Bearing Fault Classification Using Wavelet Energy and Autoencoder

Sandeep S. Udmale[1,2](✉) and Sanjay Kumar Singh[2]

[1] Department of Computer Engineering and IT,
Veermata Jijabai Technological Institute (VJTI), Mumbai 400019, Maharashtra, India
ssudmale@it.vjti.ac.in
[2] Department of Computer Science and Engineering,
Indian Institute of Technology (BHU), Varanasi 221005, Uttar Pradesh, India
{ssudmale.rs.cse17,sks.cse}@iitbhu.ac.in

Abstract. Today's modern industry has widely accepted the intelligent condition monitoring system to improve the industrial organization. As an effect, the data-driven-based fault diagnosis methods are designed by integrating signal processing techniques along with artificial intelligence methods. Various signal processing approaches have been proposed for feature extraction from vibration signals to construct the fault feature space, and thus, over the years, the feature space has increased rapidly. Also, the challenge is to identify the promising features from the space for improving diagnosis performance. Therefore, in this paper, wavelet energy is presented as an input feature set to the fault diagnosis system. In this paper, wavelet energy is utilized to represent the multiple faults for reducing the requirement of number features, and therefore, the complex task of feature extraction becomes simple. Further, the convolutional autoencoder has assisted in finding more distinguishing fault feature from wavelet energy to improve the diagnosis task using extreme learning machine. The proposed method testified using two vibration datasets, and decent results are achieved. The effect of autoencoder on fault diagnosis performance has been observed in comparison to principal component analysis (PCA). Also, the consequence has seen in the size of the extreme learning machine (ELM) architecture.

Keywords: Autoencoder · Bearing · Fault diagnosis · Wavelet energy

1 Introduction

Recently, the rapid development of sensor technology as well as communication networks has helped to develop and deploy the cyber-physical system (CPS) in advanced industrial systems. The emergence of this CPS has made the data collection much easier for various processes conditions and a significant amount of data generated for the analysis of machinery health status [1–3]. The advances in artificial intelligence techniques have assisted in extracting useful information from the substantial amount of vibration data for mitigating the issues of

fault diagnosis. As a practical means, it reduces the risk of unplanned shut-downs/breakdowns and assures the reliability as well as the safety of industrial systems. The majority of unexpected shutdowns occur due to the failure of rolling element bearing, and classifying the bearing fault as early as possible provides adequate time for maintenance planning and saves the time as well as money of industry [4–6].

Over the years, a significant amount of methods have been proposed for feature extraction and selection/transformation in bearing fault diagnosis. These approaches aim to enhance the characterization of fault features for improving fault diagnosis performance, and the remarkable advancement in the signal processing techniques has supported significantly for the various fault characterization. Therefore, the fault feature space has increased rapidly and become hybrid in nature in the last two decades due to the advancement in signal processing techniques. This advancement has introduced the various multi-domain features by the researchers to represent the multiple faults and fault severities [7–9].

Recently, a Neuro-Fuzzy system has been developed to identify the bearing health status by learning the resonant zones of vibration signals in the frequency domain [10]. The statistical feature space has been designed using the wavelet packet transform, and then distance-based criteria have been used to build the Bayesian inference fault information [11]. The Shannon entropy has assisted in choosing wavelet transform (WT) for designing the statistical feature space to calculate bearing health using artificial neural network and support vector machine [12,13]. A spectral kurtosis has assisted in extracting time-frequency energy density information to train the optimized extreme learning machine (ELM) [14]. The nonlinear and non-Gaussian characteristics of vibration data have been captured using intrinsic mode function to develop the statistical feature space for expressing multiple fault pattern. Then the fault diagnosis task has been performed by the support vector machine [15]. The wavelet-spectrum method along with kurtosis ratio reference functions has assisted in calculating the health indicator for the monitoring of bearing based on Neuro-Fuzzy classifier [16].

