [18], with satisfactory results. In this work the dmeyer (discrete Meyer wavelet) is selected and used in all conducted experiments [18]. In Fig. 2, the approximation signal at level 6 (which corresponds to the $[0,\ 39.06]$ Hz interval – 5 kHz sampling frequency) of the start-up current of the laboratory motor described in the Appendix (Table IV), is depicted at no load condition for a healthy, a one broken bar and a two broken bar situation. Note: all currents are normalized to have maximum amplitude approximately equal to one before the application of the DWT.

B. Piecewise Aggregate Approximation (PAA)

PAA, which was independently proposed in [36], and [37], divides a discrete time signal (a time series) $x = \{x[1], x[2], \ldots, x[N]\}$ of length N, into w segments of equal length (N/w) and then each segment is replaced by the average value of the segment, creating a new time series $x_{PAA} = \{\bar{x}[1], \bar{x}[2], ..., \bar{x}[w]\}$ where:

$$\bar{x}[i] = \frac{w}{N} \sum_{j=\frac{N}{w}(i-1)+1}^{\frac{N}{w}i} x[j], \text{ for } i = 1, 2, ..., w. \quad (3)$$

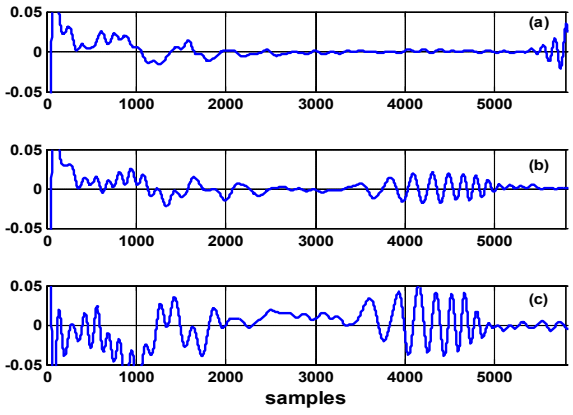

Fig. 2. The approximation signal of the start up current for: a) a healthy machine, b) a machine with one broken bar and c) a machine with two broken bars under no load condition.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2015.2463680, IEEE Transactions on Industrial Informatics

An example of the PAA representation of a time series is shown in Fig. 3. PAA is used as an effective first stage of dimensionality reduction increasing computational efficiency.

PAA assumes that N can be divided exactly by w. A modified version that can handle situations where N cannot be divided exactly by w was proposed as part of the generalized SAX [38], [39]. Assuming that the discrete time signal has a normalized sampling period ($x=\{x[nT]\}, n=1,2,...,N$ and $T=1$), an analytic expression for the calculation of generalized PAA can be written in a summation form as follows [40]:

$$\bar{x}[i]=\frac{w}{N}\left(\left(\left(\frac{N}{w}i+1\right)-\left\lfloor\frac{N}{w}i+1\right\rfloor\right)\cdot x\left[\left\lfloor\frac{N}{w}(i)+1\right\rfloor\right]\right)$$
$$+\frac{w}{N}\left(\sum_{j=\left\lfloor\frac{N}{w}(i-1)+1\right\rfloor+1}^{\left\lfloor\frac{N}{w}i+1\right\rfloor-1}x[j]\right)$$
$$+\frac{w}{N}\left(\left(1-\left(\left(\frac{N}{w}(i-1)+1\right)-\left\lfloor\frac{N}{w}(i-1)+1\right\rfloor\right)\right)\cdot x\left[\left\lfloor\frac{N}{w}(i-1)+1\right\rfloor\right]\right)$$
$$\text{for } i=1,2,...,w \qquad (4)$$

where $\lfloor y\rfloor$ is the greatest integer less than or equal to y. Eq. 4 reduces to Eq. 3 when N can be divided exactly by w.

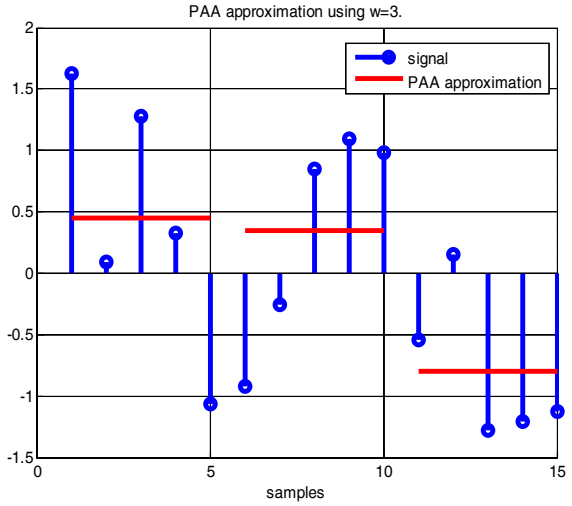

Fig. 3. An example of the application of PAA to a signal with only 15 samples (marked with open circles) using $w=3$. Each consecutive non overlapping 5-sample long segment is averaged to produce the PAA representation.

