

# Deep Learning for Fault Diagnosis based on short-time Fourier transform

Tarak BENKEDJOUH

Ecole Militaire Polytechnique  
Laboratoire LMS  
Bordj Elbahri Algiers, Algeria  
Email: bktarek@gmail.com

Noureddine ZERHOUNI

FEMTO-ST Institute UMR CNRS 6174  
UFC / ENSMM / UTBM/AS2M Department  
24, rue Alain Savary ,Besançon, France  
Email: zerhouni@ens2m.fr

Said RECHAK

Ecole Nationale Polytechnique  
Laboratoire Génie Mécanique  
Elharrach Algiers, Algeria  
Email: said.rechak@enp.edu.dz

**Abstract**—The rapid advancements of the Internet of Things (IoT) enables maintenance strategies to be applied everyday to all sectors, IoT based health management plays an important role For producing quickly, 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 the implementation of maintenance strategies for faults detection . In this paper, a novel method called deep learning based on Short-Time Fourier Transform (STFT) is developed for fault diagnosis. An experimental analysis is carried out using a dataset under different operating conditions of speed and loading to substantiate the utility of the proposed strategy. Also a multi-fault deep learning classifier based on STFT is constructed for different faults in this paper. Hence, the purpose is to design an automatic detection system for mechanical components defects based on supervised classification. The diagnosis accuracy assessment is carried out by conducting various experiments on acceleration signals collected from a rotating machinery under different operating conditions.

**Index Terms**—Deep learning, CNNs, Diagnostics, faults detection, classification, STFT

## I. INTRODUCTION

The ability to forecast machinery failure can help reducing maintenance costs, operation breakdowns and safety risks and gaining importance in industry since it may limit the loss of production due to a machine stopping [1]–[3]. Traditional signal processing techniques, including time-domain [4], time-frequency analysis [5] , linear prediction cepstral coefficients (LPCC) [6]...etc. They may result in false information, because the mechanical faults may be non-stationary and the random nature of defect growth by crack propagation in mechanical components, because each feature is effective for a defect at certain stage [7]. For improving the efficiency and effectiveness of fault diagnosis of rotating machines; Artificial intelligence have been widely developed in recent years, such as hidden Markov model (HMM) [8], expert systems [9]. Neural networks, depend upon the statistical characteristics of large-scale samples [10]. It is impossible to obtain good generalization performances when the number of training samples is insufficient. Because a neural network usually uses a gradient descent algorithm to update its weights, generalization of the method is poor. Moreover, there are some defects in neural networks that are difficult to solve. For example, the number

of hidden neurons is difficult to determine and the final weights are greatly influenced by the choice of initial weights.

The recent developments of failure prognostics have focused on implementing advanced time-frequency analysis to extract the robust features for representing the health indicator [11]. The concept of feature extraction for accurately assessing the bearing performance degradation is a critical step toward realizing an online bearing condition monitoring platform. Peng et al. [12] provided a review and summarized the development and applications of wavelet transform on machine condition monitoring and fault diagnosis over the past years.

Several research have been studied the feasibility of ANN for faults detection and isolation. However, its accuracy is highly dependent on the neural networks structure such as number of hidden layers and kernel function. for the improvement of the traditional RNN, a long short term memory based neural network scheme was proposed by Yuan et al. [13] utilizing Long Short-Term Memory neural network to get good diagnosis and prediction performance in the cases of complicated operations, hybrid faults and strong noises for RUL estimation of aero-engines in the cases of complicated operations, hybrid faults and strong noises. However, it still to develop an effective approach based on historical data such as deep learning. some properties which make intelligent conditions monitoring method difficult to develop such as: the high dimension, randomness and low SNR (signal-to-noise ratio) [14].

