Robust Feature Extraction on Vibration Data under Deep-Learning Framework: An Application for Fault Identification in Rotary Machines

Ahmad Shaheryar
University of Science and Technology Beijing.
30, Xueyuan Road, Haidian District, Beijing, P.R.China.

Xu-Cheng Yin
University of Science and Technology Beijing.
30, Xueyuan Road, Haidian District, Beijing, P.R.China.

Waheed Yousuf Ramay
University of Science and Technology Beijing.
30, Xueyuan Road, Haidian District, Beijing, P.R.China.

ABSTRACT
Mechanical failures in rotating machinery (e.g. wind turbines, generators, motor-derives etc.) may result in catastrophic failures. Different mechanical faults induce characteristic vibrations in the equipment structure. Online vibration monitoring helps mitigate catastrophic failures through early detection and identification of underlying mechanical faults. However, extracting characteristic vibration features that improve fault classification performance and are robust to various noises in the vibration signals is a challenging task. Various statistical and signal processing-based vibration-features have been proposed in the literature. These vibration-features were devised on the basis of prior knowledge about characteristics of vibration signals from different fault types. Recently, automatic feature extraction through unsupervised learning in deep neural architectures has resulted in state of art performance on image and speech recognition tasks. So, Instead of feature-engineering, we, here, hypothesized that feature learning on raw vibration signal possibly will extract vibration-features that can improve fault identification performance of subsequent classifier. To the purpose, we explored Convolutional Neural Network for unsupervised feature learning on vibration signals and Denoising Auto-Encoder for extracting vibration features that are robust and invariant to the noises in vibration signals. We proposed a Hybrid deep-model consisting of a Multi-channel Convolutional Neural Network followed by a stack of Denoising Auto-Encoders (MCNN-SDAE) with a single classification layer at the top. We compared the fault identification and classification performance of the proposed model with other models employing tradition statistical and signal processing based vibration-features. We validated the performance of all models on a benchmark vibration data collected from an experimental test-rig specifically designed to study vibration characteristics of bearing related faults.

General Terms
Pattern Recognition, Fault identification and classification, Intelligent fault monitoring.

Keywords
Vibration-features learning, equipment condition monitoring, bearing fault identification, machine learning, online monitoring.

1. INTRODUCTION
Rotating machinery such as gear boxes, shafts, turbines, generators, motor-drives etc., are vulnerable to mechanical failures due to harsh conditions of operating environment and highly dynamic load changes. Online condition monitoring is essentially required to ensure safe, reliable and economical operations of these machines [1]. It helps reduce the catastrophic failures and maintenance cost through early fault detection and identification. Several online monitoring techniques had been investigated to estimate equipment condition from vibration, acoustic and process parameter data [2]. However, vibration analysis is the most known technology applied for condition monitoring of rotating equipment [3]. Shafts, couplers, bearings, gearboxes etc are the most critical parts which subject to frequent failures. Vibration signals are often adopted for their ease of acquisition and sensitivity to a wide range of faults related to rotating machinery [4].

Fault detection and identification using analytical, signal processing and statistical-based features of vibration signal are an active area of research. In case of passive-mode, the expert knowledge-based features are manually analyzed for fault identification [5]. On the contrary, the active-mode supplies the extracted features to a machine-learning based classification/fault identification model [6]. Machine learning approach is specifically interesting due to its data-driven nature and real time performance. A typical machine learning scheme involves feature extraction and learning a classifier model on vibration-features. High dimensionality and inherent noisy nature of raw vibration-data prohibits its direct use as a feature in a fault diagnostic system is. Therefore, it's essential to reduce dimensionality by extracting features from raw vibration signal that are compact without losing characteristic information. Moreover, performance of machine learning -based classifiers relies heavily on the represent ability and quality of the features extracted from raw data. In literature, several signal processing and statistical based vibration features have been proposed, examples include wavelet packet transform (WPT), Fast Fourier Transform (FFT), cepstrum information, Short Time Fourier Transform (STFT), empirical mode decomposition (EMD), time-domain statistical features (TDSF) [5]. All these feature representations have their respective strengths and limitations and are extensively reviewed by [5]. Several fault-diagnostic models have been proposed by combining aforementioned vibration-features with suitable classifiers such as SVM, ANN, BPNN, PNN, Fuzzy Inference, ANFIS, multinomial logistic regression. Example Feature-classifier combinations include WPT-BPNN/SVM/multinomial logistic regression, TDSF-ANN/SVM/MLP etc. [7-14]. These feature-classifier combinations have mostly been investigated in context of faults related to bearings, shafts, couplings and gearboxes. So, most of the features were engineered by incorporating expert-based prior-knowledge about characteristic vibration signatures related to these faults. Though, some of the engineered-features have shown success for fault diagnostic in mechanical systems that exhibit similar vibration characteristic e.g. pump, engine etc. [15][16]. However the shallow and fault-specific nature of these features limits their performance for general vibration monitoring. Vibration signals from mechanical systems e.g. rotors, turbine-engines, Aircraft frame, loose-parts, high- pressure/velocity fluid flow systems etc. exhibits characteristic vibration-features specific to the semantics and dynamics of the underlying mechanical systems. So, instead of feature-extraction the feature-learning approach deem appropriate to capture domain specific failure-features that results in high performance. The feature-learning approach alleviates dependence on prior knowledge of the problem, and proves beneficial in tasks where it is challenging to develop characterizing features.

