University of Huddersfield Repository

Smith, Ann, Gu, Fengshou and Ball, Andrew

Maintaining model efficiency, avoiding bias and reducing input parameter volume in compressor fault classification

Original Citation

Smith, Ann, Gu, Fengshou and Ball, Andrew (2016) Maintaining model efficiency, avoiding bias and reducing input parameter volume in compressor fault classification. In: 2016 7th International Conference on Mechanical and Aerospace Engineering (ICMAE). IEEE, pp. 196-201. ISBN 978-1467388290

This version is available at http://eprints.hud.ac.uk/id/eprint/29770/

The University Repository is a digital collection of the research output of the University, available on Open Access. Copyright and Moral Rights for the items on this site are retained by the individual author and/or other copyright owners. Users may access full items free of charge; copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational or not-for-profit purposes without prior permission or charge, provided:

- The authors, title and full bibliographic details is credited in any copy;
- A hyperlink and/or URL is included for the original metadata page; and
- The content is not changed in any way.

For more information, including our policy and submission procedure, please contact the Repository Team at: E.mailbox@hud.ac.uk.

http://eprints.hud.ac.uk/

Maintaining Model Efficiency, Avoiding Bias and Reducing Input Parameter Volume in Compressor Fault Classification.

Ann Smith, Fengshou Gu, Andrew Ball
Condition and Performance Monitoring Research Group,
University of Huddersfield, HD1 3DH,
United Kingdom
a.smith@hud.ac.uk, F.Gu@hud.ac.uk, A.Ball@hud.ac.uk

***Abstract*— With the exponential growth in data collection and storage and the necessity for timely prognostic health monitoring of industrial processes traditional methods of data analysis are becoming redundant. Big data sets and huge volumes of inputs give rise to equally massive computational requirements. In this paper the differences in input parameter selection using a subset of the original variables and using data reduction techniques are compared. Each selection procedure being illustrated by both statistical methods and machine learning techniques. It is shown that the subsequent classification models are highly comparable. Finally the merits of a combined multivariate statistical and wavelet decomposition approach is considered. Techniques are applied to output signals from an experimental compressor rig.**

Keywords- Fault classification; Big data; Data compression; MSPCA

I. INTRODUCTION

As the volume of data continues to increase so do the complexities and relationships within the data. A globalised optimisation goal is necessary to avoid localised bias. Computational savings through reduction of input parameters and subsequent reduction in model complexity is essential [1]. Selecting a representative sample is also critical in avoiding bias and ensuring precision and accuracy of estimates. In this paper the problems of large data volumes are explored using the example of an industrial compressor [2, 3].

Compressors are an intrinsic part of many industrial processes whose performance and efficiency rely on early detection of compressor component deterioration. System faults can reduce performance levels and increase energy consumption in addition to potential machine damage and eventual shut-down. For example, compressors incorporated in water jet cutting systems are required to perform at the highest level in producing and maintaining the extremely high internal pressures necessary to force the water through small orifices producing exit velocities in excess of 2000 mph. Condition monitoring (CM) focuses continually on the health of a process ensuring near optimal performance for the duration of operation whilst enabling timely detection and identification of faults. Since the process is continually monitored informative non-intrusive measurements such as vibration signals externally captured at strategic points are invaluable. In this paper it is shown that the power of prognosis is further amplified by strategic selection of model input parameters and rigorous model construction.

Prior research [4] has shown that features extracted from the envelope spectra of vibration signals in the frequency domain have superior deterministic properties over their time domain equivalents in condition monitoring. Amplitudes of the harmonics being specific to process condition with a greater amplitude or displaced amplitude implying presence of a fault. Envelope spectra show only the amplitude profile of original signals and so provide a clearer insight into the underlying behaviour of processes having extraneous noise removed.

Cluster analysis (CA) [1, 5] was demonstrated to create homogeneous groups of variables. Clusters being formed in such a way that objects in the same cluster are very similar and objects in different clusters are distinct. CA gives a clear measure of variable properties whereas relevance vector machines (RVM) and support vector machines (SVM) do not.