C. Discretization

The discretization of a continuous valued signal for the creation of a sequence of symbols drawn from an alphabet Σ is performed by partitioning the range of the input signal into a number of disjoint regions. In this work the Maximum Entropy Partitioning (MEP) is adopted [41]. As its name implies, MEP creates symbols in such a way as to maximize the entropy of the created string. For a discrete time continuous valued signal with w samples and for an alphabet with cardinality $|\Sigma|$ the MEP procedure is as follows [41], [42]:

a) sort the samples of the original discrete time continuous valued signal in an ascending order,

b) define the partitioning limits starting from the first point of the sorted signal from the first step such as $\lfloor w/|\Sigma|\rfloor$ data points will lie within each section and

c) use the partition obtained in the second step, to transform the original discrete time continuous valued signal to a string of symbols (if the value of a data point lies within or on the lower bound of a partition, it is coded with the symbol associated with that partition).

A graphical representation of the procedure is depicted in Fig. 4 for the case of an alphabet with only 3 symbols ("a", "b", "c") and 150 samples (the value of the 50th sample of the sorted series defines the lower cutting point (th_1) while the value of the 100th sample defines the upper cutting point (th_2) – therefore any sample which has a value lower or equal than th_1 will be assigned the symbol "a", any sample which has a value that is higher than th_1 and lower or equal than th_2 will be assigned the symbol "b" and the rest of the samples will be assigned the symbol "c"). As it can be seen with that partition all three symbols are equipropable ("a", "b" and "c" appear 50 times each), therefore leading to a maximization of the entropy.

MEP allows for more symbols –finer resolution- in regions with higher information content, while regions with lower information content undergo a sparser partitioning with fewer symbols assigned.

D. Feature extraction – Intelligent Icons

All the previous steps have led to the creation of a symbolic sequence. This sequence is further condensed in order to come with a probabilistic representation. Along these lines the simplest approach is to use the intelligent icons [43], [30] method with or without the illustration part. The idea behind the intelligent icons is to count the occurrences of symbols or words of symbols creating this way, approximations of the underlying probability mass functions. The rational is similar to the D-Markov modeling approach employed in [24]-[26], where a transition matrix is created by counting transitions between states (where each state is just a word of length D).

For creating an intelligent icon first we assign to each letter of the alphabet a unique value k:

$$a=0, b=1, c=2, d=3. \qquad (5)$$

Each word has an index for the location of each symbol in the table of the icon. For clarity we can show them explicitly as subscripts. For example, if the first word is $\alpha_0\alpha_1$ then $k_0=a=0$ and $k_1=a=0$. In order to map a sub word to the table represented by the icon we can use a mapping like the ones presented in Fig. 5.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2015.2463680, IEEE Transactions on Industrial Informatics

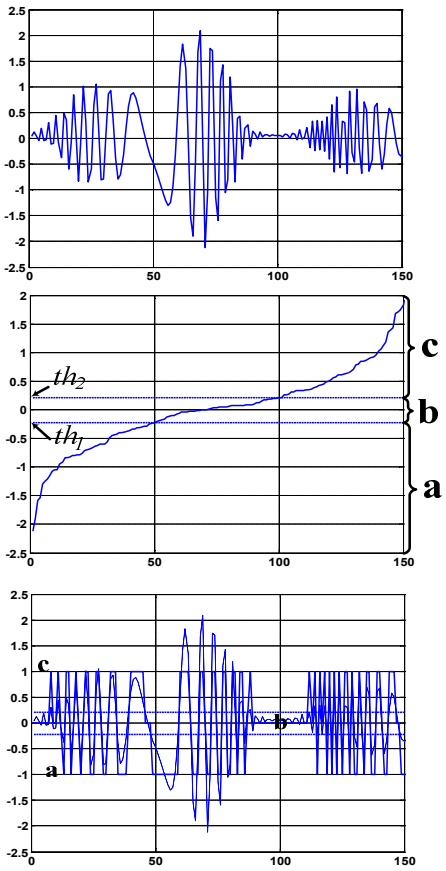

Fig. 4. Schematic explanation of the sequence creation process for the case of a three-symbol alphabet. The original signal is shown on the top of the figure In the middle the partitioning using the MEP is depicted, with the thresholds dividing the samples into three regions, with each one containing 50 samples, leading to three equiprobable symbols. The bottom caption shows the transformation of the original signal into a sequence. For illustrative purposes we only depict the sequence as a continuous plot overlaid on the original signal which is depicted in the background using a dashed line.

The following equations can be used in order to find the row and column of each sub word in the intelligent icon table:

$$col = \sum_{n=0}^{l-1} (k_n \cdot 2^{l-n-1}) \bmod 2^{l-n} , \quad (6)$$

$$row = \sum_{n=0}^{l-1} (k_n div 2) \cdot 2^{l-n-1} , \quad (7)$$

where mod represents the modulo operation and *div* the integer division operation.

For example for the sub word *aa* where $k_0 = a = 0$, $k_1 = a = 0$ and $l = 2$; substituting in Eq. (6) and Eq. (7) we get $col = 1$ and $row = 1$ respectively (the first element of the mapping of Fig. 5(b)).

Intelligent Icons, also called intelligent bitmaps [43], were originally developed to display time series in a more compact form. Along this line, Fig. 6 displays the intelligent icons for the case of a healthy machine and a machine with one and two broken bars, which is described in section III.B, for different durations of the start-up phenomenon. As it can be seen the icons are pretty distinct.

