



# Comparison Between Hidden Markov Models and Artificial Neural Networks in the Classification of Bearing Defects

Miloud Sedira<sup>(✉)</sup>, Ridha Ziani, and Ahmed Felkaoui

LMPA, Ferhat Abbas University, Setif, Algeria  
miloudsedira@yahoo.fr, ziani\_lmpa@yahoo.fr,  
a\_felkaoui@yahoo.fr

**Abstract.** In this paper a comparative study between two classification methods was presented, the first one belongs to the statistical domain in this case the Hidden Markov Models (HMM), the second is an Artificial Intelligence (AI) tool known as of Artificial Neural Networks (ANN), given their popularity in recent years and the interest shown by researchers in these methods, as to their performance and efficiency in the field of classification mainly. Indeed, the two classification tools were tested on data collected from vibratory signals on a test bench at the Bearing Data Center of Case Western Reserve University, and after being put in the appropriate form by an adequate signal processing and analysis to facilitate implementation. In this study, we have tried to identify the advantages and disadvantages of both tools in the field of classification of rotating machine defects, with the aim of accessing other work for the implementation of a classifier as effective as efficient. The results obtained are described as satisfactory and encouraging by their compatibility with those obtained by others implemented by other research but in other fields such as speech processing or image processing, which will give the character of originality to our work once completed.

**Keywords:** Bearing · Condition monitoring · Classification · HMM  
ANN · Signal processing

## 1 Introduction

The competitiveness of organizations and nations is directly affected by operational safety, the efficiency of maintenance costs and the availability of equipment conditional. Today, complex and developed production equipment requires highly sophisticated and costly maintenance strategies (Heng et al. 2009). The various demands and constraints imposed by productivity on the one hand, and the opportunities and solutions offered by the technology, notably computer technology and electronics, on the other, have fostered the continuous development of maintenance. From curative maintenance to immediate failure repair, to preventive maintenance based on scheduled maintenance operations, today arriving at predictive Maintenance (CBM) based on the anticipation and prediction of failures, resulting in a profit in time and cost and ensures benefit. Conditional Maintenance (CBM) consists of three phases: data acquisition,

data processing and decision-making. Diagnosis and prognosis are two important components of the CBM program (Jardine et al. 2006; Zhou et al. 2016; Kan et al. 2015). Vibration analysis has proven to be an effective means of diagnosing breakdowns. For decades, many methods have been proposed to extract and analyze the characteristics of the vibratory signal in order to perform the fault diagnosis (Wang et al. 2009). Considerable researches have been carried out on the development of various detection and diagnostic techniques based on vibratory data. These techniques can be classified mainly in the time domain, frequency domain, and time-frequency domain (Sanz et al. 2012). Increasingly sophisticated diagnostic methods have been used to determine the etiology of mechanical failure. Several diagnostic approaches are defined as, physical model, reliability model, machine learning model, and dependency model (Lee et al. 2014). The fault diagnosis is a procedure for mapping the information obtained in the measurement space and/or the characteristics in the characteristic space. This process is also called pattern recognition. Therefore, automatic pattern recognition is highly desirable. This can be achieved by classifying signals based on the information and/or characteristics extracted from the signals (Jardine et al. 2006).

Several defect classifiers based on characteristics extracted from vibratory signals, have been developed in this area. Recent approaches include K-nearest neighbors (KNN) the Gaussian mixture model (Wu et al. 2012), machine vector support (SVM) (Wu et al. 2012; Ziani et al. 2014; Sloin and Burshtein 2008; Zhu et al. 2009; Zhou et al. 2016; Li et al. 2015; Watanabe et al. 2013), and artificial neural network (Chen et al. 2014; Qiang et al. 2014; Wang et al. 2010; Sanz et al. 2012).

In this paper, a comparative study between two classifiers of different type is presented; the first one being an Artificial Intelligence (AI) tool, in this case Artificial Neural Networks (ANN), and the second is a probabilistic classifier consisting of Hidden Markov Models (HMM). These two tools have been used in the classification of ball bearing defects and then evaluate one with respect to the other, highlighting the advantages and disadvantages of one and the other. The following section presents the related works with our study.

