

Gearbox Fault Diagnosis Using Convolutional Neural Networks And Support Vector Machines

Zhuyun Chen

*School of Mechanical and Automobile Engineering
South China University of Technology
Guangzhou, China*

*Department of Mechanical Engineering
KU Leuven*

mezchen@gmail.com

Konstantinos Gryllias

*Department of Mechanical Engineering, KU Leuven
Dynamics of Mechanical and Mechatronic Systems
Flanders Make, Leuven, Belgium
konstantinos.gryllias@kuleuven.be*

Chenyu Liu

*Department of Mechanical Engineering
KU Leuven
Dynamics of Mechanical and Mechatronic Systems
Flanders Make
Leuven, Belgium
chenyu.liu@kuleuven.be*

Weihua Li

*School of Mechanical and Automobile Engineering
South China University of Technology
Guangzhou, China
whlee@scut.edu.cn*

Abstract—Fast and accurate fault diagnosis is important to ensure the reliability and the operation safety of rotating machinery, which is often based on vibration analysis. In this paper, a novel approach combining Convolutional Neural Networks (CNN) and a Support Vector Machine (SVM) classifier is proposed, in order not only to leverage upon the advantages of deep discriminative features (learnt by the CNN) but also to exploit the generalization performance of SVM classifiers. Firstly, the Continuous Wavelet Transform (CWT) is employed to obtain the pre-processed representations of raw vibration signals. Then a novel CNN with a square-pooling architecture is built to extract high-level features, without requiring extra training and fine-tuning and thus demanding reduced computation cost. Finally, a SVM is used as classifier to conduct the fault classification. Experiments are conducted on a dataset collected from a gearbox. The results demonstrate that the proposed method achieves competitive results compared to other algorithms in terms of computational cost and accuracy.

Keywords—Fault diagnosis, CNN, SVM, Gearboxes, Wavelets

I. INTRODUCTION

Rotating machinery is widely used in industry usually operating for long time under harsh conditions. Gears and bearings are key machine element components and are often the main cause of sudden breakdowns and machine failures. Therefore there is an ever increasing need for early, accurate and reliable fault identification and diagnosis techniques, focusing towards the increase of operational safety and the decrease of unexpected stops, accidents, environmental pollution and globally of the maintenance cost.

Deep Learning (DL) algorithms with multiple hidden layers, such as the Stacked Auto Encoder (SAE), Deep Belief Networks (DBN) and Convolutional Neural Networks (CNN) have been successfully used for feature extraction in a number of visual recognition and speech classification problems. Recently, DL algorithms have been proposed as potential tools to perform fault detection and diagnosis in an automatic way [1]. It has been demonstrated that the models are more effectively in learning complex function mapping compared to shallow architectures [2]. Moreover the application of CNNs, being typical Deep Neural Networks, have presented high classification accuracy in different diagnosis cases, mainly due to their local connections and their weight sharing strategy. Wang et al. constructed a CNN model for feature

extraction/learning and utilized Hidden Markov Models in order to classify rolling element bearing faults [3]. Guo et al. proposed a hierarchical CNN with an adaptive learning rate for bearing fault diagnosis [4]. Janssens et al. designed a 2D CNN with only one convolutional layer to extract/learn useful features from the frequency spectrum of raw signals improving the diagnosis performance [5]. On the other hand, time-frequency analysis using the Short-Time Fourier Transform (STFT) and the Wavelet Transform (WT) are developed for the rotating machinery fault diagnosis. Verstraete et al. explored different time-frequency methods to generate image representations of the raw signals and then applied a CNN for classification and fault diagnosis [6]. Wavelet Packet features have been built by Zhao et al. and a Deep Residual Network (DRN) has been proposed for fault diagnosis of planetary gearboxes [7]. In the abovementioned literature, CNN presented superior classification capabilities, compared to traditional Machine Learning methods, by using different types of inputs including raw time signals, spectrums, 2d representations etc. However, the training of CNN relies heavily on the Back Propagation (BP) algorithm to update the model adapting it to the new fault diagnosis task, which sometimes is an obstacle due to high computational cost. Additionally the network is sensitive to the selection of hyper-parameters such as the learning rate and the batch size. Furthermore, during the search for an optimum network architecture for a specific task, the new network needs to be retrained from scratch, requiring each time many computational resources and restraining the ability to search for more suitable architectures throughout a large space of parameters. Recently, some researchers indicated that a certain network architecture with random, untrained weights could perform well in different tasks. Pinto et al. explored thousands of potential convolutional pooling architectures and found that a network with random weights could achieve significant gains in vision systems [8]. Jarrett et al. found that in certain network architectures, a two stage system with random filters present competitive results [9]. A similar analysis can also be found in [10], which showed that certain convolutional pooling architectures can be inherently frequency selective and translation invariant, even with random weights. The architecture proposed in [10] was verified and successfully used for visual inspection and classification presenting good generalization ability [11].