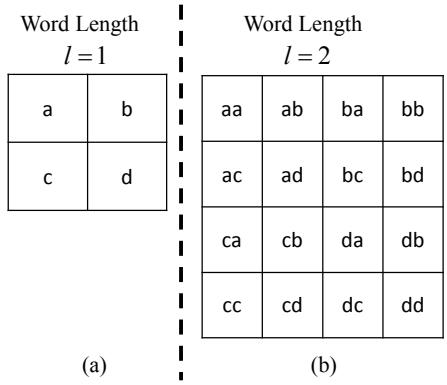

Fig. 5. Two mappings for two different word lengths, (a) $l = 1$ and (b) $l = 2$ for an alphabet $\Sigma = \{a, b, c, d\}$ of size $|\Sigma| = 4$.

However in this work, the goal is not to trade one representation (one dimensional time series) for another (a two dimensional icon) rather than to develop an automated procedure for the diagnosis of broken bars. Therefore the estimated probability mass functions of the words are treated as feature vectors to be fed into a classifier. This is a similar to the bag-of-patterns representation proposed in [44], which was inspired by the bag-of-words representation from information retrieval and text mining.

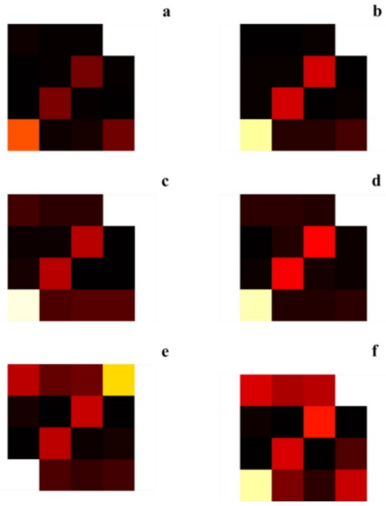

Fig. 6. Intelligent Icons representation of the approximation signal using, $w = 250$, $|\Sigma| = 4$ and word length equal to 2 for the case of a normal machine (a)-(b), a machine with one broken bar (c)-(d) and a machine with two broken bars (e)–(f). The start-up signals come from an actual machine (Section III.B). Icons (a), (c), and (e) correspond to no load and (b), (d) and (f) to half load condition.

E. Classification-Diagnosis

The final stage of the proposed approach consists of the nearest neighbor (NN) classifier the simplest member of the k-NN family [46]. This 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.

The NN classifier has the desired property [47] of being a parameter-free algorithm, leaving the tuning process entirely to the selection of the parameters involved during the application

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2015.2463680, IEEE Transactions on Industrial Informatics

of PAA, the discretization and the selection of the word length as it will be explained in the next section.

III. Experimental Procedure – Results

The proposed method was tested using both simulation data and data coming from a squirrel cage induction machine and three large industrial motors. The evaluation procedure involved, as well as the achieved results, are presented in the rest of this section.

A. Simulation Results

Simulated signals were obtained using a model derived from an analytical model of induction machine under fault that was developed in Matlab Simulink environment. The model is able to simulate induction machines with rotor asymmetries and eccentricity in different load conditions, both stationary and transient states and yielding magnitudes such as currents, speed and torque [48].

Ten healthy start-ups and ten start-ups for each one of the faulty conditions (one and two broken bars) were simulated with different durations of the transient phenomenon. Therefore a total of 30 cases were considered each one of them being different from all the rest.

To assess the performance of the approach a 10-fold (stratified) cross validation (CV) procedure was repeated 10 times (outer loop) (10×10 CV) [49]. This means that each time three cases (one from each condition) were left out and the rest 27 cases (9+9+9) were used for training, and the whole procedure was repeated ten times with reshuffling taking place between the 10 different runs of the outer loop. This way the training and testing sets are not the same in any of the 10 different runs/repetitions. This procedure is adopted in order to avoid overoptimistic or overpessimistic results due to a particular partition of the data. Moreover, for the tuning of the algorithm, i.e. for selecting the number of windows w (250, 500, 1000 or 2000), the alphabet size $|\Sigma|$ (4, 5, 6, 7 or 8) as well as the word length (2 or 3), 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 [49]. During the training process the MEP approach operated on the pooled sample of all the training signals. This way the different amplitude of the chirp like components in the case of the one and the two broken bars were taken into account. Moreover since the beginning and the ending of the recordings are similar for almost all cases, 20% of the recorded signal was excluded from further processing.

The classification results for the simulation data are depicted at the first row of Table I indicating perfect discrimination performance of the proposed approach.

B. Experimental Results-laboratory machine under different loading conditions

For the first part of experimental evaluation the squirrel cage induction motor of the Appendix (Table IV) was used coupled directly to a DC machine (load). Healthy and faulty rotors (with one and two broken bars artificially forced by drilling a hole in the corresponding bar, just in the junction point between the bar and the short-circuit end ring) were tested. A picture of one of the rotors (with one broken bar) is depicted in Fig. 7.

This first set of experiments aims to test the variability of the approach under different loading conditions. More specifically a number of start-up signals were acquired under different loading conditions with different durations of the transient phenomenon. Fig. 8 to Fig. 10 depict the approximation signal, its discrete representation as well the illustration of the extracted features in the form of an intelligent icon for three different cases (healthy- one broken bar-two broken bars) at half load condition ($|\Sigma|=4$, $w=500$ and word length equal to 2) for the machine described in the Appendix (Table IV). Stator currents throughout the experimental evaluation were sampled with a frequency of 5 kHz. Therefore the approximation signal was acquired using DWT up to level 6. The experimental data collected is summarised in Table II.

