




Gearbox Fault Diagnosis Based on Mel-Frequency Cepstral Coefficients and Support Vector Machine

Tarak Benkedjough^(✉) , Taha Chettibi^(✉), Yassine Saadouni,
and Mohamed Afroun^(✉)

Ecole Militaire Polytechnique, Bordj El-Bahri, Alger, Algérie
bktarek@gmail.com, tahachettibi@yahoo.fr, yassine.saadouni@gmail.com,
mohamed.afroun@gmail.com

Abstract. The enhancement of the machine condition monitoring process is a key issue for reliability improvement. In fact, in order to produce quickly, economically, with high quality while decreasing the risk of production break due to a machine stop, it is necessary to maintain the equipment in a good operational condition. This requirement can be satisfied by implementing appropriate maintenance strategies such as Condition Based Maintenance (CBM) and using updated condition monitoring technologies for faults detection and classification. In this context, a new method for machinery condition monitoring based on Mel-Frequency Cepstral Coefficients (MFCCs) and Support Vector Machine (SVM) is proposed to automatically detect the mechanical faults by maximized the generalization ability. Hence, the purpose is to design an automatic detection system for mechanical components defects based on supervised classification by trained to maximize the margin. The proposed approach consists in a sequence of binary classifications after extracting a set of relevant features such as temporal indicators and MFCC coefficients. The diagnosis accuracy assessment is carried out by conducting various experiments on acceleration signals collected from a rotating machinery under different operating conditions.

Keywords: Diagnostics · Fault detection
Mel-Frequency Cepstral Coefficients · Support Vector Machines
Gearbox

1 Introduction

The ability to forecast machinery failure can help reducing maintenance costs, operation breakdowns and safety risks and is gaining importance in industry since it may limit the loss of production due to a machine stopping [1]. Fault diagnosis can be seen as a problem of pattern recognition for which several

artificial intelligence methods like hidden Markov Models (HMM) [2]; artificial neural network (ANN) [3] and support vector machines [4] have been applied. A challenging problem in rotating machinery diagnostic is how to construct and evaluate an effective feature sub-space from available features that can accurately represent the fault. Implementation difficulties of rotating machinery diagnostic systems are inherent to the random nature of defect growth by crack propagation in mechanical components, because each feature is effective for a defect at certain stage [5]. Yan et al. [6] provided a review on utilizing wavelets as a powerful tool for signal analysis with the purpose of rotary machines faults diagnosis. Lei et al. [7] provide a review of applying EMD to fault diagnosis of rotating machinery. In the review, all reported applications of EMD in fault diagnosis are divided into a few main aspects based on the key components of rotating machinery, namely, rolling element bearings, gears and rotors. Liu et al. [8] propose a novel fault diagnose method based on short-time matching and SVM to overcome the limitations of traditional sparse representation and fault diagnosis methods.

Condition monitoring based classifier has existed for some time, by using a variety of features, and artificial intelligence-based approaches to distinguish between fault and normal condition. The other problem is mainly associated with selecting a features set to allow the classifier discriminate between the classes without confusion. Nyanteh et al. [9] discusses the faults in rotating machines and describes a fault detection technique using artificial neural network (ANN) which is an expert system to detect short-circuit fault currents in the stator windings of a permanent-magnet synchronous machine (PMSM).

In this paper, we analyze the use of the SVM classifier [10]. This technique used for enhancing mechanical components fault diagnosis has been developed by fusion of multiple feature extraction through support vector machine. Particularly, we investigate how best to select features from the available data in order to maximize the performance of the classifier. Another main challenge for condition monitoring performance prognostics is how to construct and evaluate an effective feature sub-space from available features extraction, which can always represent the degradation state and how the performance of dimensionality reduction (DR) techniques may be improved; various techniques for the data reduction have been proposed [11]. Several features extraction techniques are used in signal recognition systems such linear prediction coefficients (LPC), linear predictive cepstral coefficients (LPCC), perceptual linear predictive analysis (PLP), and Mel-Frequency Spectrum Coefficients (MFCC) which is currently the most popular and it is discussed in this paper.