Inspired by the previous work, this paper presents a new approach integrating the CNN model and a SVM classifier for vibration based gearbox fault diagnosis. The proposed method firstly exploits the Continuous Wavelet Transform (CWT) to provide good feature representations of raw vibration signals. Then a CNN with square-pooling architecture is used to extract high-level features. Finally, a SVM classifier is used for classification and fault diagnosis. By comparing the achieved results with the traditional CNN models and other diagnosis algorithms, it is shown that the proposed approach achieves competitive performance in terms of classification accuracy and computational cost. The remaining of the paper is organized as follows. In Section 2, the theory and the diagnosis procedure using the proposed method is introduced. In Section 3, a comprehensive experimental description and the result analysis are presented. Finally, some conclusions are drawn in Section 4.

II. THE PROPOSED METHOD

A. Support Vector Machines

SVM based on Vapnik's statistical learning theory is a popular supervised learning model which has been successfully applied for classification, regression and outlier detection tasks. The basic idea of applying SVM for classification tasks can be summarized in two steps: firstly, the input vectors are usually mapped into a high-dimensional feature space by kernel functions. Secondly, within the feature space, SVM attempts to seek an optimal hyperplane to divide the data into different classes. The separating hyperplane is obtained by maximizing the distance between two parallel hyperplanes.

SVM tries to seek a global optimized solution to avoid the problems of local minima, which presents good stable and generalization performance. A basic introduction is given here but a more detailed description about SVMs and their implementation can be found in [14]. In a training set $\{\mathbf{x}_i, y_i\}_{i=1}^N$, $\mathbf{x}_i \in \mathcal{R}^d$, where N is the number of samples, the samples are assumed to belong to one of two classes namely the positive class and the negative class, which can be represented by $y_i \in \{-1, +1\}$. The goal of SVM is to find an optimal hyperplane:

$$\begin{aligned} \mathbf{K}^T \mathbf{x}_i + b &\geq 1 \text{ for } y_i = +1, \\ \mathbf{K}^T \mathbf{x}_i + b &\leq -1 \text{ for } y_i = -1, \end{aligned} \quad (1)$$

where the weigh \mathbf{K} vector $\mathbf{K} \in \mathcal{R}^d$ and the b is the bias scalar. The vector \mathbf{K} and the scalar b can be used to define the position of the separating hyperplane. To find the optimal hyperplane, one can solve the following constrained optimization problem:

$$\begin{aligned} \text{Minimize } \phi(\mathbf{K}) &= \frac{1}{2} \mathbf{K}^T \mathbf{K} + C \sum_{i=1}^N \xi_i, \\ \text{Subject to: } & \\ y_i(\mathbf{K}^T \mathbf{x}_i + b) &\geq 1 - \xi_i, \quad \xi_i \geq 0, \quad i = 1, 2, \dots, n. \end{aligned} \quad (2)$$

where ξ is a slack variable. This variable measures the distance between the margin and the example data, which allows some data to be classified on the wrong side of the decision boundary. C is the user-specific penalize parameter of the error term. The constrained optimization problem can be transformed into a dual problem and can be effectively

solved by introducing Lagrange multipliers for constraints. In addition, SVM provides a unique solution to solve the nonlinear problem and according to the different classification problems, a different kernel function can be selected to obtain the optimal classification results.

