

Bearing Fault Detection and Diagnosis by fusing vibration data

George Georgoulas and George Nikolakopoulos

Department of Computer Science, Electrical and Space Engineering, Control Engineering Group
Luleå University of Technology
Luleå, Sweden
{geogeo, geonik}@ltu.se

Abstract—This article presents a simple method for the detection and diagnosis of bearing faults, by fusing the information coming from two accelerometers. The method relies on three simple and intuitive features, extracted from the data coming from accelerometers placed at two different locations of the system under investigation. Our preliminary results indicate that by using simple statistical measures, such as the elements of the covariance matrix of the two sensors, faults at an early stage can be detected. In the proposed scheme, the extracted features are fed to a k -nearest neighbor classifier for diagnosis purposes or to an ensemble of one-class detectors, if only the information from normal situation is available. As it is proven, based on experimental results, in both scenarios a remarkably high detection/diagnostic performance is achieved.

Keywords—Bearing fault diagnosis; seeded fault; nearest neighbor classifier; statistical descriptors.

I. INTRODUCTION

Fault detection, diagnosis and failure prognosis have received a considerable amount of attention during the past years, especially in industrial processes due to the negative impact of unexpected shutdowns, as well as due to the need for the development of more effective maintenance strategies. The latter has been marked by a shift in maintenance paradigm from Preventive Maintenance (PM) to Condition Based Maintenance (CBM) [1]. CBM relies heavily on the existence of monitoring parameters that can reflect the condition of the system. Depending on the system, different measurements can carry potential information for its condition. For the case of systems involving bearings (a typical construction is presented in Fig. 1) the most widely used systems for fault detection and diagnosis involve vibration monitoring [2] even though other measures can enhance the performance requiring however more expensive equipment [3].

Moreover, rolling element bearings faults account for 45–55% of asynchronous motor failures [4]. Therefore, it is of paramount importance to detect faults at an early stage before they lead to an irreversible failure of the system. To this end many approaches have been proposed ranging from anomaly detectors [5], [6] and fault diagnosis methods [7]–[9], to methods that try to capture the propagation of the fault within a failure prognosis framework [10].

As it was mentioned, vibration analysis is the tool of preference when it comes to bearing condition monitoring. The

reason is that a defect of any part of the bearing creates an alteration to the behavior of the bearing that is depicted in the way that the system vibrates. However, the raw signal is hardly ever used due to: a) its high dimensional nature and b) the presence of noise. To alleviate both of these problems, a feature extraction stage is usually involved which tries to capture the most relevant information embedded in the raw vibration signal.

Therefore, numerous feature extraction methods have been proposed. Some of them rely on time analysis of the monitored signal [11], while others involve frequency analysis, since the occurrence of the fault is followed by the occurrence of characteristic components in the frequency spectrum. Other methods employ time-frequency analysis, such as wavelet analysis [12] or the empirical mode decomposition (EMD) [13] to name just a few of the possible time-frequency tools. Other approaches use envelope analysis [5], features coming from the symbolic transformation of the vibration signal [9], cepstrum analysis [14], the kurtogram [15], nonlinear features [16], etc.

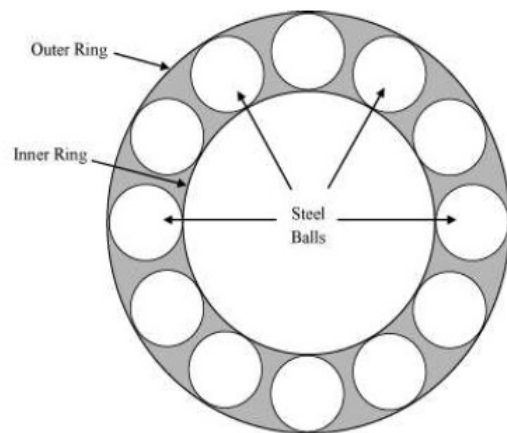


Fig. 1. A typical bearing where the outer and inner rings as well as the steel balls are depicted.

The major contribution of this article stems from the extraction of simple novel statistical descriptors, treating together the measurements coming from two accelerometers located at different positions of the test bed. The efficiency of these features for bearing condition monitoring is assessed

under two different scenarios: a) under an anomaly detection scenario, where a detector is trained using only data coming from normal operation, and b) in a diagnostic framework where the problem is treated as a classification one using data from faulty cases to train the classification algorithm.