Fig. 7. Picture of the rotor with one broken bar.

This experimental setting involved the use of data coming from two loading conditions for training and the third one for testing (i.e. no-load and half load for training and full load for testing). Therefore three sets of experiments were carried out. For the selection of the parameters an internal cross validation procedure was carried out within the training set.

The same procedures as in the case of the simulation study (pooling together of the training data for the determination of the breaking points, exclusion of the first and last 20% samples of the approximation) were employed and the results are summarized in Table I (rows 2-4). As it can be seen the method seems to be robust to load variations, capable to "extrapolate" to unseen situations.

To further test the approach a second set of experiments was carried out. This time data from only one loading condition were involved for training and the data from the other two conditions were reserved for testing. In effect the training and testing data sets of the previous experimental study were switched. The same training protocol was employed. The results are also summarized in Table I (rows 4-7). As it can be seen in almost all cases but one data from a single condition can be used effectively for building a diagnostic module. Only in the case of a model trained using only data coming from no load conditions the method mistakenly assigned a case of two broken bars to the class of one broken bars (therefore 9/10 of the two broken bar cases were correctly classified and in total 29/30 of all involved test cases).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2015.2463680, IEEE Transactions on Industrial Informatics

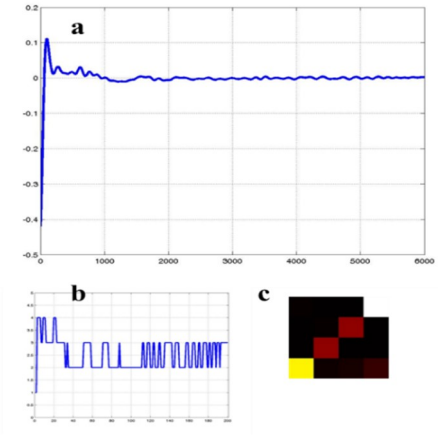

Fig. 8. Illustration of the main stages of the proposed procedure for a healthy machine operating at half load: a) approximation signal, b) Discrete representation, c) Intelligent Icon representation.

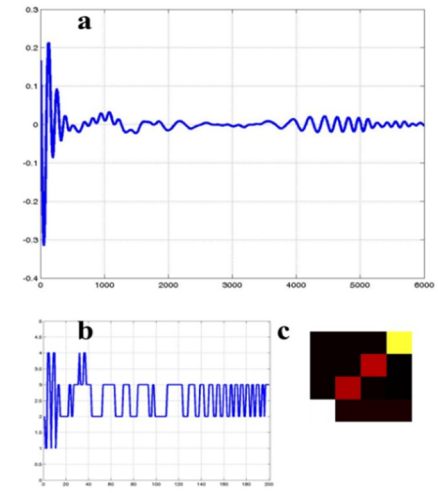

Fig. 9. Illustration of the main stages of the proposed procedure for a machine with one broken bar operating at half load: a) approximation signal, b) Discrete representation, c) Intelligent Icon representation.

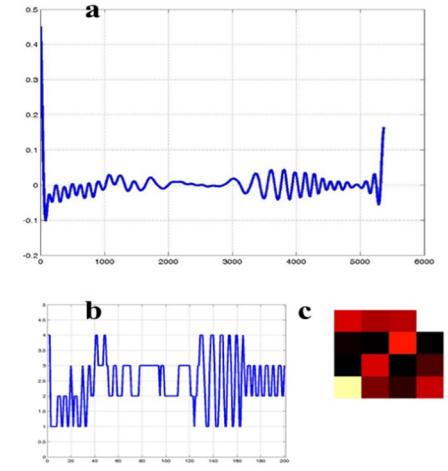

Fig. 10. Illustration of the main stages of the proposed procedure for a machine with two broken bars operating at half load: a) approximation signal, b) Discrete representation, c) Intelligent Icon representation.

TABLE I.
CLASSIFICATION ACCURACY

	Healthy	One broken bar	Two broken bars	Overall
Simulation	100%	100%	100%	100%
*ExperimentalA1	100%	100%	100%	100%
**ExperimentalA2	100%	100%	100%	100%
***ExperimentalA3	100%	100%	100%	100%
+ExperimentalB1	100%	100%	90%	96.7%
++ExperimentalB2	100%	100%	100%	100%
+++ExperimentalB3	100%	100%	100%	100%

**ExperimentalA1: full load data for testing.*
***ExperimentalA2: no load data for testing.*
****ExperimentalA3: half load data for testing.*
+ExperimentalB1: no load data for training
++ExperimentalB2: half load data for training.
+++ExperimentalB3: full load data for training.

C. Experimental Results-industrial large machines

In the previous section, we examined the efficiency of the method to diagnose the condition of an induction machine operating at different loading conditions. To corroborate the validity of the methodology in other machines and, hence, to prove its ability to generalize, several real field motor signals were used. All motors were operating in real mining facilities and the start-up current signals were obtained during one of the periodic inspections of these machines. The general characteristics of the field motors are provided in the Appendix (Tables V-VII). Note that these are very large motors that substantially differ from the motor tested in the laboratory.

TABLE II.
EXPERIMENTAL DATA FOR THE LABORATORY MOTOR

	Healthy	One broken bar	Two broken bars
No load	5	5	5
Half load	5	5	5
Full load	5	5	5

Two of the field motors (M1 and M2) had different levels of rotor asymmetry. In the case of motor M1, the rotor asymmetry was confirmed after rotor disassembly and visual inspection, while in the case of M2 there were clear symptoms of rotor asymmetry. On the other hand, motor M3 was a healthy machine.