B. Design of a CNN

The Topography Independent Component Analysis (TICA) network is used in order to build a CNN with a square-pooling architecture, learning feature representations from raw signals [10, 12]. A typical TICA network consists of three layers, including the input layer, the convolutional layer and the square-pooling layer, as shown in Fig. 1. The network learns to construct features in a topographical map by pooling together local groups of related features. TICA network can be regarded as a two-stage optimization procedure with Squares and Square-Root nonlinearities respectively at each stage.

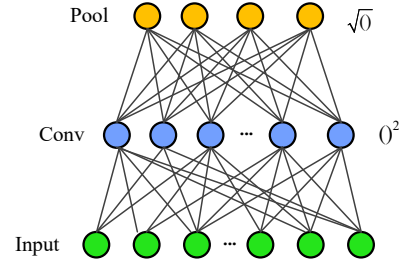


Fig. 1. The structure of TICA

In the first stage, given an input data $\{\mathbf{x}(t)\}_t^T$, the CNN firstly extracts features by computing the convolution on the raw input data with weights \mathbf{W} between the input and the hidden nodes. The weights \mathbf{W} in the convolutional layer with a set of kernels are applied on each part of the input image and then the inner product of the filter template is calculated at each location. In the second stage, the square-pooling architecture is adopted in the pooling layer. The pooling operator can be viewed as a nonlinear activation function, which includes a square-operation and a summation operation. The activations of the pooling units are equal to the sum of the squares of the units in the previous convolution layer. The weights \mathbf{V} in the second layer are the elements of the logical matrix ($V_{ij} = 0$ or 1) and are kept fixed in order to represent the topographical structure of the neurons and to create invariance to small transformations of the input in the first layer. The hidden unit output p_i refers to the pooling features, preserving important information while discarding irrelevant details by grouping the spatially neighboring neurons in a lower layer. It further receives the input signal from the layer below and returns a scalar value. Finally, the parameter \mathbf{W} can be solved by finding sparse feature representations in the second layer using the equation:

$$p_i(\mathbf{x}^t; \mathbf{W}, \mathbf{V}) = \sqrt{\sum_{k=1}^m \mathbf{V}_{ik} (\sum_{j=1}^n \mathbf{W}_{kj} \mathbf{x}_j^t)^2} \quad (3)$$

$$\begin{aligned} \min \sum_{t=1}^T \sum_{i=1}^m p_i(\mathbf{x}^{(t)}, \mathbf{W}, \mathbf{V}) \\ \text{subject to } \mathbf{W}\mathbf{W}^T = \mathbf{I} \end{aligned} \quad (4)$$

where $\mathbf{W} \in \mathcal{R}^{m \times n}$ is the input dimension, m is the number of hidden units in a layer and $\mathbf{V} \in \mathcal{R}^{m \times m}$ is a fixed matrix. Moreover $\mathbf{V}_{ik} = 1$ or 0 marks whether the pooling neuron is

connected to the convolution feature map units k or not. Furthermore, the orthonormal constraint $\mathbf{W}^T \mathbf{W} = \mathbf{I}$ forces the network to provide the diversity for the features extracting compact feature sets [11, 13].