## 2 Related Works

Several researchers around the world are working on the development of reliable and high-performance fault classifiers, with the aim of achieving diagnostic automation and subsequently the prognosis which is a primary objective of conditional maintenance (CBM). An important part of this research is focused on shape recognition tools of all types.

- (a) Many researchers have opted for statistic tools. Thus, using the Gaussian mixing models, (Wang et al. 2009) succeeded in performing a classification of the bearing defects and their degree of gravity, without resorting to the extraction of the characteristics of the vibratory signal, estimating the delay and the dimensional integration Of the time series, the vibration signal is reconstructed in the phase space and then the Gaussian mixture model (GMM) is established for each type of defect signal in the phase space. In the same context of the static tools,

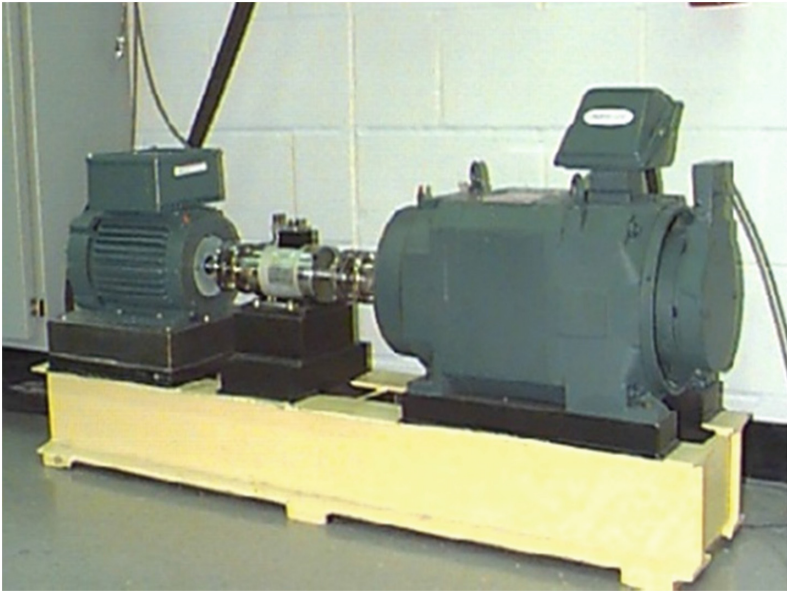
(Wang et al. 2009) used another model whose use in the field of mechanical diagnosis is very recent and remains relatively less explored, in this case the HMM, for the automatic diagnosis of bearing defects. The principle of this work is based on a given database established on the history of breakdowns listed in a catalog or dictionary. The HMM are then used to relate the data carried by a current signal with the established directory and to identify the defect by correspondence. The difficulty or inconvenience in this approach lies in the establishment of this catalog (Li et al. 2015). Also used the HMM for testing on simulation and real-life bearing fault diagnosis problem using the diversified gradient descent algorithm (DGD) to overcome the learning difficulties of HMM. This proposed formula does not require any particular form of objective function, it provides various estimates of parameters with different degrees of diversity, and it is obtained by dynamically adjusting the iterative procedure as a function of the gradient change of each parameter.