Recently, several learning methods has emerged called as deep learning that improved to learn higher level abstractions from the raw data [15]–[18], deep learning models automatically learn a hierarchical feature representation from raw data. CNNs, auto-encoders and deep belief network are the mostly known models in deep learning. and applied in many research area such as: speech recognition [19], image processing [20], [21], machinery condition monitoring and health assessment [22]–[25]. CNNs have gained a special status over the last few years as an especially promising form of deep learning. The aim of this study is to estimate the RUL of aero-engine units accurately. A good prognostic performance prediction is achieved with the proposed approach using raw feature selection, data preprocessing and sample preparation with time window.

The main contribution of this paper is to improve the accuracy of fault diagnosis of rotating machinery using STFT and deep learning. This approach is divided in two phases: features extraction phase and applying the CNNs for data classification and visualization phase. This step, called the convolution step. In CNNs, the STFT matrix is called the 'kernel' and the matrix formed by sliding the filter over the feature map and computing the dot product is called the 'Convolved Feature' or the 'Feature Map'. It is important to note that filters act as feature detectors from the original input matrix; To overcome the difficulties and applied it to fault detection and classification in this paper Pan et al. [14] the network consists of split layer, predict layer, update layer, pooling layer and full-connection layer. The split layer used to reduces the length of representations. Predict and update layer used to extract and deep representations of input data. After that, the pooling layer is used to reduce data volume, and the last step a full connection layer used for data classification. The advantages of this proposed approach is used to reduces the need for feature extraction and selection. The architecture that can be adapted to the problems easily. Firstly, the collected vibration signals are processed by STFT which decomposes them into a short frame. Finally, the performance of the proposed method was verified by a fault diagnosis case in a gearbox. The results indicated that this new method could judge and classify the multi-fault of rotating machinery quickly and effectively. This technique used for enhancing mechanical components fault diagnosis. Particularly, we investigate how can deep learning select features from STFT matrix in order to maximize the performance of the classifier.

This paper is organized as follows. Section 2 presents the description of the proposed method. Section 3 presents the STFT technique. Section 4 describes the proposed method based on CNNs for classification. Section 5 is dedicated to the experimental verification and results discussion and finally, section 6 concludes the paper.

## II. DESCRIPTION OF THE PROPOSED METHOD

The rapid advancements of the Internet of Things (IoT) enables maintenance strategies to be applied everyday to all sectors, thus creating a standard shift that is opening up significant new business opportunities. Developments are illustrated with cases of inventions from industrial manufacturing and consumer products. IoT based health management plays an important role For increasing plant reliability and availability, stabilizing the power supply with fewer current interruptions, and providing the industry with a good reputation and trust. Another advantages is ensuring that aging power infrastructure is appropriately monitored for unplanned failures.

Various conditions monitoring research works have been conducted for improving the performance classification. In Fig.2, 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

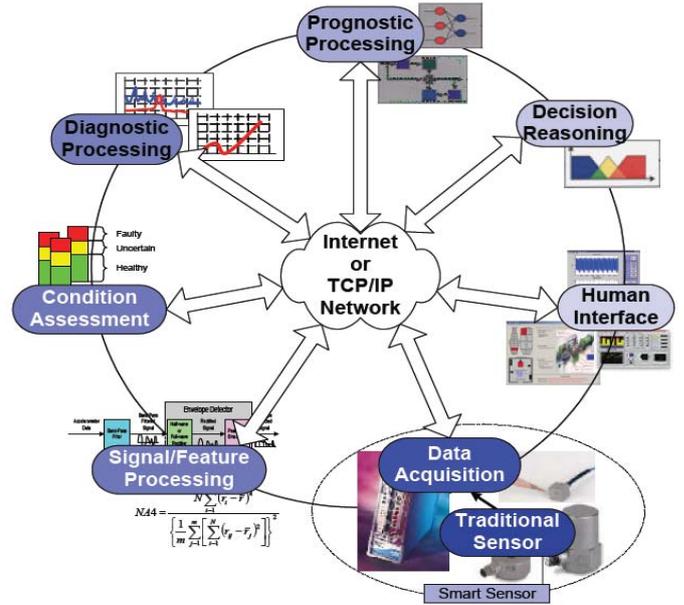


Fig. 1. Open monitoring system architecture.

acquired data and finally, in the maintenance decision-making step, effective maintenance policies will be obtained based on information analysis.