C. Hybrid structure of the proposed CNN-SVM

In this section, a hybrid architecture combining a CNN and a SVM is constructed for the fault diagnosis of gearboxes as presented in Fig. 2. The proposed network consists of two parts: (a) the feature extraction using the CNN and (b) the fault classification using the SVM. The proposed CNN includes an input layer, a convolution layer and a square-pooling layer. Compared to the standard CNN, the proposed architecture presents several differences: (1) The traditional nonlinear activation function in standard CNN is replaced with Square and Square-Root nonlinear functions respectively in the first and the second layer. (2) The local weights of the constructed CNN are restricted by the orthogonality weights. In the proposed CNN-SVM approach, the parameters of the convolutional layer are randomly initialized and are kept fixed during the training stage, for a number of reasons listed below. Firstly, we want to explore the feature learning capability of CNN with a square-pooling architecture under the random weight initialization. Thus more attention is paid on the contribution of the intrinsic properties of the architecture alone. Secondly, some works have found that certain CNN architectures with random, untrained weights can obtain competitive results [8-11]. This result has an important practical significance highlighting the importance of selecting the most suitable architecture rather than updating the weights. In this way, the testing accuracy can be improved by selecting an appropriate structure with fast speed.

D. Fault diagnosis based on the proposed CNN-SVM

After the construction of the hybrid CNN-SVM, the gearbox fault diagnosis can be implemented based on the proposed method, which is divided into three steps. Firstly, the time-frequency representation of a vibration signal acquired with CWT is adopted for the network input. Then, the convolution layer extracts 2D structure features from the input images and preserve the relationship between the pixels. The Square-pooling layer reduces the dimensionality of each map retaining the useful information. Finally, a SVM classifier is attached for effective supervised learning and fault classification. It is noted that in the CNN architecture shown in Fig. 2, the convolutional kernel size is set equal to 7. Experiments show that a small kernel size is easily disturbed by high frequency noise while a large kernel size may lead to loss of details, reducing the accuracy. In addition, the number of the output filter is set equal to 8, which is considered high enough to achieve a high and stable accuracy in the model, as by increasing the number of the filters little improvement on the accuracy is obtained.

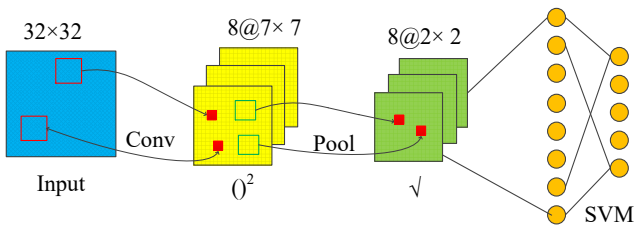


Fig. 2. The diagnosis framework of CNN-SVM model

III. EXPERIMENT DESCRIPTION AND ANALYSIS

A. Experiment description

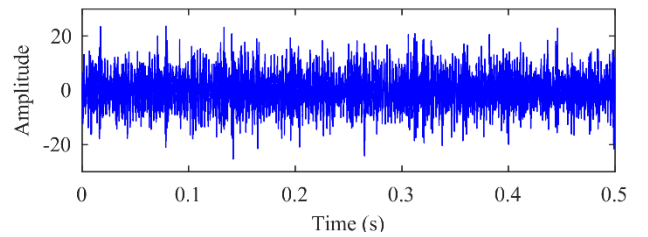
A set of gear and bearing fault experiments, conducted on an automobile five-speed gearbox, is used for validation. The test rig is shown in Fig. 3. Two severity levels of inner race bearing faults, with a depth equal to 0.2mm and 2mm are respectively introduced at the bearing of the output shaft. In addition, a minor chipped fault and a missed tooth fault have been respectively induced at the fifth-shift gear. Compound faults are also considered, containing bearing and gear faults. Totally, seven types of health conditions have been obtained. An accelerometer has been mounted on the bearing house of the output shaft to collect vibration signals with a sampling rate of 24 KHz. The output shaft load torque is set equal to 50N·m. The vibration measurements have been captured under three shaft speeds: 750rpm, 1000rpm and 1250rpm. Two time-domain signals including the gear with minor chipped tooth defect and the bearing with 0.2mm inner race defect are presented in Fig. 4. It can be observed that there are some repetitive impulsive signals captured, which could indicate the change of signal with time, but fail to localize variant and transient conditions simultaneously in time and frequency domain. In the light of its strong feature representation ability, wavelet transform is adopted at the preprocessing step to obtain good signal representation. The Morlet wavelet is used for its advantages to the detection of mechanical signals [6].