The main contribution of this paper is to use the MFCC and SVM. This approach is divided in two phases: (i) a features extraction phase by calculating the Mel-frequency cepstral coefficients and (ii) applying the SVM for data classification and visualization phase. The Support vector machine technique has been successfully applied in different applications such as in communication [12], financial time series [13] and biomedicine [14].

This paper is organized as follows. Section 2 presents the description of the proposed method. Section 3 presents the feature extraction based on MFCCs

technique. Section 4 describes the proposed method based on support vector machine for classification. Section 5 is dedicated to the experimental verification and results discussion and finally, Sect. 6 concludes the paper.

2 Description of the Proposed Method

Various conditions monitoring research works have been conducted for improving the performance classification. In Fig. 1, the three main steps of a generic condition based maintenance CBM process are indicated; namely: data acquisition, processing and maintenance decision making steps. Data acquisition step is intended to collect the data related to system health. Data processing phase is devoted to analyze the acquired data and finally, in the maintenance decision-making step, effective maintenance policies will be obtained based on information analysis.

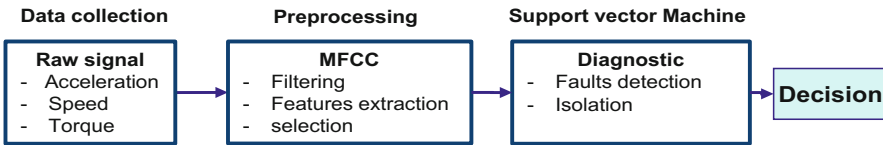


Fig. 1. Steps of a condition monitoring System.

3 Features Extraction Based on MFCC

In signal processing, The feature extraction is very important operation because the large data sets cause difficulties. Feature extraction using the MFCCs is widely known in speaker recognition. The MFCCs are commonly extracted from signals through cepstral analysis. Figure 2 shows the proposed steps of extraction of MFCCs from an raw signal. The input signal must first be broken up into small sections framed and windowed, these sections can be considered as stationary and exhibit stable characteristics. The Fourier transform is then taken and the magnitude of the resulting spectrum is warped by the Mel scale. The log of this spectrum is then taken and the DCT is applied [15].

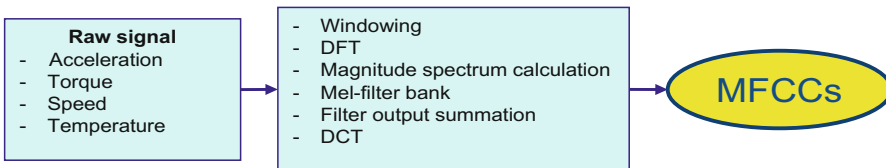


Fig. 2. Extraction of MFCCs from raw signals.

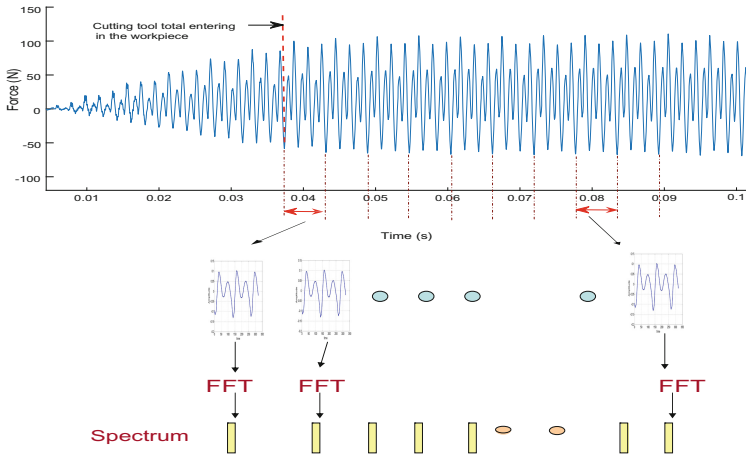


Fig. 3. Sequence of spectral vectors and time duration selection

The Input data is a raw signal in the time domain from different sensors (vibrations, force and acoustic emission) representation with duration in the order of 10 s (Fig. 3).