Along with machine learning methods, recently, a deep learning approach has been broadly employed for fault classification. The kurtogram has been evaluated using a convolutional neural network (CNN) [17,18] and recurrent neural network [19] to enhance the fault diagnosis results. The statistical feature space has been designed by capturing the vibration data from multiple sources to develop a responsive fault diagnosis method using deep neural network [20]. The 2D input feature map has been constructed using time and frequency domain statistical features to train the CNN for fault diagnosis [21,22]. The various signal processing methods have served to establish the multi-domain feature set for the learning of stacked Gaussian-Bernoulli restricted Boltzmann machines [23]. A similar type of feature set has been developed by Chen et al. to examine the performance of bearing fault diagnosis using deep Boltzmann machines, deep belief network, and stacked autoencoders (AE) [24].

The above literature signifies that the various methods have been proposed to extract the fault features for improving fault diagnosis results. The different signal processing domain features have been analyzed by introducing new features along with existing features for identification of gear and bearing fault [25]. For developing the fault features spaces, one must be master in the signal processing and then determine the promising features from the space to express various defects by designing a robust feature selection routine. Therefore, in this paper, this complex task of identifying the fault features has been simplified by introducing the wavelet energy (WE) as a feature set to represent the various faults. In addition, to enhance the fault pattern representation capability of WE, convolutional AE has applied to the feature set. Then the efficient, as well as fast classifier like ELM, has been utilized for fault diagnosis. However, Haidong et al. have proposed a similar approach of fault diagnosis with wavelet as an activation function of AE [26]. The major difference between the proposed method and approach proposed by Haidong et al. is the use of convolutional layers for the extraction of significant fault information. This information is very crucial for differentiating the faults of short duration and low amplitude value such as ball fault and healthy bearing. The effect of AE has been analyzed for achieving small-size ELM architecture in the proposed method. The two vibration datasets have been used for testifying the proposed method and compared with existing alternatives.

2   Related Work

In this section, we briefly introduce the concepts of wavelet energy and autoencoder, which underlies the proposed method.

2.1   Wavelet Energy

The effective method to represent the signal in the time-frequency domain is a WT. The delightful property of WT is that at low frequencies, it presents the significant frequency information, and at high frequencies, it provides adequate information of time [9,27]. These characteristics are vital for fault diagnosis because the vibration signal contains the high frequencies as well as low frequencies components. Also, the non-gaussianity of vibration signal has been analyzed by WT to locate the transient present in the signal. The analysis using wavelet depends on the selection of the mother wavelet for the characterization of signal, and the unique decomposition of the signal provides the advantage in signal analysis [27,28].

The wavelet family set $\Psi_{n,j}(t)$ for multiresolution analysis in $L^2(\Re)$ is an orthogonal basic, and the idea of calculating the energy is derived from the Fourier theory. Initially, the mother wavelet $\Psi(t)$ is selected with the decomposition levels N [27,28]. At different decomposition levels, energy is expressed as the energy of wavelet coefficients $C_{n,j}$. Thus, the energy is defined at different decomposition levels with m number of wavelet coefficients as

$$E_n = \sum_{j=1}^{m} |C_{n,j}|^2 \qquad (1)$$

where the jth wavelet coefficient of nth scale is $C_{n,j}$. Thus, the total energy is calculated by

$$E_{total} = \sum_{n} E_n = \sum_{n} \sum_{j} |C_{n,j}|^2 \qquad (2)$$

The above equations provide essential information at different frequency bands for characterizing the vibration signal energy distribution.

2.2 Autoencoder

Autoencoder belongs to the family of deep learning architecture in which the dimension of input and output is the same so that it will reconstruct the input at the output stage through intermediate layers with a reduced number of hidden nodes. The basic idea of AE is to compress the input data into a latent-space representation, and then reconstruct the output from this representation. Therefore, the AE network consists of encoder and decoder parts for performing dimensionality reduction and data denoising tasks [29,30].