Other variable clustering methods follow similar selection criteria. Evolutionary Genetic Algorithms (GA) or Particle Swarm Optimisation (PSO) are both population based search methods inspired by observation of the collaborative behaviour of biological populations such as birds or bees. Specifically these populations are seen to demonstrate a collective intelligence. [6] PSO has been shown to deliver comparable results to GA with significant gains in reduced computational time except for particularly complex cases. However, the majority of GA, PSO and Tipping's algorithms [7, 8] are restricted to 10 to 15 input parameters for convergence within reasonable time constraints

The aim of this paper is to compare the efficiency of models constructed using a reduced number of pre-selected input parameters to models using variable reduction techniques. Classification models constructed using both multivariate statistical methods and machine learning techniques are assessed their efficacy measured by successful classification rates. Comparisons are then drawn with multiscale PCA (MSPCA) which combines multivariate statistical methods and wavelet decomposition [9].

II. DATA ACQUISITION

Output signals for the second stage vibration measurements were collected from an accelerometer attached to the exterior of the second stage cylinder on a two-stage, single-acting Broom Wade TS9 reciprocating compressor (RC) which has two cylinders in the form of a "V". The RC was operated under healthy conditions and with four independently seeded faults (suction valve leakage (SVL), discharge valve leakage (DVL), intercooler leakage (ICL) and loose drive belt (LB)), each condition (class) being repeated 24 times. The amplitude and frequency of the first 32 envelope harmonics for the demodulated second stage vibration signal were stored for each of the 120 observations across the 5 classes.

From the vibration envelope spectrum a heterogeneous group of input parameters was selected by multivariate CA, as described in previous work. Both a multivariate statistical classifier (Discriminant analysis) and a machine learning classifier (Naïve Bayes) were constructed and their efficiency in terms of classification success rates compared.

III. MULTIVARIATE CLASSIFIERS USING REDUCED NUMBERS OF INPUT PARAMETERS

A. *Discriminant Analysis*

Discriminant analysis (DA) is a statistical method used in pattern recognition and machine learning whereby a combination of characteristic features is established with the aim of separating two or more classes or events. Categorical dependent variables are predicted by their scores on a discriminant function established using one or more continuous or binary, independent variable(s). Training data is used to estimate the parameters of the discriminant functions of the predictor variables. [1, 2].

For a set of observations $\vec{x}$ on each sample of an event with known class y in the training data set a good predictor for class y from any similar sample is sought given any further observation x. For the two class case an observation with log likelihood ratio greater than a threshold T is predicted to belong to the first class assuming the conditional probability density functions $p(\vec{x} \mid y = 0)$ and $p(\vec{x} \mid y = 1)$ are normally distributed with means μ_0, μ_1 and covariances Σ_0, Σ_1 respectively. Observations being predicted to belong to the second class if the log of the likelihood ratios are below the threshold T. The quadratic discriminant classifier is given by

$$(\vec{x}-\vec{\mu}_0)^T \sum_0^{-1} (\vec{x}-\vec{\mu}_0) + \ln|\sum_0| - (\vec{x}-\vec{\mu}_1)^T \sum_0^{-1} (\vec{x}-\vec{\mu}_1) - \ln|\sum_1| > T \quad (1)$$

If homoscedacity can be assumed then the class covariances can be assumed equal:

$$\sum_0 = \sum_1 = \sum$$

and the covariance matrices have full rank. Hence (1) simplifies to the decision criterion being based on the dot product $\vec{w}.\vec{x} > c$ for some threshold constant c

where $\vec{w} = \sum^{-1}(\vec{\mu}_1 - \vec{\mu}_0)$ and

$$c = \frac{1}{2}(T - \vec{\mu}_0{}^T \sum_0^{-1} \vec{\mu}_0 + \vec{\mu}_1{}^T \sum_1^{-1} \vec{\mu}_1$$

Thus the model is a function of a linear combination of the known observations.

100% classification rate was achieved in the DA model for the two group case 'healthy' and 'intercooler leakage' using just two input parameters, envelope harmonic features 4 and 6. Previously [4], although highly informative, envelope harmonic features 4 and 7 proved insufficient to fully separate these two groups. Thus confirming choice of input parameter is of paramount importance hence the value of pre-evaluation of characteristics for optimum selection.

B. *Naïve Bayes*

Naïve Bayes [10] is a relatively simple technique for constructing classifiers. Classification is based on estimating the conditional probability $p(C_k \mid x_1, ..., x_n)$ for n independent variables $\underline{x} = (x_1, ..., x_n)$

$$p(C_k \mid \underline{x}) = \frac{p(C_k)p(\underline{x} \mid C_k)}{p(\underline{x})}$$

Since the evidence, p(x), is not dependent on class and is effectively constant, under naïve conditional independence assumptions the probability model becomes

$$p(C_k \mid x_1, ...x_n) = \frac{1}{Z} p(C_k) \prod_{i-1}^{n} p(x_i \mid C_k)$$

Where the evidence Z = p(x) is a constant scaling factor dependent only on $x_1, ..., x_n$

The classifier based on this probability model, the (naïve) Bayes classifier is given by

$$\hat{y} = \frac{\arg\max}{k \in \{1, ..., k\}} p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

For some k that assigns the class label $\hat{y} = C_k$.