1. The first processing step is the computation of the frequency domain of (a windowed excerpt of) a signal. This is achieved by computing the Discrete Fourier Transform.
2. The second step is the computation of the mel-frequency spectrum. The powers of the spectrum obtained above onto the mel scale, using triangular overlapping windows.
3. The third step computes the logarithm of the signal; Take the logs of the powers at each of the mel frequencies.
4. The fourth step is to Take the discrete cosine transform of the list of mel log powers, as if it were a signal.
5. The fifth step tries to eliminate the information dependent characteristics by computing the cepstral coefficients. The MFCCs are the amplitudes of the resulting spectrum.

4 Data Classification by SVM

Support vector machine is a powerful technique for data classification [16]. SVM is developed from the optimal separation plane under linearly separable condition. Its basic principle can be illustrated in two-dimensional way as shown in Fig. 4.

Assume that a training set S is given by

$$S = \{x_i, y_i\}_{i=1}^n, \quad (1)$$

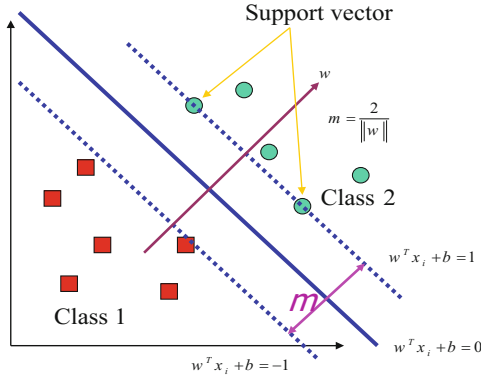


Fig. 4. Classification of data by using SVM

Where $x_i \in R^N$, and $y_i \in \{-1, +1\}$. The goal of SVM is to find an optimal hyperplane such that

$$\begin{cases} w^T x_i + b \geq 1 & \text{for } y_i = +1, \\ w^T x_i + b \leq -1 & \text{for } y_i = -1, \end{cases} \quad (2)$$

Where the weight vector $w \in R^N$, and the bias b is a scalar. If the inequality in Eq. 2 holds for all training data, it will be a linearity separable case. Therefore, in the linearity separable case, for finding the optimal hyperplane, one can solve the following constrained optimization problem:

Minimize

$$\Phi(w) = \frac{1}{2} w^T w \quad (3)$$

Subject to

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, n. \quad (4)$$

By introducing a set of Lagrange multipliers α_i, β_i for constraints 4, the problem becomes the one of finding the saddle point of the lagrangian. Thus, the dual problem becomes

Minimize

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (5)$$

Subject to

$$\sum_{i=1}^n \alpha_i y_i = 0, \quad (6)$$

$$0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n. \quad (7)$$

If $0 \leq \alpha_i \leq C$, the corresponding data points are called support vectors (SVs). SVMs map the input vector into a higher dimensional feature and thus can

solve the nonlinear case. By choosing a nonlinear mapping function $\varphi(x) \in R^M$, where $M \succ N$, the SVM can construct an optimal hyperplane in this new feature space. $K(x, x_i)$ is the inner product kernel performing the nonlinear mapping into feature space $K(x, x_i) = K(x_i, x) = \varphi(x)^T \varphi(x_i)$.

$$0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n. \tag{8}$$

Hence, the dual optimization problem becomes
 Minimize

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{9}$$

Subject to the same constraints as Eqs. 6 and 7, the only requirement on the kernel $K(x, x_i)$ is to satisfy the Mercer’s theorem [16]. Using Kernel functions, without treating the high dimensional data explicitly, unseen data are classified as follows:

$$x \in \begin{cases} \text{positive class,} & \text{if } g(x) \succ 0, \\ \text{negative class,} & \text{if } g(x) \prec 0, \end{cases} \tag{10}$$