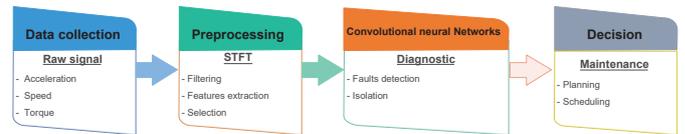


Fig. 2. Steps of a condition monitoring System.

## III. PRINCIPLE OF STFT

Short-Time Fourier Transform (STFT) is a popular method for analyzing nonstationary signals and is one of the most straightforward frequency domain. The basis for the STFT is a series of sinusoids. The visual representation in the time-frequency domain using STFT called spectrogram, and a scalogram uses the WT. Spectrograms add time to the analysis of FFT allowing the localization of both time and frequency. The STFT was introduced by Gabor and applied in many research area such as: speech processing image . First Gabors truncated Gaussian window was replace by optimal finite duration windows having better spectral properties, such as the Hamming and Kaiser windows. Second, zero-padding was introduced to increase the resolution in the frequency domain, to any desired degree.

The STFT of the signal  $x(t)$  is defined as:

$$X(t, f) = \int_{-\infty}^{+\infty} x(\tau)h(t - \tau)e^{-j2\pi f\tau} d\tau \quad (1)$$

where  $h(t)$  should be a lowpass filter, and  $\|h\|^2 = 1$ .  $X(t, f)$  can be interpreted as the correlation between  $x(\tau)$  and  $h(t - \tau)e^{-j2\pi f\tau}$ . Note that  $h(t - \tau)e^{-j2\pi f\tau}$  has its energy concentrated at time  $t$  and frequency  $f$ . Thus,  $|X(t, f)|^2$  can be viewed as the energy in  $x(t)$  at frequency  $f$  and time  $t$ . Often, one displays the energy at each time and frequency pair, i.e.,  $P(t, f) = |X(t, f)|^2$  is known as the spectrogram (SP) of  $x(t)$ . The STFT has a fundamental property that simplifies the interpretation of the resultant distribution: magnitude-wise shift invariance in both time and frequency, this is one of the most important properties.

#### IV. DEEP CNNs-BASED GEARBOX FAULTS DIAGNOSIS

This section presents the architecture of deep learning CNNs for health assessment and RUL estimation from different sensor signals. The inputs features are extracted and selected from different operating condition history. The output values are the RUL of bearings at corresponding lifetime. In honor Yann LeCun [26] an early pioneer of CNNs and the first to reduced them to practice by training them with gradient descent . for image processing, which is featured by two key properties: spatially shared weights and spatial pooling

The deep learning algorithms are machine learning technique based on distributed representations. Deep learning attempts to learn high-level features in data by using structures composed of multiple non-linear transformations. The frequently used models are CNNs and Deep Belief Network (DBN).

CNNs is also a type of feedforward neural network which is composed of alternating convolutional and subsampling layer [27], [28]. CNNs are designed to use minimal amounts of preprocessing, which is the main difference compared to other deep architectures.

Firstly, we assume that the input sequential data is  $y = [y_1, \dots, y_n]$  that  $n$  is the length of the sequence and  $y_i \in \mathbb{R}^d$  at each time step. Convolution: the dot product between a filter vector  $u \in \mathbb{R}^m$  and an concatenation vector representation  $y_{i:i+m-1}$  defines the convolution operation as follows:

$$c_i = \varphi(u^T y_{i:i+m-1} + b) \quad (2)$$

where  $b$  and  $\varphi$  denotes bias term and non-linear activation function, respectively.  $y_{i:i+m-1}$  is a  $m$  - length window starting from the  $i^{th}$  time step, which is described as:

$$y_{i:i+m-1} = y_i \oplus y_{i+1} \oplus \dots \oplus y_{i+m-1} \quad (3)$$

As defined in Eq. 3, the output scale  $c_i$  can be regarded as the activation of the filter  $u$  on the corresponding subsequence  $y_{i:i+m-1}$ . By sliding the filtering window from the beginning time step to the ending time step, a feature map as a vector can be given as follows:

$$c_j = [c_1, c_2, \dots, c_{l-m+1}] \quad (4)$$

where the index  $j$  represents the  $j^{th}$  filter. It corresponds to multi-windows as:

$\{y_{1:m}, y_{2:m+1}, \dots, y_{l-m+1:l}\}$ . Max-pooling: Pooling layer is able to reduce the length of the feature map, which can further minimize the number of model parameters. The hyper-parameter of pooling layer is pooling length denoted as  $s$ . MAX operation is taking a max over the  $s$  consecutive values in feature map  $c_j$ .

Then, the compressed feature vector can be obtained as:

$$h = [h_1, h_2, \dots, h_{\frac{l-m}{s}+1}] \quad (5)$$

where  $h_j = \max(c_{(j-1)s}, c_{(j-1)s+1}, \dots, c_{js-1})$ . Then, via alternating the above two layers: convolution and max-pooling ones, fully connected layers and LSTM layer are usually added as the top layers to make estimation.

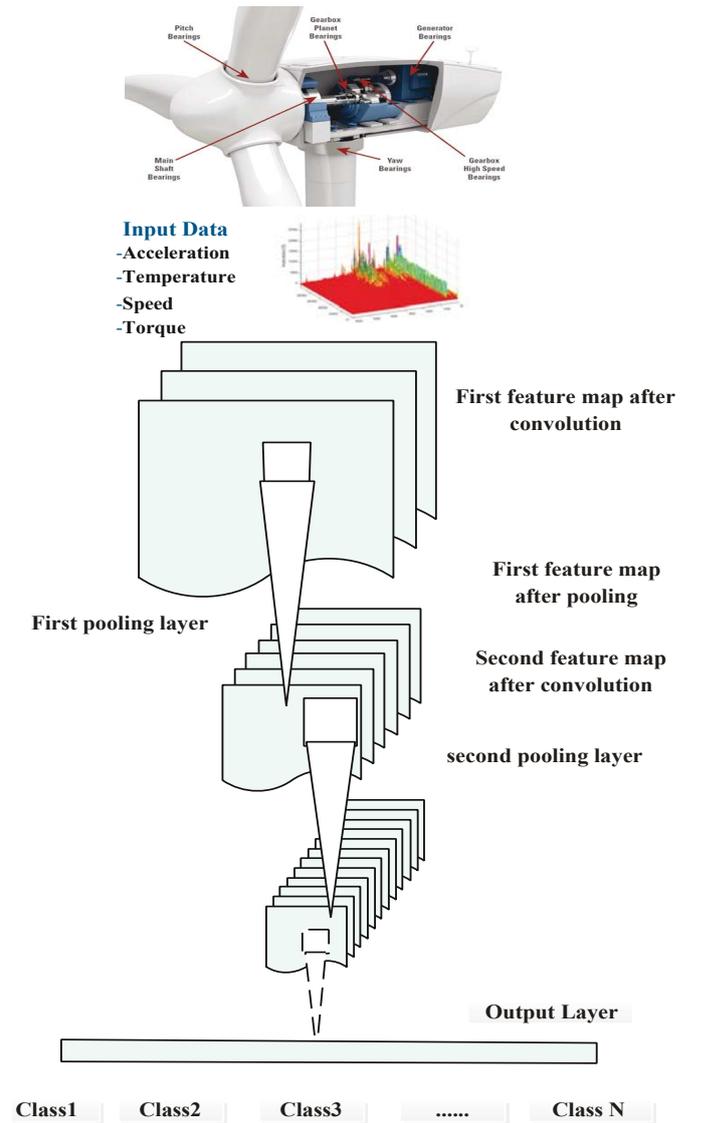


Fig. 3. Illustrations of the proposed CNNs for faults diagnosis

The proposed method diagram for the bearings health assessment and RUL estimation shown in figure 5. The method is decomposed into two main phases the preprocessing

data and training data using deep CNNs. the bearing dataset prepared for the training by computing the features extraction selected in this study using CNNs. Then training and testing datasets are prepared for deep CNNs. In the second phase, which is achieved on-line, deals with the utilization of the model generated continuously to assess the health state of the bearing and RUL prediction. All the layers use activation functions  $\tanh$ ,