TABLE 1 CLASSIFICATION SUCCESS RATES PER NUMBER OF GROUPS AND PER MODEL

	2 groups (Healthy and ICL)	5 groups
DA using 2 input parameters [4, 6]	100%	
NB using 2 input parameter [4, 6]	94%	53%
NB using 5 input parameters [2, 7, 9, 12, 17]	100%	64%
NB using 15 input parameters [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]	n/a	80%

Classification success varies extensively depending on the number of groups considered and the number of parameters incorporated in the model Table 1. A two-dimensional DA gave perfect classification in the two group case (healthy and ICL) as did a 5 parameter NB model. However, considering all five classes simultaneously requires far greater model complexity and was not achieved using DA although a 15 parameter NB model achieved 80% success rate Fig. 1.

Again model complexity hence computational efforts are significantly reduced by prior evaluation and selection of a reduced number of heterogeneous input parameters to ensure maximum explanatory power across all classes. The NB classification tree established using two input parameters [4, 6] is illustrated in Fig.2 and provides a useful visual method of classifying any further samples. For example: A case with a parameter 4 amplitude of 0.8 and a parameter 6 amplitude of 2.2 (i.e. f4 =0.8 and f6 = 2.2) would be allocated to the healthy group after passing through six decision nodes. Classification success rates were calculated for a number of models using the features indicated through the cluster analysis, the highest classification rate achieved across all five classes was 80% using 15 input features [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16]. It should be noted that 15 parameters exceeds the maximum number permitted for many algorithms.

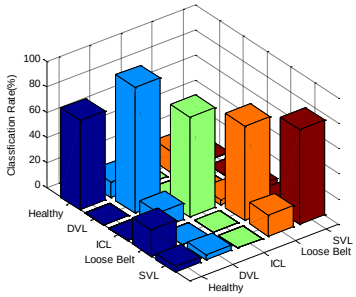

Fig. 1 Second Stage Vibration 15 Parameter Model with 80% Successful Classification Across All 5 Groups.

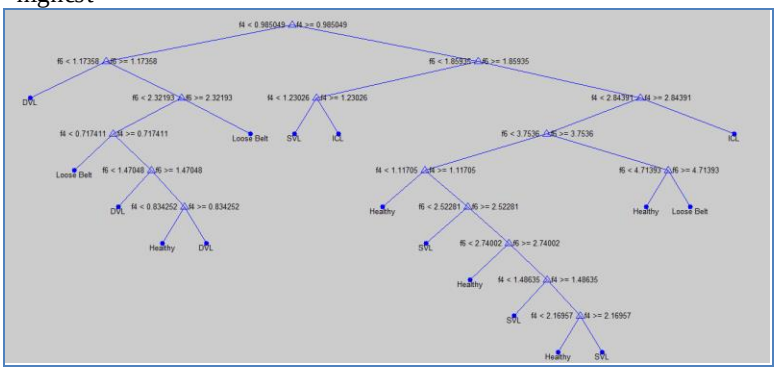

Fig. 2 Naïve Bayes Classification Tree Using Two Input Parameters: Envelope Features 4 and 6.

IV. MULTIVARIATE CLASSIFIERS USING VARIABLE REDUCTION TECHNIQUES

In contrast to methods using a reduced number of input parameters, variable reduction utilises all the original variables to generate a reduced set of new variables. These new variables are generally chosen to be orthogonal and if successful a small number will sufficiently explain a significant majority of the system variation. Again a multivariate statistical method, principal component analysis (PCA), and a machine learning method, support vector machines (SVM), are considered with respect to classification success rates and model dimensionality.

A. Principal Component Analysis

A set of uncorrelated principal components (PCs) is produced from the original correlated variables [1, 10,11]. The first PC, Z_1, accounts for the largest proportion of the variance in the sample; the second, Z_2, which is generally uncorrelated with the first the second highest, and so on. Initially there will be as many PCs as original variables and collectively they account for the total variance in the sample. The vast majority of the total variance is accounted for by the first few PCs and only a negligible amount by the rest hence these latter PCs can be dropped from further analysis so reducing the 'dimensionality' of the data set. PCA reveals the internal structure of the data in a way that best explains the variance in the data.