Where the decision function is

$$g(x_i) = y_i \left(\sum_{j=1}^N y_j \alpha_j K(x_i, x_j) + b \right), \tag{11}$$

The other different functions kernel used are:

Table 1. Different function kernel used

Function	Polynomial	Gaussian	Sigmoid
$k(z, z_i)$	$(z \cdot z_i + 1)^d$	$\exp\left(\frac{z \cdot z_i}{\sigma^2}\right)^2$	$S[v(z \cdot z_i) + c]$

5 Results and Discussion

5.1 Experimental Setup

Figure 5 illustrates the test rig used to accomplish our experience and data collection. The shaft is driven by an electric motor and the rotation speed was variated between 0 and 6000 rpm. A radial load is added to the shaft and bearings. The bearings type MB Manufacturing ER-10K have 8 ball rollers in a single row, the pitch diameter is 33.5 mm, the roller element diameter is 7.93 mm and the contact angles of 0°. The measured signals consist of two acceleration signals given by an Endevco 6259M31 Accelerometer (10 mv/g, +/-1% error, Resonance >45 KHz) which is installed in input and output position on the gearbox

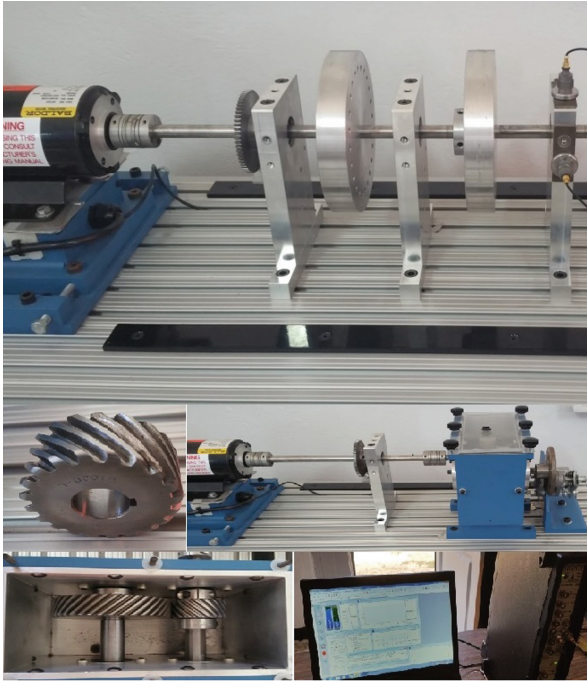


Fig. 5. Experimental setup.

housing. The data sampling rate was 66666.67 Samples per Second (200 KHz/3). The gearbox contains three shafts, 4 gears (the number of teeth is 32, 96, 48 and 80) and 6 bearings. The overall objective of the data was to specify the condition of each of the mechanical components and to specify the particular fault if it was not in a healthy state. The detail of the gearbox inside is shown in Fig. 5. A *B&K* high frequency accelerometer was mounted vertically on the housing of the test roller bearing to pick up the vertical acceleration. A filter with a cutoff frequency of 24 KHz was used to filter out the unwanted signals. Signals were then sent to the *B&K* 3560C Signal Analyzer. Readings were directly taken from the digital readout on the analyzer and a graphical representation of the data was displayed on the screen and the data were analyzed.

5.2 Experimental Verification

The diagram of the SVM method proposed for conditions monitoring is given in Fig. 6. The method is decomposed into two main steps. The first step is done off-line and aims at MFCCs generating and classification. When the SVM classifier is trained, the kernel function must be determined by user. The second step, which is achieved on-line, utilizes the trained data to predict the faults.

Figures 7 show the sensor measurements of the healthy and degraded state of the system (Acceleration) respectively.