- (b) Other researchers have used the tools of the artificial intelligence (AI) as (Ziani et al. 2014), carried out by the machine support vectors (SVM) on the same database, using Fisher criterion for the selection of the indicators extracts of the vibratory signal, in the same context (Wu et al. 2012), also used the SVM on the same database to automate the diagnosis using the characteristics extracted from the vibration signal by an entropy technique Of multi-scale permutation (MSE), for the recognition of rolling defects (Chen et al. 2014), used an ANN for the automatic diagnosis of rolling faults in the aforementioned database by characterizing the defects by a dependent characteristic vector (DFV) to designate symptom attributes. The characteristic vector is derived from the classification of the characteristics on the basis of the evaluation of the Euclidean distance. Whereas, with the same ANN but with multiple weights to form a probabilistic neural network (PNN) to diagnose rolling defects in the database of case western reserve university. Wang et al. (2010) have presented a hybrid approach of coupled pulse neural networks (PCNN) with the probability neural network (PNN) in order to perform an automatic diagnosis of a hydraulic generator. The first network is used for extracting characteristics of the vibratory signal in the time domain, while the second network is used to classify the defects of the generator in question. A multi-layer perceptron (MLP) was used by Sanz et al. (2012), for the classification of gear defects, the characteristics extracted by a wavelet packet transformation were used as a level of input data to the network. The second level was fed by the meshing data of the torsional stiffness of the gear-carrying shaft. A multi-stage algorithm was applied to supervise learning of the neural network.
- (c) Another part of the researchers evaluated the classification tools by producing guides for users of the classification, like Miao et al. (2007), which presented a comparison of a statistical tool; Which is the HMM with an AI tool in this case the SVM, the advantages and disadvantages of one with respect to the other have been highlighted. This work is an important repertoire for researchers in this field. In the same way, we present in this paper, a comparison study between two classification tools, the first belongs to the field of statistics (HMM), the second to

the domain of the AI (ANN). In the following section, we present the experimental equipment and the protocol of work procedure.

### 3 Apparatus and Experimentation

The vibratory signals that constitute the database, which supported the application of our study, were obtained from an experiment carried out at the Bearing Data Center of Case Western Reserve University (Ziani et al. 2014), and made available to unrestricted access for users on the website (<http://csegroups.case.edu/bearingdatacenter/home>). The reliability of this data base is justified by the many works published which can be cited (Heng et al. 2009; Raj and Murali 2013; Georgoulas et al. 2015; Li et al. 2014; Liu et al. 2015; Chebil et al. 2011; Rodriguez et al. 2013; Ziani et al. 2014; Watanabe et al. 2013; Wu et al. 2012; Chen et al. 2014; Tian et al. 2015; Qiang et al. 2014; Li et al. 2015).

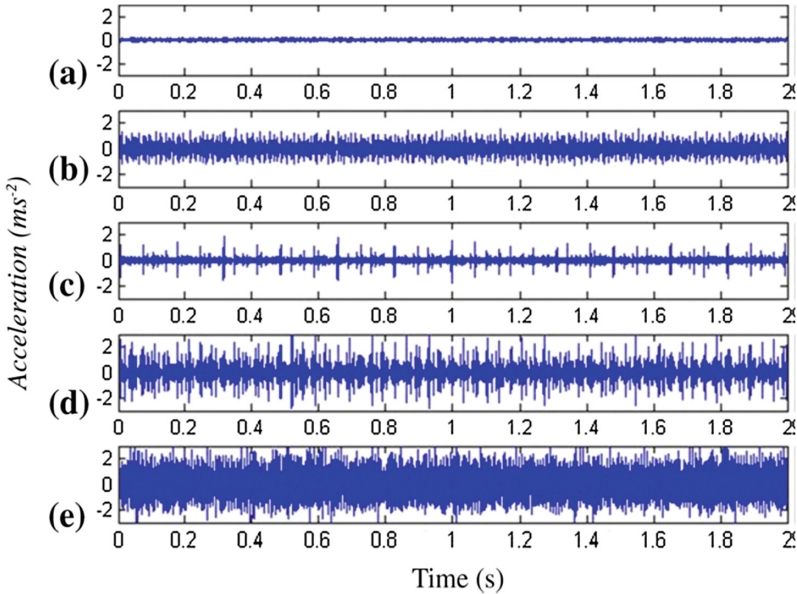


**Fig. 1.** Experimental setup

The experimental setup shown in Fig. 1 consists of a 2 hp motor (left), a torque sensor/encoder (center), a dynamometer (right) and an electronic control not shown). Single-point defects were introduced on the inner ring, outer ring and ball using electro-discharge machining with diameters of 0.007 in., 0.014 in., 0.021 in. and 0.028 in. For more details, we invite the reader to consult the website mentioned above. Vibratory signals were collected for normal bearings, and bearings with various defects

(0.007 in., 0.014 in., 0.021 in. and 0.028 in.) (Fig. 2). Data were collected at 12,000 Hz for the fan end and at 48,000 Hz for the drive end bearings.