Applying PCA to the compressor vibration signal and extracting the first two PCs gives the results shown in Fig. 3 and Table 2. Clearly in the 2 dimensional case the SVL group is entirely separate from all other measurements with the lowest scores on both the 1st and 2nd principal components. Although less obviously separated the other classes form clear clusters separable using higher dimensional models. Principal component scores by class are summarised in the table for the first two PCs. 81% of the total variation in the system is accounted for by the first three PCs with ten PCs required to achieve 95% coverage.

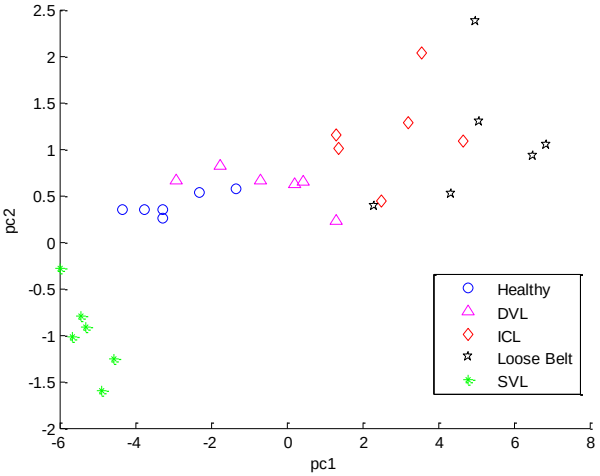

Fig. 3 Fault clustering using the first two principal components

TABLE 2 CLASS SCORES ON THE FIRST TWO PRINCIPAL COMPONENTS.

Class	Range of scores on principal component 1	Range of scores on principal component 2
Healthy	[-4.351, -1.326]	[0.262, 0.576]
SVL	[-5.980, -4.576]	[-1.595, -0.280]
DVL	[-2.926, 1.291]	[0.225, 0.818]
LB	[2.304, 6.850]	[0.394, 2.379]
ICL	[1.301, 4.677]	[0.445, 2.033]

B. Support Vector Machines

In machine learning, support vector machines (SVMs) [2, 12] are supervised learning models with associated learning algorithms that analyse data through pattern recognition, used for classification and regression analysis.

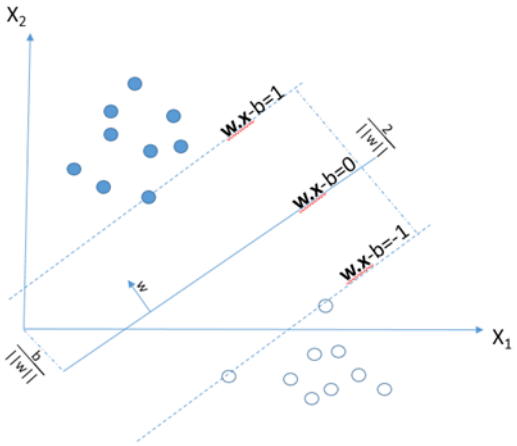

Fig. 4 SVM methodology illustrated.

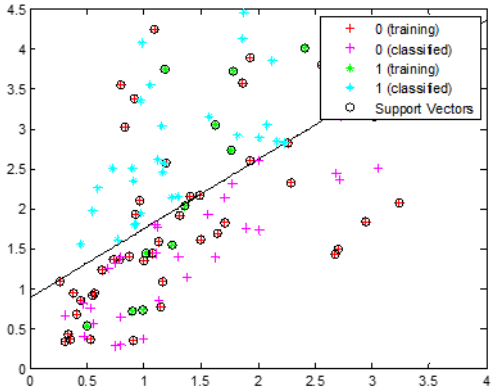

Fig. 5 SVM using Linear Kernel Function.

Given a set of training data D, a set of n points of the form $D = \{(x_i, y_i) \mid x_i \in \Re^p, y_i \in \{-1, 1\}\}_{i=1}^n$ where yi takes the value 1 or -1 indicating which class the point **xi** belongs to. Each **xi** is a p-dimensional real vector. The maximum-margin hyperplane dividing points with yi=1 from those with yi=-1 is given by the set of points x satisfying **w.x**-b = 0, Fig. 4. If the data are linearly separable, hyperplanes can be selected in such a way that they separate the data with no points between them. The region they bound, the margin is then maximised whilst ensuring no points are allowed to fall into it. The planes of the margin are given by **w.x** - b = 1 and **w.x** – b = -1. Samples falling on the margin are called the support vectors. The first class occupies the region **w.xi** – b $\geqslant$ 1 and the second class the region **w.xi** – b $\leqslant$ -1 jointly described as yi(**w.xi** – b) $\geqslant$ 1 for all 1 $\leqslant$ i $\leqslant$ n.