For this experiment, the data collected from the laboratory motor (Table II) were used to train the classifier. The assessment of the three motors based on the proposed method are summarised in Table III. We must note that sampling rate conversion took place since for M1 the sampling rate was 6 kHz, while for M2 the sampling rate was 4048 Hz and all the training signals were sampled at 5 kHz.

The trained classifier was able to detect that motors M1 and M2 had asymmetries without signalling a false alarm for M3. The results are consistent with the fact that the method assigned to M1 the class with the higher rotor asymmetry

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2015.2463680, IEEE Transactions on Industrial Informatics

trained to recognise (two broken bars), while assigning M2 to the category of one broken bar.

IV. Conclusions

An automated method for the diagnosis of broken bars using the information contained in the start-up current was presented, which can be used to signal an alarm for further inspection. The method combines DWT for the isolation of a faulty component that arises in the case of a bar brakeage and a symbolic analysis approach for transforming the information contained in the component isolated in the previous stage to useful information. The efficiency of the method was tested using both simulated and experimental data and a very simple classifier proving that once the feature engineering step is suitable [47] simple methods can be used for the final classification stage. As in most data-driven methods, the more representative data available during the training period, the better. This was verified during the different experiments involving the laboratory data where the only misclassification occurred when a reduced set of data (only one loading condition) was used for training. However even in that case, the method did not confuse healthy with faulty conditions.

The efficiency of the representation can be highlighted by inspecting Fig. 11, which depicts the projection of the data coming from the laboratory setting into three dimensions using classic multidimensional scaling. As it can be seen, even in this low dimensional space where the discrimination might be a bit compromised the three classes/conditions seem to occupy different portions of the space even though the one and two broken bar categories can be quite close at some parts of that space. This is probably the reason for the one misclassification that occurred during the "ExperimentalB1" set up. On the other hand the healthy class seems to be quite distinct from the faulty ones.

Compared to [29] this method is not only capable of detecting the presence of fault but it can also discriminate between one and two broken bars. Furthermore the method is very simple and the whole process requires minimum intervention, contrary to the method proposed in [31], where all three currents should be monitored, while the discretization was based on expert judgement and the classifier was much more involved than the simple one used in this study.

Moreover the method was tested using also real life data. Even though no additional data were available for motor M1-M3 the method was capable to discriminate the healthy motor M3 from the two faulty motors. The promising results indicate that the approach can be used at least in an anomaly detection scenario. Nevertheless further testing is needed before its adoption into industrial practice, along with testing with the simultaneous presence of other common faults.

TABLE III.
ASSESSMENT OF THE LARGE MOTORS

Motor Label	Condition
M1	Two broken bars
M2	One broken bar
M3	Healthy

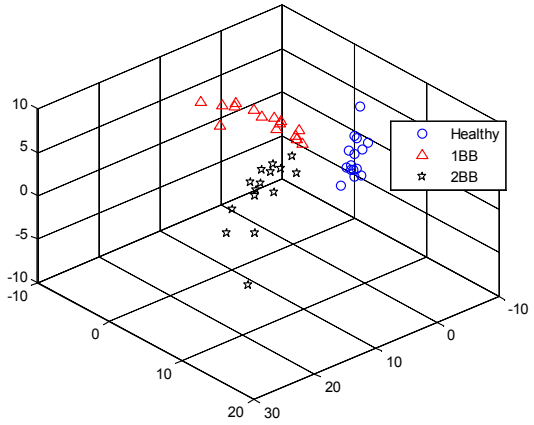

Fig. 11. Projection of the data coming from the laboratory setting into 3 dimensions, using, $w = 2000$, $|\Sigma| = 8$ and word length equal to 2. The circles correspond to the healthy condition, the triangles to one broken bar and stars to two broken bars.

Furthermore the method was tested only for completely broken bars. In future work more demanding scenarios will be pursued with only partially broken bars and under accelerated testing conditions to investigate the potential of the proposed approach to be integrated in a prognostic framework.

In addition to this, in future work the diagnostic problem will also be tested in the more demanding inverter-fed scenario as well as under the use of fault tolerant controllers which might attenuate the fault signature.

V. APPENDIX

Rated characteristics of the tested motors.

TABLE IV.
LABORATORY MOTOR

Rated Power	1.1Kw
Rated Frequency	50 Hz
Rated Voltage	400 V
Rated primary current	2.7 A
Rated speed	1410 rpm
Rated slip	0.06
Connection	Star
Number of pole pairs	2
Number of rotor bars	28
Number of stator slots	36

TABLE V.
FIELD MOTOR: M1

Application	High speed coal mill
Rated Power	320 kW
Rated frequency	50 Hz
Rated speed	740rpm
Number of poles	8

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2015.2463680, IEEE Transactions on Industrial Informatics

TABLE VI.
FIELD MOTOR: M2

Application	Mill fan
Rated Power	400 kW
Rated frequency	50 Hz
Rated speed	1492rpm
Number of poles	4

TABLE VII.
FIELD MOTOR: M3

Application	Compressor
Rated Power	1'2 MW
Rated frequency	50 Hz
Rated speed	2998 rpm
Number of poles	2 poles

REFERENCES

[1] D. B. Durocher, G. R. Feldmeier ,"Predictive versus preventive maintenance," *IEEE Trans. Ind. Appl*, vol. 10, no.5, pp. 12-21, Oct. 2004.