For our study, we only looked at the health of the fan bearing at the inner race with different health states, that is to say; the normal state, state with defect of 0.007", state with defect of 0.014", state with defect of 0.021" and state with defect of 0.028". The most important for this context is to freeze the same environment for the two cases, then evaluate the reaction of each of the two methods. The following section explains how the significant features have been obtained from the vibratory signals.



**Fig. 2.** Time domain signals acquired under 2 hp motor load for normal and faulty bearing with inner race fault. (a) Normal, (b) Fault diameter of 0.007 in., (c) Fault diameter of 0.014 in., (d) Fault diameter of 0.021 in., (e) Fault diameter of 0.028 in.

## 4 Data Analysis and Features Extraction

In fault diagnosis of rotating machines, the signature of the fault is mainly contained in the temporal vibration signals, which can be also represented in three forms: time domain, frequency domain and time-frequency domain. To ensure better characterization of the signal, it was preferred to extract the most significant features from each domain. In this context, we used the features established in the framework of a previous work by Ziani et al. (2014) knowing that it is the same experimental framework, the objective of our study is the comparison between two classifiers notwithstanding the diagnosis and finally this work is closely related to Ziani et al. (2014). Thus, in the time domain, seven statistical features were established: mean, peak factor, asymmetry, kurtosis and the central statistical moments of order 5 to 7 (Ziani et al. 2014). In the

frequency domain, five (5) characteristics were calculated: the sum of the peaks of the power spectral density (PSD), calculated in 4 frequency bands centered on the fault frequency and its harmonics, and then in the total band. The third-order wavelet packet decomposition (WPD) in the time-frequency domain allowed us to calculate Kurtosis and energies for each coefficient; this allowed us to build the sets of learning and test data, which is illustrated in the following section.

### 5 Building Data Sets

Based on features extracted in the previous section, learning and test data sets were constructed for classification of bearing defects Table 1. The vibratory signals have been divided into 80 samples, for each rolling state or each class, there are 16 samples or sequences, divided into two groups of 8 samples each; One group reserved for learning and the other for the test. For a supervised classification, states or classes with their attributes have been designated as:

Classes or states: State 1, State 2, State 3, State 4, State 5, assigned attributes: Normal, default 0.007", default 0.014", default 0.021" and default 0.028".

**Table 1.** Data sets configuration

|         |                 | Features        |                 |                 |   |   |   |   |                   |
|---------|-----------------|-----------------|-----------------|-----------------|---|---|---|---|-------------------|
|         |                 | I <sub>11</sub> | I <sub>12</sub> | I <sub>13</sub> | . | . | . | . | I <sub>140</sub>  |
| Classes | C <sub>1</sub>  | .               | .               | .               | . | . | . | . | .                 |
|         | C <sub>2</sub>  | .               | .               | .               | . | . | . | . | .                 |
|         | .               | .               | .               | .               | . | . | . | . | .                 |
|         | C <sub>40</sub> | .               | .               | .               | . | . | . | . | I <sub>4040</sub> |
| .       | .               | .               | .               | .               | . | . | . | . |                   |
| .       | C <sub>80</sub> | .               | .               | .               | . | . | . | . | I <sub>8040</sub> |

Learning data

Test data

The classes in question with their attributes will serve as input data for the learning of the two classifiers, this is what will be illustrated in the next section.