The rest of the paper is structured as follows. In Section II the feature extraction as well as the anomaly detection and fault diagnosis techniques are presented. Section III presents the experimental set up as well as the different sets of experiments carried out. The experimental results are summarized in Section IV, while Section V concludes the paper.

II. METHODS

A. Feature Extraction

In a previous work [8] it was shown that in case we have at our disposal two measured vibration signals, informative features can be extracted by using an ellipsoid representation of the measurements. In this work a different approach is used to characterize the two dimensional measurements through a simpler method, which can also be quite robust to noise levels. It is well known that when we have two (or more) signals a way to characterize them individually, as well as in combination, is through their covariance matrix, which summarizes both the variance of each signal, as well as the degree that they are (linearly) related:

$$\mathbf{C} = \begin{pmatrix} \text{var}[X_1] & \text{cov}[X_1, X_2] \\ \text{cov}[X_2, X_1] & \text{var}[X_2] \end{pmatrix} \quad (1)$$

where

$$\text{var}[X] = \langle X^2 \rangle - \langle X \rangle^2 \quad (2)$$

and

$$\text{cov}[X_1, X_2] = \langle (X_1 - \langle X_1 \rangle)(X_2 - \langle X_2 \rangle) \rangle \quad (3)$$

with $\langle \cdot \rangle$ being the expectation operator.

A first empirical assessment of the potential utilization of the elements of the covariance matrix (in fact three out of four since $\text{cov}[X_1, X_2] = \text{cov}[X_2, X_1]$) can be done by observing the different subplots of Fig. 2, where the scatter plots of the raw vibration measurements are depicted (each vibration measurement corresponds to one of the axis) for four different conditions of a bearing (healthy condition and three different types of faults under the same load). As it can be seen, both the spread of the point-cloud along the axes (which can be quantified by the variances) as well as its "orientation" (which can be quantified by the covariance) is affected by the presence of a fault. The experimental measurements come from the testbed described in Section III.

Similar conclusions about the applicability of the method can be drawn from Fig. 3, where the three extracted features, noted as $[\text{var}[X_1], \text{var}[X_2], \text{cov}[X_1, X_2]]$, are depicted. In this case X_1, X_2 correspond to the measurements coming from the

first and the second sensor respectively, again for four different health conditions but this time with different loading.

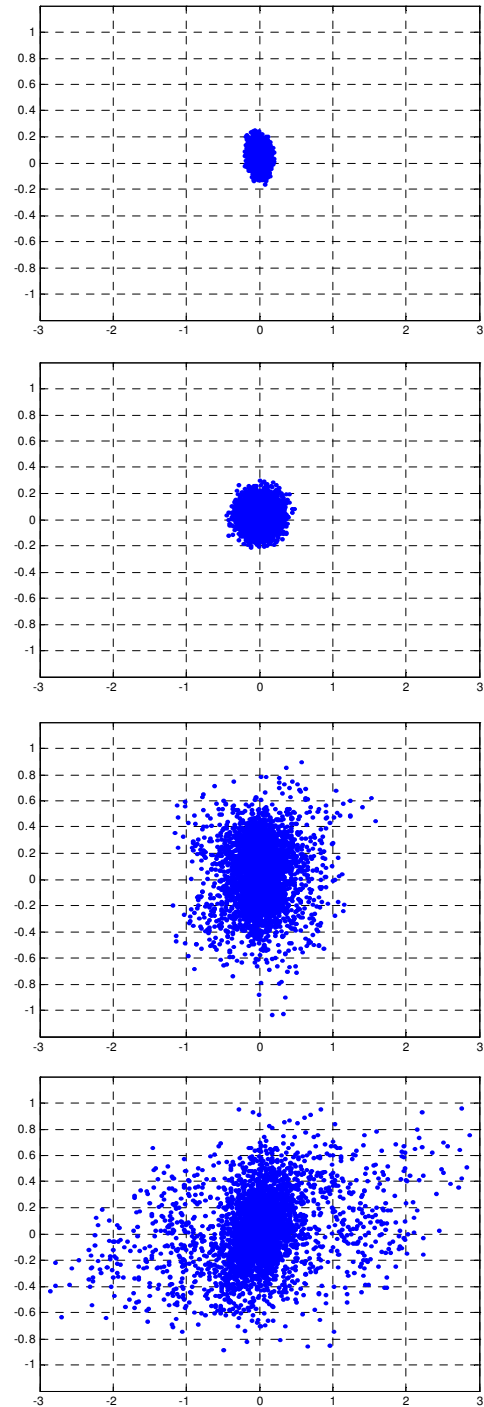


Fig. 2. Scatter plots of the raw measurements for the case of (from top to bottom): normal bearing, ball bearing fault, inner race fault and outer race faults. All recordings correspond to a load of 1 hp. The size of all faults is 0.007 inches. Each axis corresponds to the measurements coming from one of the two accelerometers