## V. RESULTS AND DISCUSSION

### A. Experimental setup

To demonstrate the effectiveness of our developed classification framework, figure 4 illustrates the experimental setup used to accomplish our experience and data collection. The shaft is driven by an electric motor and the rotation speed was varied between 0 and 6000 rpm. 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 angle is  $0^\circ$ . The measured signals consist of two acceleration signals given by an Endevo 6259M31 Accelerometer ( $10\text{mv/g}$ ,  $\pm 1\%$  error, Resonance  $> 45\text{KHz}$ ) which is installed in input and output position on the gearbox housing. The data sampling rate was 66666.67 Samples per Second ( $200\text{KHz}/3$ ). 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.

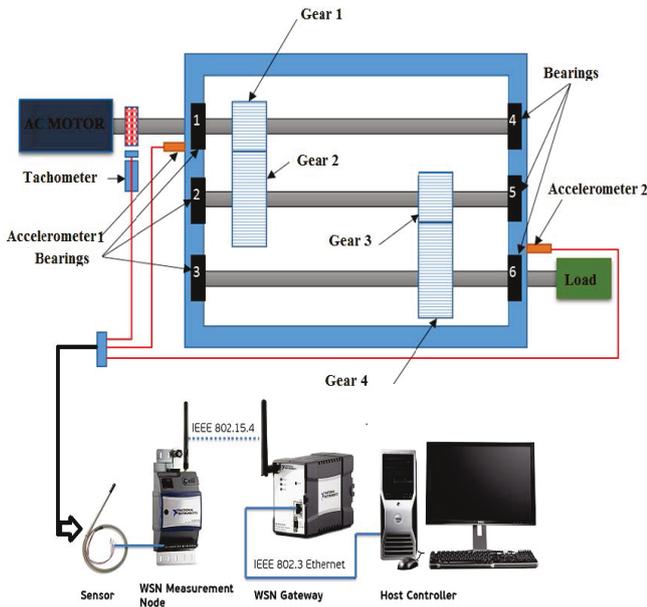


Fig. 4. Experimental setup.

### B. Experimental verification

In this section, a detailed description of our proposed algorithm is given. Before stating the approach, a flowchart of the classification system is depicted in Figure 5.

The diagram of the CNNs method proposed for conditions monitoring shown in Figure 5. The method is decomposed into two main steps. The first step is done off-line and aims at STFTs generating and classification. When the CNNs 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.