[2] W. T. Thomson, and M. Fenger, "Current signature analysis to detect induction motor faults," *IEEE Industry Applications Magazine,* pp. 26-34, July/August 2001.

[3] V. Sousa Santos, P. R. V. Felipe, J. R. Gomez Sarduy, N. Lemozy, A. Jurado, and E. C. Quispe, "Procedure for Determining Induction Motor Efficiency Working Under Distorted Grid Voltages," *IEEE Trans. Energy Convers.*, vol. 30, no. 1, pp. 331-339, 2015.

[4] Allianz Insurance, "Allianz Survey", VDE Colloquium, Germany, June 28, 2001.

[5] P. Zhang, Y. Du, T. G. Habetler, and B. Lu, "A Survey of Condition Monitoring and Protection Methods for Medium-Voltage Induction Motors," *IEEE Trans. Ind. Appl*, vol.47, no.1, pp.34-46, Jan.-Feb. 2011.

[6] M. Riera-Guasp, J. A. Antonino-Daviu, and G. A. -Capolino, "Advances in electrical machine, power electronic and drive condition monitoring and fault detection: State of the art," *IEEE Trans. Ind. Electron.* , vol. 62, no. 3, Mar. 2015.

[7] M. Sahrauoi, A. J. M. Cardoso, and A. Ghoggal, "The use of modified Prony method to track the broken rotor bar characteristic frequencies and amplitudes in three-phase induction motors," *IEEE Trans. Ind. Appl*., vol 51, no 3, 2136-2147, May/June 2015.

[8] S. H. Kia, H. Henao, and G.-A. Capolino, "Efficient digital signal processing techniques for induction machine fault diagnosis," in *Proc. IEEE WEMDCD , Paris, France*, pp. 233–246, 2013.

[9] K. N. Gyftakis, and J. C. Kappatou, "The zero-sequence current as a generalized diagnostic mean in Δ-connected three-phase induction motors," *IEEE Trans. Energy Convers.*, vol. 29, no. 1, pp. 138-148, 2014.

[10] B. M. Ebrahimi, M. J. Roshtkhari, J. Faiz, and S. V. Khatami, "Advanced Eccentricity Fault Recognition in Permanent Magnet Synchronous Motors Using Stator Current Signature Analysis," *IEEE Trans. Ind. Electron.*, vol. 61, no. 4, pp. 2041–2052, Apr. 2014.

[11] M. B. K.Bouzid, and G. Champenois, "New expressions of symmetrical components of the induction motor under stator faults," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4093-4102, Sept. 2013.

[12] M.J. Picazo-Rodenas, J. Antonino-Daviu, V. Climente-Alarcon, R. Royo-Pastor, and A. Mota-Villar, "Combination of non-invasive approaches for general assessment of induction motors," *IEEE Trans. Ind. Appl*., vol 51, no 3, pp. 2172-2180, May/June 2015

[13] M. Riera-Guasp, J. A. Antonino-Daviu, J. Roger-Folch, and M. P. Molina Palomares, "The Use of the Wavelet Approximation Signal as a Tool for the Diagnosis of Rotor Bar Failures," *IEEE Trans. Ind. Appl*., vol. 44, pp. 716-726, 2008.

[14] F. Filippetti, A. Bellini, and G. Capolino, "Condition monitoring and diagnosis of rotor faults in induction machines: State of art and future perspectives," *In Proc. IEEE Workshop on Electrical Machines Design Control and Diagnosis* (WEMDCD), pp. 196-209, 2013

[15] G. M. Joksimovic, J. Riger, T. M. Wolbank, N. Peric, and M. Vasak, "Stator-current spectrum signature of healthy cage rotor induction machines." *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4025-4033, 2013.

[16] J. A. Antonino-Daviu, M. Riera-Guasp, J. R. Folch, and M. Pilar Molina Palomares, "Validation of a new method for the diagnosis of rotor bar failures via wavelet transform in industrial induction machines," *IEEE Trans. Ind. Appl.,* vol. 42, pp. 990-996, 2006.

[17] J. Park, B. Kim, J. Yang, K. Lee, S.B. Lee, E. J. Wiedenbrug, M. Teska, and S. Han, "Evaluation of the Detectability of Broken Rotor Bars for Double Squirrel Cage Rotor Induction Motors," *in proc. of the IEEE ECCE,* pp. 2493-2500, Sept. 2010.

[18] M. Riera-Guasp, J. A. Antonino-Daviu, M. Pineda-Sanchez, R. Puche-Panadero, and J. Perez-Cruz, "A general approach for the transient detection of slip-dependent fault components based on the discrete wavelet transform," *IEEE Trans. Ind. Electron*., vol. 55, no. 12, pp. 4167–4180, Dec. 2008.

[19] E. Cabal-Yepez, A.G., Garcia-Ramirez, R. J. Romero-Troncoso, A. Garcia-Perez and R. A. Osornio-Rios, "Reconfigurable Monitoring System for Time-Frequency Analysis on Industrial Equipment Through STFT and DWT," *IEEE Trans. Ind. Inf*., vol. 9, no. 2, pp. 760-771, 2013.