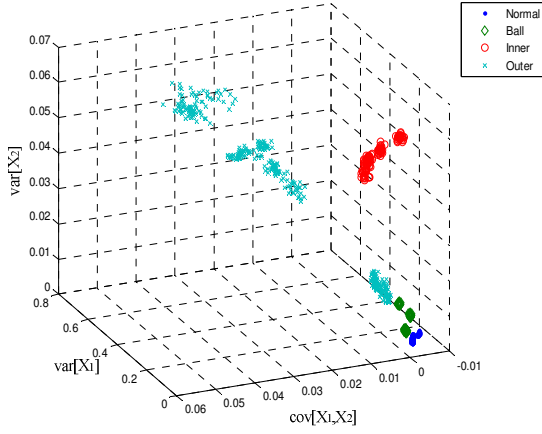


Fig. 3. Scatter plot of the three extracted features, for four different conditions, with different loadings pooled together. The higher spread of the outer race faults is due to the fact that different locations were also pooled together.

B. Anomaly Detection

Anomaly detection is the crudest approach to condition monitoring, since it only detects that there is something out of the usual happening to the system. On the other hand, when the gathering of faulty conditions is difficult (for a number of reasons: too costly to run accelerated fault tests or seeded fault tests, unknown number of faulty conditions etc.), anomaly detection offers a way to have an alarm, indicating that something is going wrong and has the potential to offer the user the chance to run more appropriate (usually costly) procedures in order to pinpoint the fault.

Thus, a number of approaches have been proposed over the years to tackle the problem of anomaly detection, which is strongly related to the notions of novelty detection and outlier detection [17]-[19]. Among the different approaches a one-class formulation [20] within an ensemble scheme [21] has been proven quite successful [6] for bearing fault detection using however a more computational demanding feature extraction stage.

In this work following the promising results of the ensemble detector presented in [6], three simple one-class classifiers are combined for the final detection: a Principal Component Analysis (PCA) anomaly detector, a Nearest Neighbor (NN) anomaly detector and a Gaussian anomaly detector. The combination is performed through a standard majority-voting scheme (at least two out of three detectors should detect a faulty condition). In the following paragraphs the underlying assumptions made by each one of the detectors are presented in brief. All detectors were implemented using the Data Description Toolbox [22].

Nearest Neighbor Anomaly Detector

The NN anomaly detector assumes that normal data appear in dense clusters, while anomalies appear at a distance quite far apart from those clusters. More specifically, having a new data point \mathbf{x}_{new} , it is decided whether it corresponds to an anomaly or

not, based on the distance to its nearest neighbor \mathbf{y} , after normalizing it by the distance of \mathbf{y} to its nearest neighbor \mathbf{z} :

$$D = \frac{\|\mathbf{x}_{new} - \mathbf{y}\|}{\|\mathbf{y} - \mathbf{z}\|} \quad (4)$$

In most of the cases the distance is measured using the Euclidian norm.

Gaussian anomaly detector

The Gaussian anomaly detector assumes a Gaussian distribution of the normal class with a probability density function given by:

$$p(\mathbf{x}; \boldsymbol{\mu}, \mathbf{C}) = \frac{1}{(2\pi)^{d/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (5)$$

where d is the dimension of the input space, $\boldsymbol{\mu}$ is the mean value and \mathbf{C} is the covariance matrix, which are estimated based on the training data.

Using a threshold for p , (eq. 5), data that are quite away from the center are declared as anomalies. The threshold can be computed analytically given a desired false negative rate [20]. However in most cases it is calculated using the available data.

PCA Anomaly Detector

PCA anomaly detector uses the reconstruction error as a means to detect anomalies, where the reconstruction error is defined as:

$$e_r = \|\mathbf{x}_{new} - (\mathbf{W}\mathbf{W}^T)\mathbf{x}_{new}\|^2 \quad (6)$$

where \mathbf{W} is a $d \times l$ matrix, whose columns are the eigenvectors corresponding to the l largest eigenvalues of the covariance matrix \mathbf{C} as in conventional PCA [23], [24].