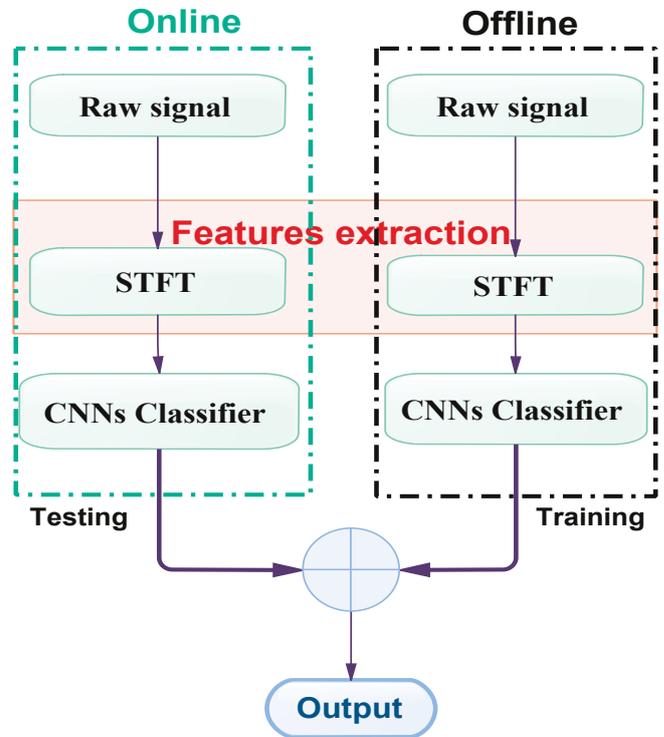


Fig. 5. Framework of the faults detection procedure

Figures 6 show the sensor measurements of the healthy and degraded state of the system (Acceleration) respectively.

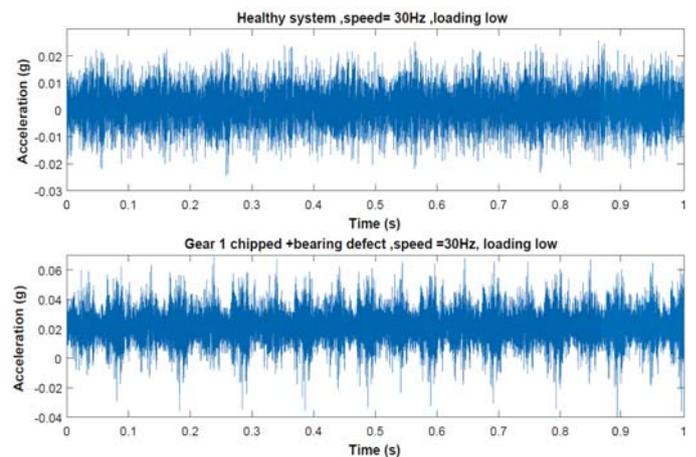


Fig. 6. Acceleration signals measurement of healthy system (top) and defect (bottom)

We decompose the monitoring signals of each loading data

above two conditions with STFTs 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 STFTs components. 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 fault types and numbers are shown in I and the same defect with helical gear. Here we consider healthy as a special fault class.

TABLE I  
THE DIFFERENT FAULTS CLASS

Part Class	Gear				Bearing		
	32T	96T	48T	80T	IS:IS	ID:IS	OS:IS
Spur 1	G	G	G	G	G	G	G
Spur 2	C	G	E	G	G	G	G
Spur 3	G	G	E	G	G	G	G
Spur 4	G	G	E	B	Ball	G	G
Spur 5	C	G	E	B	Inner	Ball	Ooter
Spur 6	G	G	G	B	Inner	Ball	Outer
Spur 7	G	G	G	G	Inner	G	G
Spur 8	G	G	G	G	G	Ball	Outer

Good (G); Broken (B); C (C); Eccentric (E); Input Shaft (IS); Output Shaft (OS); Idler Shaft (ID).

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 CNNs.

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 STFTs that can well represent the condition of rotating machinery, a feature selection method based on the performance classification is shown in table II.

in order to show the effectiveness of the approach, algorithm performance has been compared with SVM and MFCC (see [29] ) respectively. The fault detection rates on our test dataset are shown in Fig.7. As a binary classification, our algorithm has achieved ideal results. deep learning method can learn the data's structural and distributed characteristics, and fully tap the local features and global features.

In Table II, classification process by CNNs performed on the (STFTs) for features extraction. The classification ratio of this process among 88.52% until 100%.

TABLE II  
PERFORMANCES CLASSIFICATION COMPARISON

Gear	Defect	B	C	I	O
Spur	Rate %(SVM) [29]	100	100	99.28	100
Helical	Rate %(CNNs)	100	100	100	100
Spur	Rate %(SVM) [29]	88.52	93.64	95.28	90.71
Helical	Rate %(CNNs)	98.56	99.45	100	97.89

The multi-fault classifier is trained by 50 training samples, which include 8 fault classes, with 6 samples for each fault class. The faults and fault number are shown in Table I. We

compute directly using original data without preprocessing the signal to extract its features. There are 8 fault classes (include healthy class). Choosing a kernel function for classifiers has considerable impact on classification results.