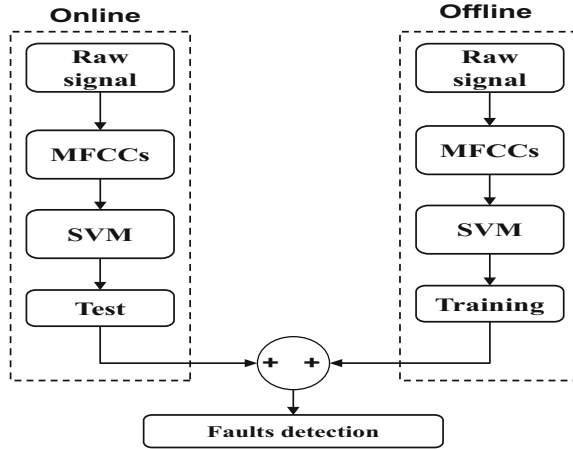


Fig. 6. Framework of the faults detection procedure

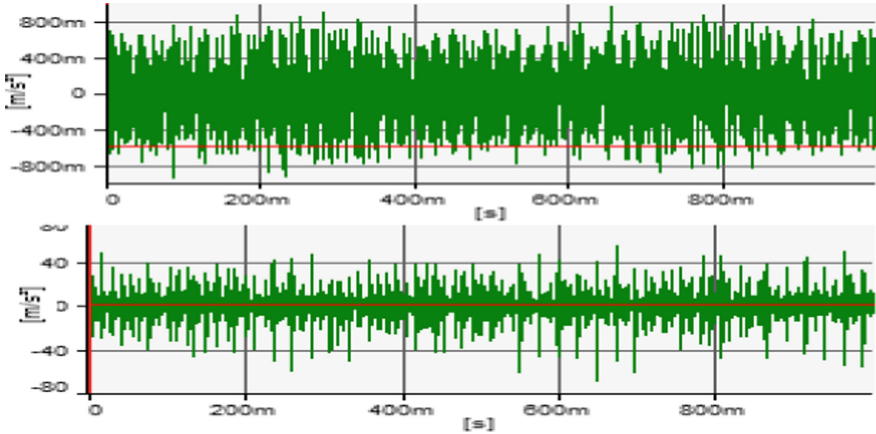


Fig. 7. Acceleration signals measurement of healthy system (top) and bearing defect (bottom) (speed 30 Hz)

We decompose the monitoring signals of each loading data above two conditions with MFCCs method for computing the feature extraction. It is noticed by signal analysis that the defect information of bearings and gears is mainly included in the first three MFCCs components. The above discussion deals with binary classification where the class labels can take only two values: +1 and -1. To find more than two classes in fault diagnosis of rotating machinery there are several fault classes such as bearing faults, gears broken, chipped, misalignment...etc. The different classes used in this paper are shown in Table 2.

Table 2. The different faults class

Case	Gear1	Gear 2	Bearing 1	Bearing 2	Shaft
1	Good	Good	Good	Good	Good
2	Chipped	Good	Good	Good	Good
3	Good	Good	Good	Inner	Eccentric
4	Good	Broken	Good	Inner	Good
5	Chipped	Broken	Outer	Good	Eccentric
6	Good	Broken	Outer	Good	Imbalance
7	Chipped	Good	Good	Inner	Good
8	Good	Good	Outer	Good	Imbalance

Broken (B); Chipped (C); Eccentric (E); Imbalance (I)

The total 13 features (16 signals for input and output) are calculated from 13 feature parameters of time domain. These parameters are MFCCs and the speed motor. The normal conditions of the system as $y = -1$ and the one with the defect as $y = +1$. The decision function $f(x)$ obtained by the linear kernel function and according to Eqs. (3) and (6) the parameters of classifier SVM, $\alpha = [0.0030, 0, 0.0056, 0, 0, 0, 0, 0, 0, 0.0126, 0, 0, 0, 0]^T$, $\omega = 0.1628$ and $b = 2.4856$. For gears defect, the parameters of the SVM classifier, $\alpha = [0.0070, 0, 0.0028, 0, 0, 0, 0, 0, 0, 0.0223, 0, 0, 0, 0]^T$, $\omega = 0.1421$ and $b = -3.4291$. It can be seen from Table 5 that SVM classifier based on MFCCs can still classify the three conditions of bearings (inner race defect, outer race defect and ball defect) which confirm fully that the SVM based MFCCs can be applied successfully to the faults recognition even in cases where only limited training samples are available.