[20] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A Symbolic Representation of Time Series, with Implications for Streaming Algorithm,". in *Proc of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. San Diego, CA, pp. 2-11, 2003.

[21] C. S. Daw, C. E. A. Finney, and E. R. Tracy, "A review of symbolic analysis of experimental data." *Review of Scientific Instruments*, vol. 74, no. 2, pp. 915-930, 2003.

[22] A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection," *Signal Processing*, vol. 84, no. 7, pp. 1115-1130, 2004.

[23] R. Samsi, V. Rajagopalan, and A. Ray, "Wavelet-based symbolic analysis for detection of broken rotor bars in inverter-fed induction motors," in *IEEE American Control Conference*, 2006.

[24] R. Samsi, V. Rajagopalan, and A. Ray, 2006. Wavelet-Based Symbolic Analysis for Detection of Broken Rotor Bars in Inverter-Fed Induction Motors," in *IEEE American Control Conference*, pp. 3032-3038, 2006.

[25] R. Samsi, A. Ray, and J. Mayer "Early detection of stator voltage imbalance in three-phase induction motors," *Electric Power Systems Research*, vol. 79, no. 1, pp. 239-245, 2009.

[26] R. A. Gupta, A. K. Wadhwani, and S. R. Kapoor, "Early estimation of faults in induction motors using symbolic dynamic-based analysis of stator current samples," *IEEE Trans. Energy Convers*, vol. 26, no. 1, pp. 102-114, 2011.

[27] S. Chakraborty, E. Keller, A. Ray, and I. Mayer, "Detection and estimation of demagnetization faults in permanent magnet synchronous motors," *Electric Power Systems Research*, vol. 96, pp. 225-236, 2013.

[28] Y. Li, P. Chattopadhyay, and A. Ray, "Dynamic data-driven identification of battery state-of-charge via symbolic analysis of input–output pairs," *Applied Energy,* vol. 155, pp. 778-790, 2015.

[29] P. Karvelis, I. P. Tsoumas, G. Georgoulas, C. D. Stylios, J. A. Antonino-Daviu, and V. Climente-Alarcon, "An intelligent icons approach for rotor bar fault detection". in *Proc. Annual Conference of the IEEE Industrial Electronics Society (IECON), IECON 2013,* pp. 5526-5531, 2013.

[30] E.Keogh, L. Wei, X. Xi, S. Lonardi, J. Shieh, and S. Sirowy, "Intelligent Icons: Integrating Lite-Weight Data Mining and Visualization into GUI Operating Systems," in *Proc. of Sixth International Conference on Data Mining, ICDM '06*, pp. 912-916, 2006.

[31] G. Georgoulas, I. P Tsoumas, J. A. Antonino-Daviu, V. Climente-Alarcón, C. D. Stylios, E. D. Mitronikas, and A. N. Safacas, "Automatic pattern identification based on the Complex Empirical Mode Decomposition of the startup current for the diagnosis of rotor asymmetries in asynchronous machines," *IEEE Trans. on Ind. Electronics*, vol. 61, vo. 9, pp 4937-4946, Sep. 2014.

[32] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed., Academic Press, Oxford, 1999.

[33] J. Antonino-Daviu, M. Riera-Guasp, M. Pineda-Sanchez, and R. Pérez, "A critical comparison between DWT and Hilbert–Huang-based methods for the diagnosis of rotor bar failures in induction machines," *IEEE Trans. Ind. Appl.* , vol. 45, no. 5, pp. 1794–1803, 2009.

[34] J. Antonino-Daviu, S. Aviyente, E. G, Strangas, and M. Riera-Guasp, "Scale invariant feature extraction algorithm for the automatic diagnosis of rotor asymmetries in induction motors," *IEEE Trans. Ind. Inf*., vol. 9, no. 1, pp 100-108, 2013.

[35] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms, and Applications*, 4th ed. NJ, Prentice-Hall, 2007.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TII.2015.2463680, IEEE Transactions on Industrial Informatics

[36] E. Keogh, K. Chakrabarti, M. Pazzani, and S, Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *J. Knowledge and Information Systems*, vol. 3, pp. 263-286, 2000.

[37] K. Yi, and C, Faloutsos, "Fast time sequence indexing for arbitrary Lp norms," *in Proc. of the 26th International Conference on Very Large Databases*, Cairo, Egypt, 2000.

[38] SAX Homepage. /http://www.cs.ucr.edu/~eamonn/SAX.htm

[39] J. Lin, E. Keogh, L. Wei, and S. Lonardi, "Experiencing SAX: a novel symbolic representation of time serie,". Data Mining and Knowledge Discovery, vol. 15, no. 2, pp. 107-144, 2007.

[40] G. Georgoulas, P. Karvelis, T. Loutas, and C. D. Stylios, "Rolling element bearings diagnostics using the Symbolic Aggregate approXimation," *Mech. Syst. Signal Process.*, vol. 60, pp. 229-242, 2015.

[41] V. Rajagopalan and A. Ray, "Wavelet-based space partitioning for symbolic time series analysis," in *Proc. of 44th IEEE Conference on Decision and Control, European Control Conference. CDC-ECC '05.*, pp. 5245–5250, 2005.

[42] V. Rajagopalan, and A. Ray, "Symbolic time series analysis via wavelet-based partitioning," *Signal Processing*, vol. 86, no. 11, pp. 3309-3320, 2006.