In the proposed method, WE is treated as 2D data and therefore, to extract the more significant information for the fault representation by dimensionality reduction, convolutional AE has operated on WE E_i [29]. The encoder assists in determining the reduced hidden representation $\hat{E} \in R^{d_{\hat{E}}}$ by mapping the $E_i \in R^{d_E}$ with the help of the feed-forward process of the network as

$$\hat{E} = \varphi_L \left(\cdots \varphi_2 \left(\varphi_1 \left(E, \delta^{(1)} \right), \delta^{(2)} \right) \cdots, \delta^{(L)} \right) \qquad (3)$$

where δ is the learning parameters of a network for each stage L and φ are the various convolutional and pooling operation performed by the network at each L. The decoder reconstructs the $E' \in R^{d_E}$ from $\hat{E}$ as

$$E' = \varphi_1 \left(\cdots \varphi_{L-1} \left(\varphi_L \left(\hat{E}, \delta^{(L)} \right), \delta^{(L-1)} \right) \cdots, \delta^{(1)} \right) \qquad (4)$$

Finally, the training of AE is a minimization of error between E, and E' and it is given as

$$\underset{\varphi, \delta}{argmin} \left(E, E' \right) \qquad (5)$$

AE illustrates a promising potential to study the meaningful features from E for fault diagnosis.
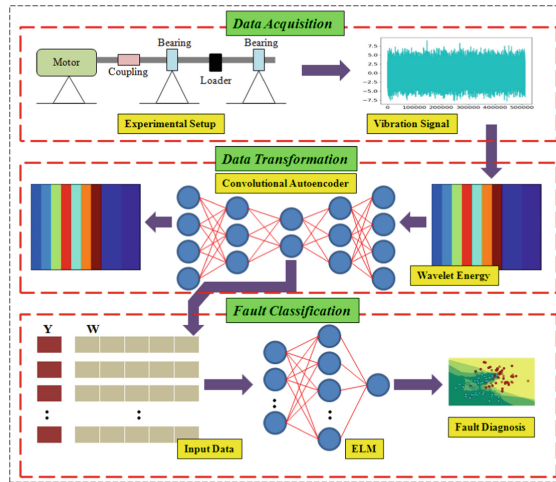
Fig. 1. Proposed method framework.

3 Proposed Method

Several signals generated through machines provide adequate information about the health state of the machine by analyzing them. Therefore, in this paper, a novel bearing fault classification method is proposed by incorporating WE and AE for identifying the distinguishing features, and then various fault patterns are learned by ELM. The proposed bearing fault classification method is manifested in Fig. 1, and it is consists of the following phases: (1) Data acquisition, (2) Feature transformation, and (3) Fault classification.

It has been found in the literature that the vibration data has been utilized extensively for the development of fault diagnosis solutions. As a result, a two bearing vibration set has assisted in testifying the proposed method. The detail information of data acquisition with experimental setup and dataset description as well as analysis is provided in [17, 19, 31–33].

The traditional signal processing methods could not able to handle a large amount of industrial data effectively for identifying the various fault patterns. As a result, it causes a bottleneck for precise and timely evaluation of bearing health conditions. In addition, assuming the idealization and simplifications of vibration data extend the inappropriate review of the bearing health. As an effect, the reliability of the overall bearing health diagnosis system reduces. Therefore, in this paper, to determine the various faults, WE have utilized it as a feature. Hence, in the second stage, the raw data has been transformed into WE. Further, to mine important information about the various defects and improve the fault diagnosis performance, the AE based dimension reduction technique has been operated on WE feature in the data transformation stage. It identifies the more distinguishing features to articulate the various features. The W contains the

Table 1. Performance analysis of the proposed method.

Algorithm	Training accuracy (%)	Testing accuracy (%)	F-score	Computational time (s.)
Dataset-1				
ELM (Sigmoid)	98.33 ± 1.17	96.34 ± 1.08	0.96 ± 0.01	170.11 ± 0.01
ELM (RBF)	98.61 ± 0.54	97.39 ± 0.77	0.97 ± 0.01	167.21 ± 0.01
MLP	96.11 ± 0.0	92.91 ± 0.0	0.92 ± 0.0	165.5 ± 0.01
Dataset-2				
ELM (Sigmoid)	98.13 ± 1.57	95.14 ± 3.27	0.95 ± 0.03	125.71 ± 0.01
ELM (RBF)	98.11 ± 1.1	95.94 ± 1.79	0.96 ± 0.02	125.24 ± 0.01
MLP	90.63 ± 0.0	86.75 ± 0.0	0.86 ± 0.0	125.24 ± 0.01

input data, and Y is the corresponding class label. In the final stage, ELM has trained using W and Y to study the different fault diagnosis patterns.