Utilising all 32 envelope harmonics of the second stage vibration signal, classification into 'healthy' and 'faulty' using a support vector machine (SVM) classifier gave 80% classification success Fig. 5, no clear linear separation is apparent between healthy and non-healthy cases.

V. VOLUME REDUCTION BY SIGNAL SIMPLIFICATION PRE MODELLING

Compression and simplification of signals offer alternative means of reducing data quantities. Signal compression which offers a crude de-noising technique in producing a smoothed simplified signal is the basis of multiscale PCA. A simplified multivariate signal being reconstructed using a simplified representation at each of a number of resolution levels. The most significant PCs being selected at each level.

A. Multiscale Principal Component Analysis

Multiscale principal component analysis (MSPCA) [13, 14, 15] combines the ability of PCA to produce a set of uncorrelated variables with that of wavelet analysis to extract deterministic features. The wavelet coefficients of the PCA are calculated at each scale and selected results are combined. Only those scales showing significant events are combined thus the process both de noises and simplifies the original multivariate signal. The technique is appropriate for modelling data with dynamic events due to its' multiscale nature, hence its' suitability for process fault detection. PCA captures the correlation and maximum variance between measurements and wavelet analysis captures the within measurement correlation. Thus both the variable correlation and the signal trend are accounted for by MSPCA. The complementary strengths of each procedure resulting in maximum information being extracted from complex multivariate measurements. The aim of multiscale PCA is to reconstruct a simplified multivariate signal, starting from an original multivariate signal and using a simple representation at each of a specified number of resolution levels. Multiscale principal components analysis generalises the PCA of a multivariate signal represented as a matrix by simultaneously performing a PCA on the matrices of details at different levels. A PCA is performed on the coarser approximation coefficients matrix in the wavelet domain as well as on the final reconstructed matrix. By selecting the numbers of retained principal components, interesting simplified signals can be reconstructed. Rules for retention of PCs are akin to those of PCA for example Kaiser's rule retains all PCs with eigenvalues greater than the mean eigenvalue i.e. those contributing greater than average explanatory power.

Signal compression was executed on the first four envelope harmonics, Fig. 6, and clearly illustrates the signal simplification achievable. The relative mean squared error for the first seven PCs is very good from a compression perspective all values being close to the maximum 100% i.e. 97.2383 98.3299 87.5594 93.8073 87.0964 92.6310 and 97.2287. Seven components were retained initially according to Kaiser's rule, $\lambda > \bar{\lambda}$. As expected, the rule keeps two principal components, both for the PCA approximations and the final PCA, but one principal component is kept for details at each of the five levels. Removal of the first three of these seven PCs which are primarily composed of noise with small contributions to the signal provides an effective albeit rather crude de-noising process.

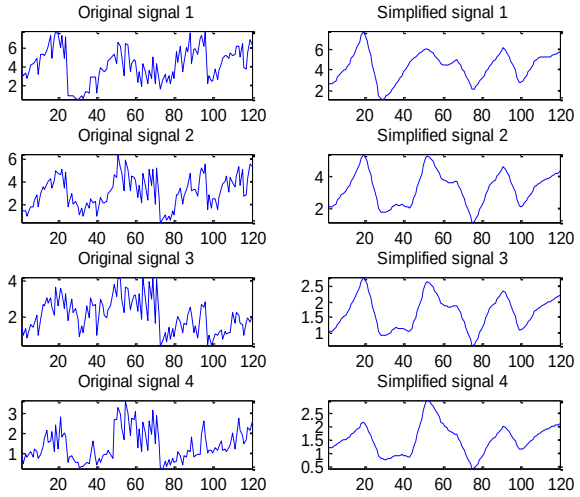

Fig. 6 Improved Multiscale PCA Signal Comparisons for the First Four Envelope Harmonics Across all 120 Cases.

VI. DISCUSSION AND CONCLUSIONS

The aim of this paper was to compare and contrast methods for optimal selection of input parameters and subsequent analysis in order to generate classifiers with improved fault classification accuracy.