C. Fault Diagnosis

In the case of sufficient training data (data covering all potential fault conditions), a data driven condition monitoring process can move a step further from the simple anomaly detection scheme and build a diagnostic system following a standard pattern classification approach, with a plethora of classification methods that could be used in this case. However taken into account the structure of the feature space, as it is revealed in Fig. 3, it seems that different classes form quite concrete clusters. Therefore a simple NN method seems a viable candidate.

The NN classifier is a very simple nonlinear classifier. It is based on the general principle that “nature does not make jumps” (“Natura non facit saltus”) [25]. This means that data belonging to the same class it is highly likely to have similar attributes (feature values). Similarity is usually measured using the Euclidian distance. Therefore given a labeled training set, the NN classifier assigns a new unseen example to the class of its most similar (nearest) neighbor [26].

III. EXPERIMENTAL VALIDATION

A. Test Bed

For testing the utility of the system the well-known data set from Case Western Reserve University [27] is used, which contains seeded bearing faults, introduced using electro-discharge machining with various fault diameters. Only faulty conditions at the drive end were tested, for which measurements from both the drive end and the fan end sensors are available. Furthermore all the outer race faults are treated as belonging to one super-class of outer race faults (three different fault locations were actually engaged during the experiment). The experimental set-up is shown in Fig. 4.

From each experiment a number of data points are extracted using a sliding window without overlapping of duration equal to approximately 20 revolutions. The sampling frequency was 12kHz.

B. Detection setting

For the detection setting, normal condition data produced from different loading conditions along with data coming from three faulty conditions with different loadings and with the smallest available fault size (0.007 inches) are used. By this setting, the efficiency of the method for detecting what could be considered an “incipient” fault (a fault with size as small as 0.007 inches) is put to test.

In order to estimate the performance of the proposed method, 10 fold cross validation (CV) procedure is used, which is repeated 10 times (10x10 CV) each time randomly reshuffling the available data [28]. This means that 9/10 of normal data samples are used for training and 1/10 for testing for each fold. The results are summarized in the following section.

C. Diagnosis setting

For testing the diagnostic capabilities of the proposed method, a slightly modified set of tests than those proposed in [7] is performed as it follows.

Test A: a four class diagnostic problem considering only incipient faults. Practically the same data, as for the case of detection described above, using however during the training process apart from the normal-class data, also data belonging to the faulty classes. In other words the learning method (the NN classifier) is given information not only about the normal operating condition, but also about the faulty operating conditions, for the minimum size of the seeded fault (0.007 inches). In order to get as unbiased an estimate of the performance of the algorithm as possible 10x10 CV is employed (the standard 10 fold CV procedure is repeated 10 times after reshuffling of the data and the results are summed across the 10 repetitions).

Test B: a ten class diagnostic problem: In this setting apart from the normal condition class, there are nine more classes: ball fault of size 0.007, ball fault of size 0.014, ball fault of size 0.021, inner race fault of size 0.007, inner race fault of size 0.014, inner race fault of size 0.021, outer race fault of size 0.014 and finally inner race fault

of size 0.021. In other words for each one of the main three fault categories (ball fault, inner race fault, outer race fault) three subcategories are considered based on the dimension (severity) of the fault. The same procedure (10x10 CV) as in the previous setting is employed for estimating the performance.

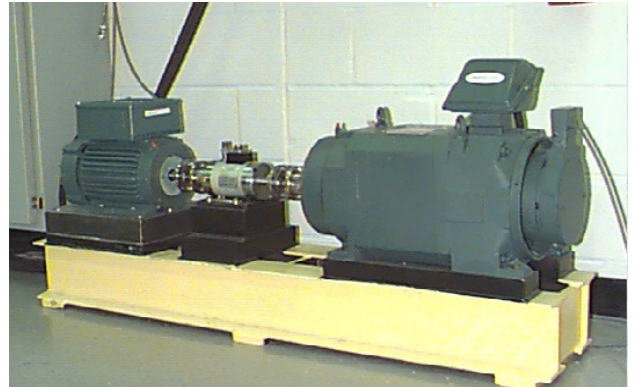


Fig. 4. The experimental test bed (courtesy of Professor K. Loparo).

IV. RESULTS

The results of the achieved detection and diagnostic performances are summarized in Tables I-IV and V, VI respectively.