3.1   Benefits of Proposed Method

– The use of WE as an input feature has reduced the task of constructing the feature space using various signal processing methods.
– Further, there is no need to find promising features from the feature space. As a result, feature extraction and selection process become simple.
– The convolutional AE has assisted in dimension reduction of input data and identifies the more distinguishing feature to represent the fault.
– The use of AE helps in reducing the requirement of L, and hence, small size ELM architecture has been attained by preserving the accuracy of the model.
– In addition, AE has assisted in improving the fault diagnosis performance.

4   Results and Discussion

4.1   Simulation Environment

The two REB datasets are used to verify the effectiveness of the proposed fault diagnosis method. Dataset-1: The machine fault simulator is used to generate the vibration signal under different working conditions in a supervised manner [31] and Dataset-2: Publicly available bearing dataset from Case Western Reserve University (CWRU) [32]. The simulation environment (i7-CPU with the frequency of 3.6 GHz, 8.0-GB RAM, and Ubuntu 16.04 operating system) has setup to analyze the fault diagnosis results. The 50 trails are performed to calculate the average as well as standard deviation value for training and testing performance along with F-score and computational time. The K-fold cross-validation method has been employed to deliver the data partitioning exercise. The AE consists
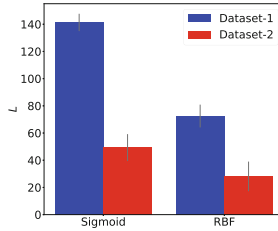
Fig. 2. Value of L for ELM the performance in Table 1.

Table 2. Performance analysis of the proposed method without autoencoder.

Algorithm	Training accuracy (%)	Testing accuracy (%)	F-score	Computational time (s.)
Dataset-1				
ELM (Sigmoid)	93.85 ± 0.81	90.48 ± 1.02	0.90 ± 0.01	2.29 ± 0.01
ELM (RBF)	91.69 ± 0.01	89.37 ± 0.01	0.88 ± 0.0	2.61 ± 0.01
MLP	85.83 ± 0.0	84.31 ± 0.0	0.84 ± 0.0	2.32 ± 0.01
Dataset-2				
ELM (Sigmoid)	90.14 ± 1.57	79.41 ± 2.24	0.79 ± 0.02	0.57 ± 0.01
ELM (RBF)	88.42 ± 0.01	84.11 ± 0.01	0.83 ± 0.0	0.36 ± 0.01
MLP	85.18 ± 0.0	80.81 ± 0.0	0.79 ± 0.0	0.16 ± 0.01

of six layers of convolutional and pooling. The first three layers are used for encoding and the last three for decoding. The bottom layer of encoding obtains the low-level features, and a top layer acquires the high-level features of faults to train the classifier [26]. For the comparison, the stopping RMSE is set to 0.2, and sigmoid as well as RBF type of nodes are considered in ELM [34].

4.2 Result Analysis

Table 1 illustrates the performance of the proposed method. For both the vibration datasets, the ELM demonstrate the training and testing performance respectively of 98.0% and 95.0% with RMSE equal to 0.2. The RBF types of nodes show the acceptable performance in comparison to the sigmoid type of nodes for both the datasets. The best performance of 97.39% has been recorded for the dataset-1 and, similarly, 95.94% for dataset-2. The SD value of performance is around 1.0 for dataset-1 and 2.0 for dataset-2. The F-score values of different algorithms are above 0.95.

Also, the hidden node requirement of a sigmoid node is high compared to RBF nodes, and it is almost 300.0% more in both the datasets as shown in Fig. 2. However, the SD value of L for a sigmoid node is better than RBF nodes, and it is approximately below 10. It is also noted that the computational time

Table 3. Performance analysis of the proposed method with PCA.