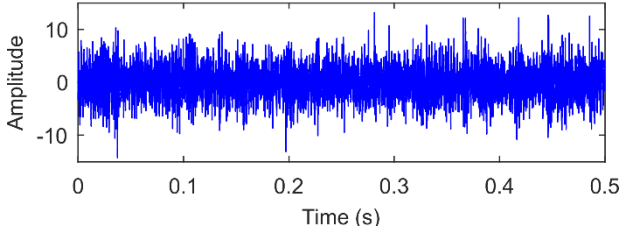
Fig. 3. The gearbox test rig

TABLE I. DESCRIPTION OF THE GEARBOX DATASET

Fault Types		Number of samples	Class Label
Gears	Bearings		
Health	Health	300	1
Minor chipped tooth	Health	300	2
Missing tooth	Health	300	3
Health	0.2mm	300	4
Minor chipped tooth	0.2mm	300	5
Missing tooth	0.2mm	300	6
Missing tooth	2mm	300	7



(a) Time domain signal of minor chipped tooth defect



(b) Time domain signal of 0.2mm inner race defect

Fig. 4. Time signals of two fault types (0.5 sec)

In order to calculate the time-frequency representations from the vibration measurement, 12000 data points (0.5 seconds) are used from the time-series signals to form one sample, allowing the capture of the local characteristics. Each sample is transformed into a magnitude scalogram using the CWT with 1024 scales. The dimension is further downsized into 32×32 using the bicubic interpolation algorithm balancing the tradeoff between computation cost and accuracy. 100 samples are generated under each speed. Therefore, 300 samples have been obtained for each health condition under three different speeds. The complete dataset, used for the algorithm verification, is composed of 2100 samples (i.e. 300 samples \times 7 health conditions), as presented in Table I. In order to verify the effectiveness of the proposed method, 30% of the samples are randomly selected for training and the rest are used for testing.

B. Analysis of feature effectiveness

The ability of the proposed CNN with random weights to extract robust features improving the classification performance is evaluated. The constructed CNN is firstly used to extract good features, which are further fed into a classifier for diagnosis. Liblinear SVM [14] is adopted to conduct the multiclass classification task and only one parameter should be selected. In addition, another typical classifier the Softmax is also utilized for verification. Ten trails are carried out to reduce the random error introduced by the parameter initialization. All the experiments have been carried out in a PC with Matlab 2017b environment running in a core Intel i7 2.8GHz CPU. The average testing accuracies of final results are listed in Table II.

TABLE II. DESCRIPTION OF THE GEARBOX DATASET

CNN with different classifier	Testing accuracy (%)	Testing deviation (%)
Softmax	99.68	0.21
SVM	99.72	0.22

Analyzing the results, it can be seen that both SVM and Softmax achieve a high accuracy and a small standard deviation. SVM achieves a 99.71% accuracy, which is slight higher than the 99.68% of the Softmax. In addition, the SVM has less parameters to be selected compared to Softmax, which provides a relative better generalization performance. Thus it is demonstrated that the extracted CNN features have strong diagnosis discriminative and generalization ability.

C. Analysis of different CNN architectures

A comparison between the constructed CNN with a square-pooling architecture and the standard CNN model is carried out in order to verify the superiority of the constructed