## 6 Application of Hidden Markov Models and Artificial Neural Networks in the Classification of Bearing Defects

### 6.1 Theoretical Background

#### 6.1.1 HMM

In recent years, we have witnessed a growing interest in HMM, given their powerful ability to model time-varying signals in many real-world applications (Rabiner 1989; Kan et al. 2015; Li et al. 2015). An HMM is a statistical approach based on the Markov chain principle to model signals that evolve through a finite number of states. A Markov chain is a sequence of states where each depends only on the event that immediately precedes it (Kan et al. 2015). For more theoretical details see Rabiner (1989). An HMM is a doubly stochastic process, with an underlying process hidden from finite states, which associates itself with the observation process (Miao et al. 2007; Zhu et al. 2009; Lee et al. 2014). An HMM is characterized by: An initial state distribution vector  $\pi$ , a transition matrix  $A$ , an emission matrix  $B$  (1). An HMM is designated by:

$$\lambda = (A, B, \pi) \quad (1)$$

#### 6.1.2 ANN

ANN are a pattern recognition tool belonging to the concept of AI. The basic idea of this concept has been extracted from the functioning of the human brain. They have excellent learning and generalization skills nevertheless, they can deviate from complicated or imprecise data (Kan et al. 2015) because they are generally used for the classification of static inputs without sequential processing. They can also create their own representation of the received information. These exceptional features make ANN a powerful tool for modeling data (Kan et al. 2015). The minimization of the objective function, also called the cost function, is the main learning problem encountered in the ANN classification. One of the most used learning algorithms is back propagation (Pacheco et al. 2016; Sanz et al. 2012).

### 6.2 Application

#### 6.2.1 HMM

Typically, an HMM is assigned to each class and its parameters are estimated from a learning database using the maximum likelihood (ML) method. The recognition of an observed sequence which represents an unknown class can then be done by estimating the parameters of the HMM related to it, hence the unobservable state sequence, can be estimated by the observation sequence (Wang et al. 2010; Rabiner 1989; Zhou et al. 2016). The basis established in Sect. 4. contains 80 samples of 40 features for the five states defined above, the learning was done on 3 steps, the first with 10 features, the second with 20 features, and the third with all the Features i.e. 40 features. Indeed for

each class or state an HMM is computed, characterized by its, log-likelihood (LL), transition matrix and emission matrix. The results are summarized in Table 2.

$$LL = \log P(O|\lambda). \quad (2)$$

$\lambda$  is the HMM, O is the observation, LL is the log likelihood of HMM, P is the probability

**Table 2.** HMM Recognition performance

| HMM                         | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ | $\lambda_4$ | $\lambda_5$ |
|-----------------------------|-------------|-------------|-------------|-------------|-------------|
| State                       | 1           | 2           | 3           | 4           | 5           |
| Recognition performance (%) |             |             |             |             |             |
| 10 features                 | 95.75       | 97.66       | 98.28       | 97.08       | 96.93       |
| 20 features                 | 98.09       | 98.07       | 96.14       | 98.07       | 96.37       |
| 40 features                 | 98.23       | 94.47       | 93.96       | 93.62       | 98.23       |

### 6.2.2 ANN

A multilayer perceptron neural network (MLP) has been used in the present work to identify the different classes based on the severity of the defects. ML consist of an input layer of source nodes, one or more hidden layers of compute nodes and an output layer (Sanz et al. 2012). The same context observed for HMMs in Sect. 6.1.2 was applied for ANNs; the results of the classification are illustrated in Table 3.

**Table 3.** ANN Recognition performance

|             | ANN based classifier |
|-------------|----------------------|
|             | Test success (%)     |
| 10 features | 99.00                |
| 20 features | 97.00                |
| 40 features | 95.50                |

## 7 Comparison Between HMM and ANN Based Classifiers

The comparison is made first based on the results generated by each of the classifiers, secondly on the context of data manipulation and the computational computation with the difficulties encountered. On this aspect, the ANN offers more facilities for modeling and generalization in relation to HMM, but the latter remain robust to the variation of data in quantity and quality (nature and type of values). The results of this comparison are shown in Table 4.