For the gears faults identification with multiple-class (crack teeth, broken teeth and shipped ... etc.), generalizing method can be introduced to decompose the multiple-class problems into two-class problems which then can be trained with SVM.

In general, vibration signals of healthy bearings are Gaussian in distribution. The value of speed and load, therefore the value of the kurtosis is close to three for the vibration signals of a healthy system.

To select the optimal feature MFCCs that can well represent the condition of rotating machinery, a feature selection method based on the performance classification is shown in Tables 3 and 4.

Table 3. Motor speed influence for the classification

MFCCs+ speed			MFCCs		
Broken	Chipped	Eccentric	Broken	Chipped	Eccentric
100	99.285	100	100	97.857	100

Table 4. Window size influence for the classification

Window (ms)	MFCCs+ speed			Kurtosis+ speed		
	B	C	E	B	C	E
40	100	94.28	100	72.85	78.57	78.57
60	100	96.42	100	69.28	78.57	79.28
100	100	97.14	100	67.85	78.57	81.42
120	100	98.51	100	67.85	78.57	81.42
140	100	99.28	100	67.14	78.57	80.71

Table 5. Kernel used for classification

Kernel RBF $\sigma =$	Window (ms)	MFCCs+ speed			MFCCs+ RMS+ speed		
		B	C	E	B	C	E
(0.007)	100	100	100	99.285	100	100	99.285
(0.008)	120	100	100	99.285	100	100	99.285
(0.011)	140	100	100	99.285	100	100	99.285

The results shown in Table 3 compare the classification rate when including the motor speed as features with MFCCs. The classification ratio increases with the different kinds of faults. Note that the duration time of windowing equal to ($w = 140$ ms) and the kernel is RBF with ($\sigma = 0.002$).

In Table 4, classification process by SVM performed on the original feature (MFCCs) added the motor speed and compared with the fourth moment order (Kurtosis). The classification ratio of this process among 67.14% until 100%. The bad performance of this classification is due to the existence of irrelevant and useless features such as kurtosis.

Table 4 compares the classification rates for different windows size with different features used in this study by using the fourth moment order and the speed motor compared with MFCCs and speed. In this study, the RBF kernel are used as the basic kernel function of SVMs. The goal of this guideline is to identify optimal choice of the kernel parameter that the classifier can accurately classify the data input with a good classification rates.

In the specialized literature, no method is available for choosing the best kernel function. The most appropriate kernel function and the values of kernel function parameters (σ) for RBF. The selection of RBF kernel width is one of the major problems in SVMs for good performance of classification. For choosing the optimum values of the parameters (σ) of the RBF kernel, a large number of studies has been carried out by varying the values of parameters.

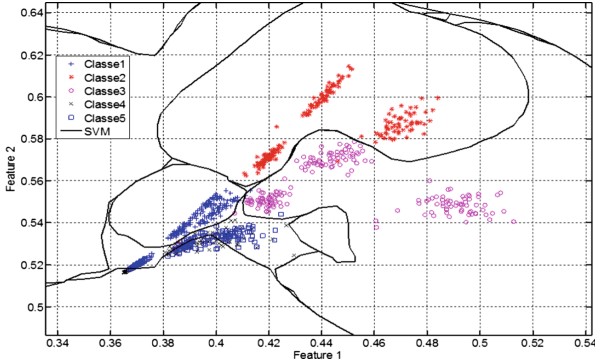


Fig. 8. Faults classification using RBF kernel

Table 5 compares the classification rates for different kernel function shown in Table 1. The Radial basis function (RBF) kernel gives a good classification results with a small number of the support vectors and learning time. The experiments are performed on three data sets with 60(%) training samples and 60(%) test samples (Fig. 8).

It is worth noting that the Gaussian kernel is the only kernel function used in our experiments. In fact, on each dataset we perform search for optimal combination of kernel width and the number of principal components for transformation. To speed up the search, we discard any eigenvector whose corresponding eigen value is smaller than 10^4 . To achieve this, the SVM based on MFCCs is proposed; as it is a very powerful tool that can determine a good classification of the system.