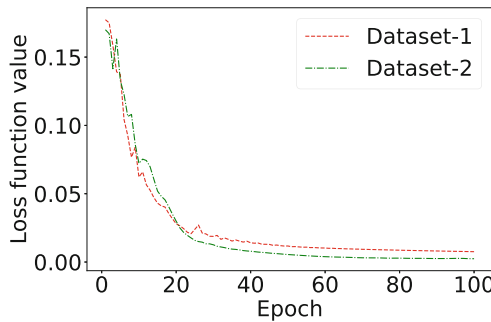| Algorithm	Training accuracy (%)	Testing accuracy (%)	F-score	Computational time (s.)
Dataset-1				
ELM (Sigmoid)	98.12 ± 0.51	95.75 ± 0.82	0.95 ± 0.01	5.16 ± 0.01
ELM (RBF)	95.13 ± 0.01	89.74 ± 0.01	0.90 ± 0.0	3.64 ± 0.01
MLP	90.03 ± 0.0	87.73 ± 0.0	0.86 ± 0.0	3.44 ± 0.01
Dataset-1				
ELM (Sigmoid)	92.64 ± 1.18	86.19 ± 0.06	0.86 ± 0.06	3.36 ± 0.01
ELM (RBF)	91.21 ± 0.01	88.16 ± 0.01	0.88 ± 0.0	3.16 ± 0.01
MLP	88.64 ± 0.0	84.16 ± 0.01	0.83 ± 0.0	2.59 ± 0.01

Fig. 3. Training speed comparison of the datasets for autoencoder.

requirement of RBF nodes is less as compared to sigmoid nodes. This impact has been observed due to the less demand for RBF hidden nodes in comparison to sigmoid nodes for achieving similar performance. This effect has been found in both the data sets. In addition, the proposed method has been compared with MLP. Table 1 indicates that the MLP performance is not acceptable in comparison to ELM and it is almost 5.0% to 9.0% less than ELM performance.

Further, to analyze the effect of AE, the WE has been presented as an input to the classifier, and L value for the ELM has been utilized from the Fig. 2 to illustrate the generalize performance for comparison. Table 2 demonstrates the results for the proposed method without the AE step. It has been found that the overall performance has decreased for all the types of nodes. The significant effect has been noticed for the dataset-2 as compared to dataset-1. The total decrement in the training performance is approximately 6.0% and 10.0%, respectively, for the dataset-1 and dataset-2. Similarly, for the testing performance, it has been 6.0% and 15.0%, respectively, for the dataset-1 and dataset-2. The effect of performance degradation has observed on the F-score value. Besides F-score,

Table 4. Statistical feature set performance comparisons of ELM and MLP.

Algorithm	Training accuracy (%)	Testing accuracy (%)	F-score	Computational time (s.)
Dataset-1				
ELM (Sigmoid)	72.83 ± 3.65	65.31 ± 3.92	0.65 ± 0.04	1.78 ± 0.4
ELM (RBF)	78.19 ± 0.28	76.45 ± 0.21	0.76 ± 0.01	1.69 ± 0.01
MLP	35.41 ± 0.18	34.58 ± 0.1	0.29 ± 0.01	1.32 ± 0.17
Dataset-2				
ELM (Sigmoid)	87.09 ± 2.34	78.64 ± 3.91	0.77 ± 0.03	1.18 ± 0.06
ELM (RBF)	87.5 ± 0.1	86.36 ± 0.1	0.85 ± 0.1	0.96 ± 0.02
MLP	39.77 ± 0.08	38.63 ± 0.04	0.38 ± 0.0	0.51 ± 0.1

the computational time of the method without AE is better than the routine with AE.

Further, PCA has been widely utilized to extract sensitive features and reduced feature size in bearing [35]. Therefore, to analyze the effect of AE, the WE has been presented as an input to principal component analysis (PCA), and then to the classifier. L value for the classifier has been utilized from the Table 1 to display the generalize performance for comparison. Table 3 illustrates the results for the proposed method with PCA as a dimension reduction step. It has been notified from the Tables 1, 2 and 3 that the AE and PCA improve the fault diagnosis performance. PCA based proposed method performance is not acceptable in comparison to AE based proposed method. Similar to AE, in PCA based approach, RBF provides better performance than sigmoid nodes. Also, the computational time of the PCA based method is better than the AE-based method. This effect of computational time has seen due to the utilization of various convolutional and pooling layers.