**Table 4.** Performance comparison between HMM and ANN based classifiers

|             | HMM based classifier  | ANN based classifier |
|-------------|-----------------------|----------------------|
|             | Test success mean (%) | Test success (%)     |
| 10 features | 93.34                 | 99.00                |
| 20 features | 97.50                 | 97.00                |
| 40 features | 89.61                 | 95.50                |

## 8 Conclusion

A comparison between the hidden Markov models HMM and the artificial neural networks ANN is presented in this paper, for classification of ball bearing defects. This is a study of pattern recognition techniques for designing an efficient and robust classification system. Based on the results of the analysis, we found that the performance of the ANN decreases with the increase in the number of input data, while the HMM remain robust on this aspect. All the ANN are easier for computational modeling and manipulation in relation to HMM. This leads us to explore them deeply in future work.

**Acknowledgements.** This work is partially supported by the laboratory of applied precision mechanics (LMPA), Ferhat Abbas University, Setif, Algeria. The authors thank Professor K. LOPARO of case western university for providing the data. Also, the authors gratefully acknowledge the reviewers for their valuable comments and valuable suggestions, which greatly contributed to the improved presentation of this work.

## References

- Jardine, A.K.S., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **20**, 1483–1510 (2006). <https://doi.org/10.1016/j.ymssp.2005.09.012>
- Heng, A., Zhang, S., Tan, A.C.C., Mathew, J.: Rotating machinery prognostics: state of the art, challenges and opportunities. *Mech. Syst. Signal Process.* **23**, 724–739 (2009). <https://doi.org/10.1016/j.ymssp.2008.06.009>
- Sloin, A., Burshtein, D.: Support vector machine training for improved hidden markov modeling. *IEEE Trans. Signal Process.* **56**(1), 172–188 (2008)
- Wang, C., Zhou, J., Wang, Y., Huang, Z., Kou, P., Zhang, Y.: Hybrid neural network based fault diagnosis of rotating machinery. In: 2010 3rd International Congress on Image and Signal Processing (CISP2010). IEEE (2010). 978-1-4244-6516-3/10/©2010
- Raj, A.S., Murali, N.: Morlet wavelet UDWT denoising and EMD based bearing fault diagnosis. *Electronics* **17**(1), 1–8 (2013)
- Pacheco, F., de Oliveira, J.V., Sánchez, R.-V., Cerrada, M., Cabrera, D., Li, C., Zurita, G., Artés, M.: A statistical comparison of neuro classifiers and feature selection methods for gearbox fault diagnosis under realistic conditions. *Neurocomputing* **194**, 192–206 (2016). <https://doi.org/10.1016/j.neucom.2016.02.028>