However, there does not exist general rules for choosing the kernel function, the best kernel function depends on the classification problem considered. the limitation of the other technique such as support vector machine that the theory only really covers the parameters determination for a given regularisation value and kernel parameters . In a way the SVM moves the problem of over-fitting from optimizing the parameters to model selection. Sadly kernel models can be quite sensitive to over-fitting the model selection criterion [30]. In general, SVM with kernal trick is equivalent to deep learning with only one hidden unit and cannot handle multilayers of features, as deep learning does. This difference between svm and deep learning becomes important if you have a big data sets with complicated structures. for the same problem of our study the svm is limited to 10 classes for classification.

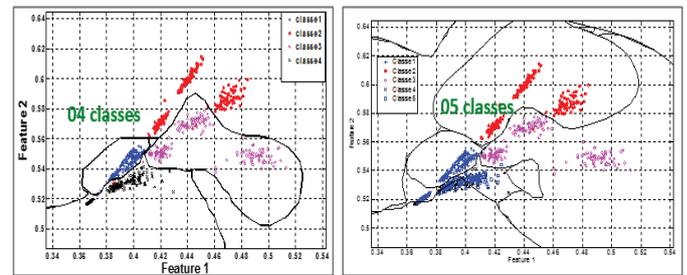


Fig. 7. Faults classification for different classes

In figure 7 shown the comparison between the proposed method. The CNNs based on STFT successfully adapted to classification task and achieved better accuracy.

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 CNNs based on STFTs is proposed; as it is a very powerful tool that can determine a good classification of the system.

To further verify the accuracy of the proposed STFTs, the experiment of testing samples of eight classes shown in 4 collected for different loading and speed. The previous trained CNNs classifier is employed to test the robustness of the feature extraction algorithm. Table II lists the classification accuracies of the four defects (B, C, I, O), respectively. As shown in the table, CNNs combining with the STFT works well in general. The recognition rate of faults detection is higher to 99%.

that is 100%. Employing the proposed STFTs features, the classification performance is generally better than using the original STFTs and the motor speed. The experiment results in

this work indicated that the CNNs based on STFTs method has better effectiveness than traditional artificial neural network. The multi-class faults classifier designed in this study has many advantages: simple algorithm, good classification and high efficiency. It is very suitable for online monitoring and diagnosis. CNNs provides us a new and useful method for developing intelligent diagnosis.

## VI. CONCLUSION

To construct an efficient monitoring system, we investigated the acceleration signal based fault detection of gearbox in this paper. An enhanced vibration feature extraction algorithm based on the popular STFT has been developed. The multi-class classification algorithm has been then built on the CNNs technique. Real vibrations signals of eight representatives fault components have been collected to test the effectiveness of the proposed algorithm. The proposed method were developed based on the acceleration signals measurements the robustness of the algorithm on vibration signals captured under different loading and speed motor has been presented. However, the automatic feature selection using CNNs is better than the other selection techniques. The real collected testing results shown that our algorithm is able to provide promising performance in faults detection gearbox equipment classification. The potential of STFT-CNNs has been highlighted for classification. Particularly, the simulation results of CNNs classier have veried that the proposed method has good efficiency in classifying eight types of defect with different characteristics.