[43] N. Kumar, V. N. Lolla, E. J. Keogh, S. Lonardi, and C. Ratanamahatana, "Time-series Bitmaps: a Practical Visualization Tool for Working with Large Time Series Databases," in *the Fifth SIAM International Conference on Data Mining SDM*, 2005.

[44] J. Lin, R. Khade, and Y. Li, "Rotation-invariant similarity in time series using bag-of-patterns representation," *Journal of Intelligent Information Systems*, vol. 39, no. 2, pp. 287-315, 2012.

[45] H. J. Koskimaki, P. Laurinen, E. Haapalainen, L. Tuovinen, and J. Roning, "Application of the Extended nn Method to Resistance Spot Welding Process Identification and the Benefits of Process Information," *IEEE Trans. Ind. Electron*., vol. 54, no. 5, pp. 2823-2830, 2007.

[46] X. Jin, M. Zhao, T.W. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Trans. on Ind. Electronics*, vol. 61, no. 5, pp. 2441-2451, 2014.

[47] P. Domingos, "A few useful things to know about machine learning". *Communications of the ACM*, vol. 55, no. 10, pp. 78-87, 2012.

[48] M. Pineda-Sanchez, V. Climente-Alarcon, M. Riera-Guasp, R. Puche-Panadero and J. Pons-Llinares, "Enhanced Simulink Induction Motor Model for Education and Maintenance Training," *Systemics, Cybernetics and Informatics*, vol. 10, no. 2, 2012

[49] N. Japkowicz, and M. Shah, *Evaluating learning algorithms: a classification perspective*. Cambridge University Press, 2011.

Petros Karvelis received the B.Sc., M.Sc. and Ph.D. in Computer Science from the University of Ioannina, Greece, in 2001, 2004 and 2012 respectively. He is currently a post-doctorate research fellow at Knowledge and Intelligent Computing laboratory, Dept. of Computer Engineering, Technological Educational Institute of Epirus, Greece. His research interests include machine learning, image processing, biomedical engineering, and decision support systems.

George Georgoulas received the Dipl.-Eng. and Dr.-Eng. degrees in Electrical and Computer Engineering from the University of Patras, Greece, in 1999 and 2006 respectively. He is currently a post-doctorate research fellow at Knowledge and Intelligent Computing laboratory, Dept. of Computer Engineering, Technological Educational Institute of Epirus, Greece. His main scientific interests include machine learning, signal processing, failure detection, diagnosis and decision support systems.

Ioannis P. Tsoumas (M '04) was born in Athens, Greece. He received the Dipl.-Eng. and Dr.-Eng. degrees in electrical and computer engineering from the University of Patras, Greece, in 2000 and 2007 respectively.

He is currently with Siemens AG, Industry Sector, Drive Technologies Division, Large Drives, Nuremberg, Germany. His scientific interests include electrical machines and drives condition monitoring and fault diagnosis, modulation techniques for power converters, converters and drive systems efficiency, as well as motor-converter interaction. Dr. Tsoumas is member of the Association for Electrical, Electronic and Information Technologies (VDE) and of the Technical Chamber of Greece.

Jose Antonino-Daviu (S'04/M'08/SM'12), received his M.S. and Ph. D. degrees in Electrical Engineering, both from the Universitat Politecnica de Valencia, in 2000 and 2006, respectively. He was working for IBM during 2 years, being involved in several international projects. Currently, he is Associate Professor in the Department of Electrical Engineering of the mentioned University, where he develops his docent and research work. He has been invited professor in Helsinki University of Technology (Finland) in 2005 and 2007, Michigan State University (USA) in 2010, Korea University (Korea) in 2014 and Universite Claude Bernard Lyon 1 (France) in 2015. He has over 100 publications in international journals, conferences and books.His primary research interests are condition monitoring of electric machines, wavelet theory and its application to fault diagnosis and design and optimization of electrical installations and systems.

Vicente Climente-Alarcon (S'11/M'12), received his M.Sc. degrees in Chemical and Industrial Engineering in 2000 and 2011, and his Ph.D. degree in Electrical Engineering in 2012, all from the Universitat Politecnica de Valencia (Spain). He has worked as Assistant Professor in the School of Industrial Engineering of the mentioned university, on research tasks in the area of condition monitoring of electrical machines, and externally as a consultant in automation and management of power systems. Currently he is carrying out postdoctoral research at the Department of Electrical Engineering and Automation, Aalto University, Espoo, Finland. His scientific interests include the study of transients in electrical machines.

Dr. Chrysostomos Stylios (M'97/SM'15), is an Associate Professor at Dept. of Computer Engineering, Technological Educational Institute of Epirus, Greece. The academic year 2014-2015 is a visiting researcher at Electrical and Computer Engineering Dept., University of Alberta, Canada and at Computer Science Dept, University of Texas at El Paso. He received his Ph.D from the Department of Electrical & Computer Engineering, University of Patras (1999) and the diploma in Electrical & Computer Engineering from the Aristotle University of Thessaloniki (1992). He has published over 160 journal and conference papers and book chapters. His main scientific interests include: Fuzzy Cognitive Maps, Soft Computing, Computational Intelligence Techniques, Signal Processing methods and Decision Support Systems. Dr. Stylios is senior member of IEEE and member of the TC 8.2 and TC 5.4 of IFAC.