- Georgoulas, G., Karvelis, P., Loutas, T., Stylios, C.D.: Rolling element bearings diagnostics using the symbolic aggregate approximation. *Mech. Syst. Signal Process.* **60–61**, 229–242 (2015). <https://doi.org/10.1016/j.ymssp.2015.01.033>
- Wang, G.F., Li, Y.B., Luo, Z.G.: Fault classification of rolling bearing based on reconstructed phase space and Gaussian mixture model. *J. Sound Vib.* **323**, 1077–1089 (2009). <https://doi.org/10.1016/j.jsv.2009.01.003>
- Li, H., Zhao, J., Zhang, X., Teng, H., Yang, R., Hao, L.: Bearing fault diagnosis method using envelope analysis and euclidean distance. *Indones. J. Electr. Eng.* **12**(3), 1887–1894 (2014). <https://doi.org/10.11591/telkomnika.v12i3.4186>
- Zhou, H., Chen, J., Dong, G., Wang, R.: Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model. *Mech. Syst. Signal Process.* **72–73**, 65–79 (2016). <https://doi.org/10.1016/j.ymssp.2015.11.022>
- Liu, H., Wang, X., Lu, C.: Rolling bearing fault diagnosis based on LCD–TEO and multifractal detrended fluctuation analysis. *Mech. Syst. Signal Process.* **60–61**, 273–288 (2015). <https://doi.org/10.1016/j.ymssp.2015.02.002>
- Chebil, J., Hrairi, M., Abushikhah, N.: Signal analysis of vibration measurements for condition monitoring of bearings. *Aust. J. Basic Appl. Sci.* **5**(1), 70–78 (2011)
- Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., Siegel, D.: Prognostics and health management design for rotary machinery systems – reviews, methodology and applications. *Mech. Syst. Signal Process.* **42**, 314–334 (2014). <https://doi.org/10.1016/j.ymssp.2013.06.004>
- Sanz, J., Perera, R., Huerta, C.: Gear dynamics monitoring using discrete wavelet transformation and multi-layer perceptron neural networks. *Appl. Soft Comput.* **12**, 2867–2878 (2012). <https://doi.org/10.1016/j.asoc.2012.04.003>
- Zhu, K., San Wong, Y., Hong, G.S.: Multi-category micro-milling tool wear monitoring with continuous hidden Markov models. *Mech. Syst. Signal Process.* **23**, 547–560 (2009). <https://doi.org/10.1016/j.ymssp.2008.04.010>
- Rabiner, L.R.: A tutorial on hidden Markov models and selection. *Proc. IEEE* **77**(2), 257–286 (1989)
- Kan, M.S., Tan, A.C., Mathew, J.: A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mech. Syst. Signal Process.* **62–63**, 1–20 (2015). <https://doi.org/10.1016/j.ymssp.2015.02.016>
- Miao, Q., Huang, H.-Z., Fan, X.: A comparison study of support vector machines and hidden Markov models in machinery condition monitoring. *J. Mech. Sci. Technol.* **21**, 607–615 (2007)
- Rodriguez, P.H., Alonso, J.B., Ferrer, M.A., Travieso, C.M.: Application of the Teager-Kaiser energy operator in bearing fault diagnosis. *ISA Trans.* **52**, 278–284 (2013). <https://doi.org/10.1016/j.isatra.2012.12.006>
- Ziani, R., Felkaoui, A., Zegadi, R.: Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized Fisher’s criterion. *J. Intell. Manuf.* **28**(2), 405–417 (2014). <https://doi.org/10.1007/s10845-014-0987-3>
- Watanabe, S., Nakamura, A., Juang, B.-H.F.: Structural bayesian linear regression for hidden Markov models. *J. Signal Process. Syst.* **74**(3), 341–358 (2013). <https://doi.org/10.1007/s11265-013-0785-8>
- Wu, S.-D., Wu, P.-H., Wu, C.-W., Ding, J.-J., Wang, C.-C.: Bearing fault diagnosis based on multiscale permutation entropy and support vector machine. *Entropy* **14**, 1343–1356 (2012). <https://doi.org/10.3390/e14081343>
- Chen, X., Zhou, J., Xiao, J., Zhang, X., Xiao, H., Zhu, W., Fu, W.: Fault diagnosis based on dependent feature vector and probability neural network for rolling element bearings. *Appl. Math. Comput.* **247**, 835–847 (2014). <https://doi.org/10.1016/j.amc.2014.09.062>

- Tian, Y., Ma, J., Lu, C., Wang, Z.: Rolling bearing fault diagnosis under variable conditions using LMD-SVD and extreme learning machine. *Mech. Mach. Theory* **90**, 175–186 (2015). <https://doi.org/10.1016/j.mechmachtheory.2015.03.014>
- Qiang, Y., Chen, L., Hua, L., Gu, J., Ding, L., Liu, Y.: Research on the classification for faults of rolling bearing based on multi-weights neural network. *Int. J. Smart Sens. Intell. Syst.* **7**(3), 1004–1023 (2014)
- Li, Z., Fang, H., Huang, M.: Diversified learning for continuous hidden Markov models with application to fault diagnosis. *Expert Syst. Appl.* **42**, 9165–9173 (2015). <https://doi.org/10.1016/j.eswa.2015.08.027>