## REFERENCES

- [1] E. Zio, *An introduction to the basics of reliability and risk analysis*. World scientific, 2007, vol. 13.
- [2] P. Baraldi, R. Canesi, E. Zio, R. Seraoui, and R. Chevalier, "Genetic algorithm-based wrapper approach for grouping condition monitoring signals of nuclear power plant components," *Integrated computer-Aided engineering*, vol. 18, no. 3, pp. 221–234, 2011.
- [3] T. Benkedjouh, N. Zerhouni, and S. Rechak, "Tool wear condition monitoring based on continuous wavelet transform and blind source separation," *The International Journal of Advanced Manufacturing Technology*, pp. 1–13, 2018.
- [4] A. Sharma, M. Amarnath, and P. Kankar, "Feature extraction and fault severity classification in ball bearings," *Journal of Vibration and Control*, vol. 22, no. 1, pp. 176–192, 2016.
- [5] D. Bordoloi and R. Tiwari, "Support vector machine based optimization of multi-fault classification of gears with evolutionary algorithms from time–frequency vibration data," *Measurement*, vol. 55, pp. 1–14, 2014.
- [6] S. Yang, J. Cao, and J. Wang, "Acoustics recognition of construction equipments based on lpcc features and svm," in *Control Conference (CCC), 2015 34th Chinese*. IEEE, 2015, pp. 3987–3991.
- [7] Z. Su, B. Tang, Z. Liu, and Y. Qin, "Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine," *Neurocomputing*, vol. 157, pp. 208–222, 2015.
- [8] A. Sadhu, G. Prakash, and S. Narasimhan, "A hybrid hidden markov model towards fault detection of rotating components," *Journal of Vibration and Control*, p. 1077546315627934, 2016.
- [9] P. Jayaswal, S. Verma, and A. Wadhvani, "Development of ebp-artificial neural network expert system for rolling element bearing fault diagnosis," *Journal of Vibration and Control*, vol. 17, no. 8, pp. 1131–1148, 2011.
- [10] B. Merainani, C. Rahmoune, D. Benazzouz, and B. Ould-Bouamama, "A novel gearbox fault feature extraction and classification using hilbert empirical wavelet transform, singular value decomposition, and som neural network," *Journal of Vibration and Control*, p. 1077546316688991, 2017.
- [11] D. Goyal and B. Pabla, "The vibration monitoring methods and signal processing techniques for structural health monitoring: A review," *Archives of Computational Methods in Engineering*, vol. 23, no. 4, pp. 585–594, 2016.
- [12] Z. Peng and F. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography," *Mechanical systems and signal processing*, vol. 18, no. 2, pp. 199–221, 2004.
- [13] M. Yuan, Y. Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using lstm neural network," in *Aircraft Utility Systems (AUS), IEEE International Conference on*. IEEE, 2016, pp. 135–140.
- [14] J. Pan, Y. Zi, J. Chen, Z. Zhou, and B. Wang, "Liftingnet: a novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 6, pp. 4973–4982, 2018.
- [15] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, 2017.
- [16] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.
- [17] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3, 2014.
- [18] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.
- [19] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen *et al.*, "Deep speech 2: End-to-end speech recognition in english and mandarin," in *International Conference on Machine Learning*, 2016, pp. 173–182.
- [20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [21] S. Kumar, A. Pandey, K. S. R. Satwik, S. Kumar, S. K. Singh, A. K. Singh, and A. Mohan, "Deep learning framework for recognition of cattle using muzzle point image pattern," *Measurement*, vol. 116, pp. 1–17, 2018.
- [22] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mechanical Systems and Signal Processing*, vol. 72, pp. 303–315, 2016.
- [23] G. S. Babu, P. Zhao, and X.-L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life," in *International conference on database systems for advanced applications*. Springer, 2016, pp. 214–228.
- [24] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring: A survey," *arXiv preprint arXiv:1612.07640*, 2016.
- [25] P. Tamilselvan, Y. Wang, and P. Wang, "Deep belief network based state classification for structural health diagnosis," in *Aerospace Conference, 2012 IEEE*. IEEE, 2012, pp. 1–11.
- [26] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Advances in neural information processing systems*, 1990, pp. 396–404.
- [27] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.
- [28] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, 2016.
- [29] T. Benkedjouh, T. Chettibi, S. Yassine, and M. Afroun, "Gearbox fault diagnosis based on mel-frequency cepstral coefficients and support vector machine," in *Computational Intelligence and Its Applications*. 6th IFIP TC 5 International Conference, CIIA 2018, Oran, Algeria, May 8–10, 2018, pp. 220–231.
- [30] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *Journal of Machine Learning Research*, vol. 11, no. Jul, pp. 2079–2107, 2010.