CNN model in feature learning capability. The standard CNN shares the same network architecture and parameters with the constructed CNN. The parameters of the convolution kernels in the standard CNN are randomly initialized without further pre-training. The samples are input into the different CNN models for feature extraction. Then the SVM is employed for the fault classification. Two different pooling schemes including the standard CNN with mean pooling and the standard CNN with max pooling are implemented for comprehensive comparison. Ten trails are carried out and the average results (Mean and Standard Deviation (SD)) are shown in Table III. It can be seen that the standard CNN with the average pooling architecture achieves a 95.24% accuracy and a large standard deviation of 2.93%. The standard CNN with the max pooling scheme receives a 97.08% accuracy with a standard deviation of 2.71%, providing better results than the average pooling operation. This is possibly due to the fact that the max pooling operation attempts to extract the most important features by taking a large local magnitude. This is similar to the operation of squaring convolutional features in the proposed method. On the other hand the average pooling takes all the low magnitudes into consideration. After the average pooling operation, the contrast of the new feature map will be reduced, leading to loss of useful discriminative features. In contrast, the proposed method achieves better results compared to the standard CNN architectures, demonstrating the superiority of the architecture design.

TABLE III. RESULT COMPARISON WITH DIFFERENT CNN ARCHITECTURES

SVM integrating CNN	Testing mean accuracy (%)	Testing SD (%)
CNN with average pooling	95.24	2.93
CNN with max pooling	97.08	2.71
The proposed method	99.72	0.22

D. Comparison with different methods

In order to further illustrate the superiority of the proposed method, three classical classification algorithms including the SVM and the standard Artificial Neural Network (ANN) and Deep Learning CNN are tested on the same datasets for comparison. (1) SVM [14]. The Liblinear SVM is adopted to conduct multiclass classification tasks. For the regularization term C , a large range of values from $\{0.001, 0.01, 0.1, 1, 10\}$ are evaluated by searching the given parameter space. (2) ANN. A three hidden layers network is constructed. The network architecture is 1024-500-200-100-7, where the input is 1024 and the hidden nodes in three layers are 500, 200 and 100, respectively. The learning rate is set to 0.1 with a batch size of 30. The network is trained with 200 epochs. (3) CNN. The same convolutional kernel and number of filters are set as at the constructed CNN. Max pooling is used in the pooling layer operation. The learning rate is set equal to 0.1. The weights in CNN are fully trained with 200 epochs. For the SVM and ANN, The time-frequency image is converted from 32×32 into one vector representation (1024 data points), which is suitable for the network input. Ten repetition experiments are conducted and the final results are averaged. The ten trails for the test result comparison of different methods are presented in Fig. 5. The average diagnosis accuracy and the computational cost of the training computations are listed in Table IV and the confusion matrix

in the first trial are also displayed in Fig. 6 for comprehensive analysis.

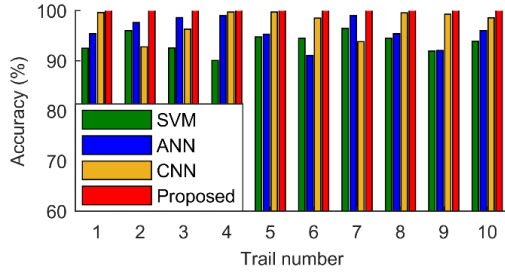


Fig. 5. Accuracy (%) for each trial for different methods

TABLE IV. THE ACCURACY AND COMPUTATION COMPARISON BETWEEN DIFFERENT METHODS

Model	Accuracy \pm STD (%)	Average Time (s)
SVM	93.87 \pm 1.95	3
ANN	95.97 \pm 2.77	43
CNN	98.30 \pm 2.62	87
Proposed	99.72 \pm 0.22	11

From the Fig. 5, Fig. 6 and Table IV, it can be seen that the traditional machine learning method SVM achieves 93.87% testing accuracy with a standard deviation of 1.95%, but requires only 3s as computation cost, which demonstrates its advantage of shallow architecture. The ANN also obtain testing accuracy of 95.97% with a large standard deviation of 2.77%. More specifically, compared to SVM and ANN, the standard DL algorithm CNN obtains 98.30% accuracy, showing the advantages of deep architecture in learning discriminative representations. However, one disadvantage is that it requires a long training time equal to 87s. In contrast, the proposed method performs better in testing accuracy and standard deviation and requires only 11s to train the model, which is a large advantage compared to standard CNN. This is mostly because the traditional CNN has many hyper-parameters needed to be carefully set i.e. learning rate, momentum and training iterations. Thus, its performance is easily influenced by the parameter selection. In addition, the model with deep architecture has many weights to be optimized during the training stage, which also increase the risk of over-fitting problem. The proposed method modifies the structure of the CNN with random weights to extract features, allowing for fast parameter selections reducing the human intervention.