Clustering the envelope harmonics into homogeneous groupings using Euclidean distance enabled a heterogeneous input parameter set to be selected which was shown to produce highly successful classifiers using both discriminant analysis and naïve Bayes methods. In terms of model efficiency using the second stage vibration envelope harmonics a 100% successful classification was achieved for a two group DA with just two input parameters whereas a five parameter model was required using NB. Also high classification rates were achieved across all five classes with a 15 parameter NB classifier. In contrast variable reduction techniques utilised all 32 harmonics to create a new set of input parameters. Models constructed using the technique of SVM realised 80% successful classification into the two groups 'healthy' and 'faulty'. PCA analysis also showed potential higher order modelling capabilities although only the inter-cooler leak class was completely separated from the other four classes in the two-dimensional model.

Modelling with a reduced set of input variables has been shown to achieve high efficiency with respect to classification success and has the additional advantage of preserving the underlying variable structure which is not the case for variable reduction methods.

Variable reduction techniques provide an alternative approach wherein all the original variables are reconstructed as a smaller number of principal components (PCs) or support vectors. Each of these being a weighted combination of all the original variables none of the original variables need be omitted from the analysis. Sufficient reduced variables are incorporated in models to give desired accuracy with respect to percentage of variation accounted for.

Undoubtedly there are advantages to both classical multivariate statistical methods and machine learning techniques equally both have their limitations particularly applied to large data sets with numerous classification groups. Gleaning the benefits of each and combining in a single compound analysis such as MSPCA can offer significant gains in both precision and input volume reduction. Likewise both preselection of input parameters and variable reduction techniques contribute significantly towards the construction of highly efficient, unbiased classifiers. If signals can be further reduced prior to consideration through compression further volume reductions are possible.

VII. REFERENCES

[1]. B. S. Everitt, G. Dunn (2001) 'Applied Multivariate Data Analysis', 2nd Edition. Arnold, London, ISBN 0 340 74122 8

[2]. B. S. Yang, W. W. Hwang, D. J. Kim, A. Chit Tan (2005) 'Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines.' , Mechanical Systems and Signal Processing, vol. 19, p. 371-390.

[3] Volk, W. (1969). Applied statistics for engineers. 2d ed. New York: McGraw-Hill.

[4]. M. Ahmed, A. Smith, F. Gu, A.D. Ball (2014) 'Fault Diagnosis of Reciprocating Compressors Using Relevance Vector Machines with A Genetic Algorithm Based on Vibration Data' Proceedings of the 20th International Conference on Automation & Computing'.

[5]. Manley, Bryan F. J. (1986) 'Multivariate Statistical Methods: A Primer', 3rd Edition. Chatfield and Colins ISBN 9781584884149.

[6]. Hassan, R., Cohanim, B. de Weck, O., Vente G. (2005) A comparison of particle swarm optimization and the genetic algorithm

[7]. M.E. Tipping (2001) 'Sparse Baysian learning and the relevance vector machine.' The journal of Machine Learning Research, vol. 1, pp. 211-244

[8]. M.E.Tipping, A. Faul, (2002) 'Analysis of sparse Baysian learning' Advances in neural information processing systems, vol. 14, pp. 383-389

[9]. Misra, M., Yue, H.H., Qin, S.J. & Ling, C. (2002), "Multivariate process monitoring and fault diagnosis by multi-scale PCA", Computers and Chemical Engineering, vol. 26, no. 9, pp. 1281-1293.

[10]. Smith, A., Gu, F., Ball, A. (2015) "Selection of Input Parameters for Multivariate Classifiers in Proactive Machine Health Monitoring by Clustering Envelope Spectrum Harmonics". Applied Mechanics and Materials, 798. pp. 308-313. ISSN 1662-7482

[11] Bro, Rasmus, and Age K. Smilde. (2014) "Principal component analysis." *Analytical Methods* 6.9: 2812-2831.

[12] Pan, Shuang, et al. (2016) "Fault Diagnosis System of Induction Motors Based on Multiscale Entropy and Support Vector Machine with Mutual Information Algorithm." *Shock and Vibration* 2016.

[13] Bakshi, B.R., (1998) Multiscale PCA with application to multivariate statistical process monitoring. American Institute of Chemical Engineers.AIChE Journal, 44(7), pp. 1596.

[14]. Wickerhauser, M.V. & Books24x7, I. (1994), Adapted wavelet analysis from theory to software, A K Peters, Natick, Mass;New York;.

[15] Marsland, S., (2009) Machine learning: an algorithmic perspective. Chapman & Hall/CRC machine learning & pattern recognition series