Tables I-IV shows the aggregated confusion matrices (CF) of the three detectors as well as the ensemble detector (Note: a CF is a table containing information about the actual and the predicted classes as these have been assigned by a classification or detection system. All performance measures are derived by manipulating the different elements of this matrix. For example the overall accuracy equals the sum of the diagonal elements of the matrix divided by the sum of all elements of the matrix. Usually, the columns are labeled with the predicted classes and the rows with the actual/true classes even though the transpose is also encountered in the literature. Therefore, each entry $CF(i, j)$ in the matrix denotes the number of cases that actually belong to class i and have been assigned to class j by the classification algorithm. The better the classifier the less non-zero off-diagonal elements exist, with the perfect classification system corresponding to a diagonal matrix. As it can be seen even without the ensemble scheme the detection accuracy it is very high (the off diagonal elements are a very small fraction of the total number of cases) with the detectors not missing any faulty conditions and producing only very few false alarms. The use of the ensemble scheme eliminates false alarms further improving the reliability of the detector.

Table V presents the aggregated CF for Test A. As it can be seen this is a diagonal matrix indicating that the classifier performs optimally. Test B is difficult to be represented in the form of a confusion matrix due to the large number of elements. Therefore, instead of the CF, the overall classification accuracy is summarized in Table VII along with the accuracy of Test A for completeness.

V. CONCLUSIONS

In this article, a very simple yet quite effective feature extraction scheme for bearing fault detection and diagnosis is proposed. The method relies on the fusion of the information coming from two sensors through the estimation of the covariance matrix of the two signals, an approach which is probably among the simplest ones for fusing information of two different sources (sensors). The different classes-health conditions have been well characterized by this feature set for this specific test case. This can be also visually verified due to the three dimensional nature of the feature space as it is depicted in Fig. 3.

The extracted features are combined with very simple detection and diagnostic algorithms with very promising results. The ensemble scheme is capable of further increasing the detection accuracy of the individual detectors. Nevertheless, further experimentation is needed before safer conclusions can be drawn, since the method was tested only under seeded fault scenarios.

In future work we will try to test the effectiveness of the method in real life long degradation testings. Apart from that, as it was mentioned, the proposed scheme is probably the simplest data fusion approach that can be established. Therefore, in future work other features/descriptors that can be extracted from the "cloud points" will be tested along with the possibility to use embedding to alleviate the need for more than one sensors.

TABLE I. NN DETECTOR

Detection/Confusion Matrix		Estimated class	
		Normal	Faulty
True class	Normal	4110	11
	Faulty	0	152900

TABLE II. GAUSSIAN DETECTOR

Detection/Confusion Matrix		Estimated class	
		Normal	Faulty
True class	Normal	4109	10
	Faulty	0	152900

TABLE III. PCA DETECTOR

Detection/Confusion Matrix		Estimated class	
		Normal	Faulty
True class	Normal	4105	15
	Faulty	0	152900

TABLE IV. ENSEMBLE DETECTOR

Detection/Confusion Matrix		Estimated class	
		Normal	Faulty
True class	Normal	4120	0
	Faulty	0	152900

TABLE V. CONFUSION MATRIX FOR TEST A

Confusion Matrix		Estimated class			
		Normal	Ball	Inner	Outer
True class	Normal	4120	0	0	0
	Ball	0	1170	0	0
	Inner	0	0	1180	0
	Outer	0	0	0	3530

TABLE VI. OVERALL ACCURACY

	Test A	Test C
Accuracy	100%	98.91%

ACKNOWLEDGMENT

The authors would like to thank Professor Kenneth Loparo for granting permission to use Fig. 4 and also for its general contribution to the field of bearing fault detection by making available the seeded fault bearing data set [27].

This work was partially supported by the Horizon 2020 Framework program DISIRE under the Grant Agreement 636834.