6 Conclusion

In this paper, we applied the combination of MFCCs and SVMs for intelligent fault diagnosis of rotating machinery. MFCCs were successfully applied for feature extraction step. However, the training feature using SVM is better than the other features such as kurtosis and the root mean square of signals. The feature extraction is an important step in fault diagnosis process. The proposed method were developed based on the acceleration signals measurements. In this paper; the potential of MFCCs-SVM has been highlighted for classification. Particularly, the simulation results of SVM classifier have verified that the proposed method has good efficiency in classifying eight types of defect with different characteristics. SVMs based MFCCs for multi-class classification is applied to the faults classification. The results show that SVMs achieved high performance in using multi-class classification strategy for one-against-all.

Acknowledgement. This research work is partially supported by the ATRST “Agence Thématique de Recherche en Science et Technologie” (Algeria). (Project code: 129/2016/P8/LMS).

References

1. Zio, E.: An Introduction to the Basics of Reliability and Risk Analysis, vol. 13. World Scientific, Singapore (2007)
2. Geramifard, O., Xu, J.-X., Panda, S.K.: Fault detection and diagnosis in synchronous motors using hidden markov model-based semi-nonparametric approach. *Eng. Appl. Artif. Intell.* **26**(8), 1919–1929 (2013)
3. Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufer, M., Verstockt, S., Van de Walle, R., Van Hoecke, S.: Convolutional neural network based fault detection for rotating machinery. *J. Sound Vib.* **377**, 331–345 (2016)
4. Jedliński, L., Jonak, J.: Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform. *Appl. Soft Comput.* **30**, 636–641 (2015)
5. Su, Z., Tang, B., Liu, Z., Qin, Y.: Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine. *Neurocomputing* **157**, 208–222 (2015)
6. Yan, R., Gao, R.X., Chen, X.: Wavelets for fault diagnosis of rotary machines: a review with applications. *Sig. Process.* **96**, 1–15 (2014)
7. Lei, Y., Lin, J., He, Z., Zuo, M.J.: A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mech. Syst. Signal Process.* **35**(1), 108–126 (2013)
8. Liu, R., Yang, B., Zhang, X., Wang, S., Chen, X.: Time-frequency atoms-driven support vector machine method for bearings incipient fault diagnosis. *Mech. Syst. Signal Process.* **75**, 345–370 (2016)
9. Nyanteh, Y., Edrington, C., Srivastava, S., Cartes, D.: Application of artificial intelligence to real-time fault detection in permanent-magnet synchronous machines. *IEEE Trans. Ind. Appl.* **49**(3), 1205–1214 (2013)
10. Saimurugan, M., Ramachandran, K.: A comparative study of sound and vibration signals in detection of rotating machine faults using support vector machine and independent component analysis. *Int. J. Data Anal. Techn. Strat.* **6**(2), 188–204 (2014)
11. Van der Maaten, L., Postma, E., Van Den Herik, H.: Dimensionality reduction: a comparative review. *J. Mach. Learn. Res.* **10**, 1–41 (2009)
12. Qian, Z.-L., Juan, D.-C., Bogdan, P., Tsui, C.-Y., Marculescu, D., Marculescu, R.: A support vector regression (SVR)-based latency model for network-on-chip (NoC) architectures. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **35**(3), 471–484 (2016)
13. Law, T., Shawe-Taylor, J.: Practical Bayesian support vector regression for financial time series prediction and market condition change detection. *Quant. Financ.* **17**(1), 1–14 (2017)
14. Du, W., Cheung, H., Johnson, C.A., Goldberg, I., Thambisetty, M., Becker, K.: A longitudinal support vector regression for prediction of ALS score. In: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), pp. 1586–1590. IEEE (2015)
15. Kinnunen, T., Li, H.: An overview of text-independent speaker recognition: from features to supervectors. *Speech Commun.* **52**(1), 12–40 (2010)
16. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)