From Tables 1, 2 and 3, it can be concluded that the proposed method reduces the size of ELM architecture. With the same number of L, the method with PCA and without AE has unable to attain a similar performance. Thus, the combined effect of WE and AE has been seen on the overall performance of the system. Also, Fig. 3 demonstrates the training speed of the AE using both the datasets.

In addition, the fault diagnosis system development Samanta and Al-Balushi by utilizing statistical features have been analyzed for the comparison with WE feature set [36]. Tables 1 and 4 indicates that the performance of the proposed method is better than the statistical feature set, and it is improved by 20.0% [19]. Also, the result notifies that the proposed method illustrates the satisfactory results with wavelet energy.

5   Conclusion

In the modern industry, the rotating machinery has been broadly employed for various applications. Therefore, the health monitoring of equipment is an

essential task for the overall functioning of the system. Hence, bearing fault classification performance has improved by introducing the wavelet energy as an input feature set in this paper. Also, it simplifies the feature extraction and selection process of fault diagnosis. In addition, it proves the wavelet energy as a useful input feature vector for fault diagnosis. Further, the use of autoencoder has assisted in identifying the promising features from wavelet energy to achieve the small size ELM design by preserving the accuracy of the solution.

References

1. Hossain, M.S., Muhammad, G.: Cloud-assisted industrial internet of things (IIoT) - enabled framework for health monitoring. Comput. Netw. **101**, 192–202 (2016)
2. Ren, L., Cheng, X., Wang, X., Cui, J., Zhang, L.: Multi-scale dense gate recurrent unit networks for bearing remaining useful life prediction. Future Gener. Comput. Syst. **94**, 601–609 (2019)
3. Kan, C., Yang, H., Kumara, S.: Parallel computing and network analytics for fast industrial internet-of-things (IIoT) machine information processing and condition monitoring. J. Manuf. Syst. **46**, 282–293 (2018)
4. Bellini, A., Filippetti, F., Tassoni, C., Capolino, G.A.: Advances in diagnostic techniques for induction machines. IEEE Trans. Ind. Electron. **55**(12), 4109–4126 (2008)
5. Henriquez, P., Alonso, J.B., Ferrer, M.A., Travieso, C.M.: Review of automatic fault diagnosis systems using audio and vibration signals. IEEE Trans. Syst. Man Cybern. Syst. **44**(5), 642–652 (2014)
6. Dai, X., Gao, Z.: From model, signal to knowledge: a data-driven perspective of fault detection and diagnosis. IEEE Trans. Ind. Inf. **9**(4), 2226–2238 (2013)
7. Kan, M.S., Tan, A.C., Mathew, J.: A review on prognostic techniques for non-stationary and non-linear rotating systems. Mech. Syst. Sign. Process. **62**, 1–20 (2015)
8. Choudhary, A., Goyal, D., Shimi, S.L., Akula, A.: Condition monitoring and fault diagnosis of induction motors: a review. Arch. Comput. Methods Eng. **26**(4), 1221–1238 (2018)
9. El-Thalji, I., Jantunen, E.: A summary of fault modelling and predictive health monitoring of rolling element bearings. Mech. Syst. Sign. Process. **60**, 252–272 (2015)
10. Marichal, G., Artés, M., Prada, J.G., Casanova, O.: Extraction of rules for faulty bearing classification by a neuro-fuzzy approach. Mech. Syst. Sign. Process. **25**(6), 2073–2082 (2011)
11. Zhang, S., Mathew, J., Ma, L., Sun, Y.: Best basis-based intelligent machine fault diagnosis. Mech. Syst. Sign. Process. **19**(2), 357–370 (2005)
12. Kankar, P., Sharma, S.C., Harsha, S.: Fault diagnosis of ball bearings using continuous wavelet transform. Appl. Soft Comput. **11**(2), 2300–2312 (2011)
13. Kankar, P., Sharma, S.C., Harsha, S.P.: Rolling element bearing fault diagnosis using wavelet transform. Neurocomputing **74**(10), 1638–1645 (2011)