REFERENCES

- [1] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, and B. Wu, *Intelligent Fault Diagnosis And Prognosis For Engineering Systems*, John Wiley & Sons 2006.
- [2] P.D. McFadden, and J.D. Smith, "Vibration monitoring of rolling element bearings by the high-frequency resonance technique-a review," *Tribology International*, vol. 17, no. 1, pp.3-10, 1984.
- [3] A.M. Al-Ghamd, and D. Mba, "A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp.1537-1571, 2006.
- [4] C. Lanham, "Understanding the tests that are recommended for electric motor predictive maintenance," Baker Instrument Company, 2006.
- [5] 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 16th Mediterranean Conference on Control and Automation*, 2008, Greece, pp. 1792-1797.
- [6] G. Georgoulas, T. Loutas, C.D. Stylios, and V. Kostopoulos, "Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition," *Mechanical Systems and Signal Processing*, vol. 41, no. 1, pp. 510-525, 2013.
- [7] Y. Lei, Z. He, Y. Zi, and Q. Hu, "Fault diagnosis of rotating machinery based on multiple ANFIS combination with Gas," *Mechanical Systems and Signal Processing*, vol. 21, no. 5, pp. 2280-2294, 2007.
- [8] M. O. Mustafa, G. Georgoulas, and G. Nikolakopoulos, "Bearing fault classification based on Minimum Volume Ellipsoid feature extraction." In *proceedings of IEEE International Conference on Control Applications (CCA)*, 2013, pp. 1177-1182.
- [9] G. Georgoulas, P. Karvelis, T. Loutas, and C. D. Stylios, "Rolling element bearings diagnostics using the Symbolic Aggregate approximation," *Mechanical Systems and Signal Processing*, vol 60-61, pp. 229-242, 2015.
- [10] T. H. Loutas, D. Roulias, and G. Georgoulas, "Remaining Useful Life Estimation in Rolling Bearings Utilizing Data-Driven Probabilistic E-Support Vectors Regression," *IEEE Transactions on Reliability*, vol. 62, no. 4, pp. 821-832, 2013.
- [11] A.K. Mahamad, and T.Hiyama, "Development of Artificial Neural Network Based Fault Diagnosis of Induction Motor Bearing," in *Proceedings of 2nd IEEE International Conference on Power and Energy PECon*, Johor Bahru, 2008, pp. 1387-1392.
- [12] S. Abbasion, A. Rafsanjani, A. Farshidianfar, and N. Irani, "Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine," *Mechanical Systems and Signal Processing*, vol. 21, no. 7, pp. 2933-2945, 2007.
- [13] Y. Lei, J. Lin, Z. He, and M. J. Zuo, "A review on empirical mode decomposition in fault diagnosis of rotating machinery," *Mechanical Systems and Signal Processing*, vol. 35, no. 1, pp. 108-126, 2013.
- [14] C. S. Park, Y. C. Choi, and Y. H. Kim, "Early fault detection in automotive ball bearings using the minimum variance cepstrum," *Mechanical Systems and Signal Processing*, vol. 38, no. 2, pp. 534-548, 2013.

- [15] D. Wang, P. W. Tse, and K. L. Tsui, "An enhanced Kurtogram method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 35, no. 1, pp. 176-199, 2013.
- [16] Y. Qian, R. Yan, and M. Shan, "Damage assessment of mechanical systems based on recurrence quantification analysis," *In proceedings IEEE Conference on Prognostics and System Health Management (PHM)*, 2012, pp. 1-5.
- [17] M. Markou, and S. Singh, "Novelty detection: a review-part 1: statistical approaches," *Signal Processing*, vol. 83, no. 12, pp. 2481-2497, 2003.
- [18] M. Markou, S. Singh, "Novelty detection: a review-part 2: neural network based approaches," *Signal Processing*, vol. 83, no. 12, pp. 2499-2521, 2003.
- [19] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1-58, 2009.
- [20] D. M. J. Tax, "One-class classification; concept-learning in the absence of counter-examples," Ph.D. thesis, Delft University of Technology, 2001.
- [21] L. Rokach, "Ensembled-based classifiers," *Artificial Intelligence Review*, vol. 33, pp. 1-39, 2010.
- [22] The Data Description Toolbox, <http://homepage.tudelft.nl/n9d04/dd_tools.html>.
- [23] I. Jolliffe, *Principal component analysis*. John Wiley & Sons, 2005.
- [24] G. Georgoulas, M.O. Mustafa, I.P. Tsoumas, J.A. Antonino-Daviu, V. Climente-Alarcon, C.D. Stylios, and G. Nikolakopoulos, "Principal Component Analysis of the start-up transient and Hidden Markov Modeling for broken rotor bar fault diagnosis in asynchronous machines," *Expert Systems with Applications*, vol. 40, no. 17, pp.7024-7033, 2013.
- [25] R. Battiti and M. Brunato, *The LION way. Machine Learning Plus Intelligent Optimization*, 2013.
- [26] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*. 4th ed, Academic Press, 2009.
- [27] K. A. Loparo, Bearing Vibration Dataset, Case Western Reserve University <<http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website>> (last accessed 13.01.2014).
- [28] N. Japkowicz, N. and M. Shah, *Evaluating learning algorithms: a classification perspective*, Cambridge University Press, 2011