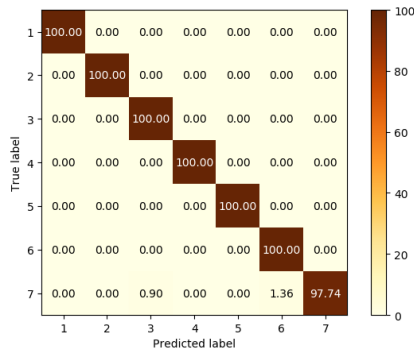


Fig. 6. The confusion matrix of the first trial

IV. CONCLUSION

A novel method for gearbox fault diagnosis based on CNN and SVM has been presented in this paper. The acquired data from an automobile experimental test rig is firstly preprocessed using the CWT to produce robust 2D feature representations. Then a novel CNN with square-pooling architecture is introduced to extract high-level abstract features. Finally, the classification process is implemented using a SVM. In this case, CNN with random weights provides high classification accuracy in two typical classifiers including SVM and Softmax. The results are much better than the traditional CNN architectures, where weights are not trained. In additional, the method also presents superiority in terms of classification accuracy and training speed compared to the standard ANN and CNN requiring extra time to update the parameters of their models.

REFERENCES

- [1] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213-237, January 2019.
- [2] 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, May 2016.
- [3] S. H. Wang, J. W. Xiang, Y. T. Zhong and Y. Q. Zhou, "Convolutional neural network-based hidden Markov models for rolling element bearing fault identification," *Knowl.-Based Syst.* Vol. 144, no. 15. pp. 65-76, March 2018.
- [4] 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, November 2016.
- [5] V. Janssens, B. Slavkovic, K. Vervisch, M. Stockman, S. Loccufier, S. S. Verstockt and Van Hoecke, "Convolutional neural network based fault detection for rotating machinery," *Journal of Sound and Vibration*, vol. 377, pp. 331-345, September 2016.
- [6] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings". *Shock and Vibration*, August 2017.
- [7] M. Zhao, M. Kang, B. Tang and M. Pecht, "Deep Residual Networks with Dynamically Weighted Wavelet Coefficients for Fault Diagnosis of Planetary Gearboxes," *IEEE Trans. Ind. Electron.*, vol. 65, pp. 1-1, October 2017.
- [8] N. Pinto and D. D. Cox, "Beyond simple features: A large-scale feature search approach to unconstrained face recognition," In *IEEE International Conference on Automatic Face and Gesture Recognition*, May 2011.
- [9] K. Jarrett, K. Kavukcuoglu and Y. LeCun, "What is the best multi-stage architecture for object recognition?," In *Computer Vision, 12th International Conference on IEEE*, pp. 2146-2153, September 2009
- [10] A. M. Saxe, P. W. Koh, Z. Chen, M. Bhand, B. Suresh and A. Y. Ng. "On Random Weights and Unsupervised Feature Learning," In *ICML*, pp. 1089-1096, June 2011.
- [11] Z. Wang and T. Oates, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," In *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*, vol. 1, January 2015.
- [12] A. Hyvarinen, J. Hurri, and P. Hoyer, *Natural Image Statistics: A probabilistic approach to early computational vision*, Springer Science & Business Media, May 2009.
- [13] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619-628, January 2018.
- [14] R. E. Fan, K. W. Chang, C. J. Hsieh, X. R. Wang, C. J. Lin, "LIBLINEAR: A library for large linear classification," *Journal of machine learning research*, pp.1871-1874, August 2018.