14. Udmale, S.S., Singh, S.K.: A mechanical data analysis using kurtogram and extreme learning machine. Neural Comput. Appl. 1–13 (2019). https://doi.org/10.1007/s00521-019-04398-0

15. Soualhi, A., Medjaher, K., Zerhouni, N.: Bearing health monitoring based on Hilbert-Huang transform, support vector machine, and regression. IEEE Trans. Instrum. Meas. **64**(1), 52–62 (2015)

16. Liu, J., Wang, W., Golnaraghi, F.: An enhanced diagnostic scheme for bearing condition monitoring. IEEE Trans. Instrum. Meas. **59**(2), 309–321 (2010)

17. Udmale, S.S., Patil, S.S., Phalle, V.M., Singh, S.K.: A bearing vibration data analysis based on spectral kurtosis and ConvNet. Soft. Comput. **23**(19), 1–19 (2019)

18. Udmale, S.S., Singh, S.K., Singh, R., Sangaiah, A.K.: Multi-fault bearing classification using sensors and ConvNet-based transfer learning approach. IEEE Sens. J. 1–12 (2019). https://doi.org/10.1109/JSEN.2019.2947026

19. Udmale, S.S., Singh, S.K., Bhirud, S.G.: A bearing data analysis based on kurtogram and deep learning sequence models. Measurement **145**, 665–677 (2019)

20. Tao, J., Liu, Y., Yang, D.: Bearing fault diagnosis based on deep belief network and multisensor information fusion. Shock Vibr. **2016**, 9 (2016)

21. Chen, Z., Li, C., Sanchez, R.V.: Gearbox fault identification and classification with convolutional neural networks. Shock Vibr. **2015**, 10 (2015)

22. Lu, C., Wang, Z., Zhou, B.: Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. Adv. Eng. Inf. **32**, 139–151 (2017)

23. Li, C., Sánchez, R.V., Zurita, G., Cerrada, M., Cabrera, D.: Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning. Sensors **16**(6), 895 (2016)

24. Chen, Z., Deng, S., Chen, X., Li, C., Sanchez, R.V., Qin, H.: Deep neural networks-based rolling bearing fault diagnosis. Microelectron. Reliab. **75**, 327–333 (2017)

25. Dhamande, L.S., Chaudhari, M.B.: Compound gear-bearing fault feature extraction using statistical features based on time-frequency method. Measurement **125**, 63–77 (2018)

26. Shao, H., Jiang, H., Li, X., Wu, S.: Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine. Knowl. Based Syst. **140**, 1–14 (2018)

27. Zhang, X., Yan, Q., Yang, J., Zhao, J., Shen, Y.: An assembly tightness detection method for bolt-jointed rotor with wavelet energy entropy. Measurement **136**, 212–224 (2019)

28. Pan, Y., Zhang, L., Wu, X., Zhang, K., Skibniewski, M.J.: Structural health monitoring and assessment using wavelet packet energy spectrum. Saf. Sci. **120**, 652–665 (2019)

29. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)

30. Wang, W., Huang, Y., Wang, Y., Wang, L.: Generalized autoencoder: a neural network framework for dimensionality reduction. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 496–503, June 2014

31. Udmale, S.S., Singh, S.K.: Application of spectral kurtosis and improved extreme learning machine for bearing fault classification. IEEE Trans. Instrum. Meas. **68**(11), 1–12 (2019)

32. CWRU: Case Western Reserve University Bearing Data Center Website (2009). https://csegroups.case.edu/bearingdatacenter/home

33. Smith, W.A., Randall, R.B.: Rolling element bearing diagnostics using the Case Western Reserve University data: a benchmark study. Mech. Syst. Sig. Process. **64–65**, 100–131 (2015)
34. Huang, G.B., Chen, L., Siew, C.K.: Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans. Neural Netw. **17**(4), 879–892 (2006)
35. Dong, S., Luo, T.: Bearing degradation process prediction based on the PCA and optimized LS-SVM model. Measurement **46**(9), 3143–3152 (2013)
36. Samanta, B., Al-Balushi, K.: Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. Mech. Syst. Sig. Process. **17**(2), 317–328 (2003)