Recently, abstract level feature extraction through unsupervised learning in deep multi-layered neural models has resulted in state of art performance on image and speech recognition tasks [17][18]. To the purpose of machinery fault-diagnostics, researchers investigated the potential of deep models to extract more abstract representations on traditional vibrations-features as well as raw-data [19]. Consequently, a variety of deep-models for vibration based equipment condition monitoring have appeared in literature. It includes Deep Belief Networks (DBN) [20-23], Auto-Encoder ELM [24], Stacked Denoising Auto-encoders (SDA) and CNN-based fault-models [25][26]. These researches suggest deep-architectures to be more effective at fault recognition than shallow ones. However, in these methods, features still need to be selected manually at first while deep-models serve as non-linear classifiers. WPT-DBN[20][22] and TDSF-DBN [27] are the two notable deep-models that employ engineered-features as base-input for bearing-fault location/severity identification and detection of early weak-faults in rolling bearing, respectively. A deep-model that can automatically extract the discriminative features from data is desired for general applicability. To the purpose, here, we proposed a deep-hybrid model composed of the Convolutional Neural Network and stacked denoising-autoencoder for unsupervised feature learning and classification on multi-channel vibration data. The proposed model is conceptualized to address the specific challenges as outlined in section 5.3. The article is organized as follow. Section 5.4 identifies the potential deep learning architectures in context of vibration-features based mechanical fault modeling and address relevant challenges thereby posed by vibration signals. Section 5.5 discusses the feature learning on vibration signals under CNN architectures. The architecture of the proposed deep-hybrid model is elaborated in section 5.6. The MCNN-SDAE model is validated on a CWRU-benchmark vibration data-set collected from an experimental test-rig that was specifically setup to study bearing fault diagnostic methods [28].

2. CHALLENGES IN VIBRATION-BASED FAULT DIAGNOSTICS

Vibration-based fault diagnosis is a challenging task especially for the case of rotating machinery. Some of the difficulties are due to inadequacy of engineered features to capture non-linear fault dynamics hidden in the vibration-data. Vibration signals are often non-stationary with different time-frequency characteristics which further complicate the feature-representation. Here we identified some key contributors to those challenges.

1- Frequency spectrum of the vibration signal is often analyzed to detect presence of bearing related faults. Bearing specific frequencies i.e. BPFO, BPFI and BPF are calculated with the assumption that the rolling elements just roll on the raceways and do not exhibit sliding behavior. However, this assumption seldom holds. In practice, a bearing roll-element undergoes a combination of rolling and sliding. Consequently, the calculated bearing-frequencies may differ from the actual frequencies by a small percentage. This rolling-slipping behavior of bearing manifests itself in the form of frequency shifts. Usually these frequency-shifts are dynamic and exhibit a non-linear behavior against different bearing-faults and their severities.

2- In case of bearing or gear faults the early vibration signals are often non-stationary and are dominated by vibrations from other components in the equipment and transmission path. So, the beneficial information in vibration signals may get distorted, thereby resulting in a reduced recognition rate. Obtaining useful information from a signal polluted by noise is essential for effective fault diagnosis methods.

3- Multiple simultaneous faults can obfuscate important frequencies

4- Interference from additional sources of vibration, i.e. bearing looseness may also obscure valuable features.

Further, the following deficiencies in classical diagnostic models limit the classification performance especially under above-mentioned challenging scenarios.

1- The features employed in the diagnostic model are manually extracted on the basis of prior knowledge about different fault types and the corresponding suitable signal processing techniques that can extract salient features to characterize the underlying faults. So the extracted features are specific to a particular diagnosis issue and might not be suitable for other fault types.

2- Many diagnostic models, reported in the literature, uses classifiers that have shallow architectures. It limits the model-ability to model complex non-linear relationships for effective fault diagnosis.

3. POTENTIAL DEEP-LEARNING ARCHITECTURES

Extracting features from vibration signals that are robust and global is a challenging task. Instead of extracting and selecting features manually, methods that can adaptively mine the distinctive features hidden in the measured signals are needed to reflect different health conditions of corresponding machinery. Deep learning [17] has the potential to address the aforementioned deficiencies in current intelligent diagnosis methods. The deep learning is outstanding in its ability to model high-level abstractions in the data by using architectures composed of multiple non-linear learning layers. It learns the discriminative features and is helpful in tasks where it is difficult to manually develop the characterizing features.

The deep-learning research has proposed several architectures e.g., Restricted-Boltzmann-Machine (RBM) and Denoising Auto-encoder (DAE) and their variants, to estimate underlying statistical structure in inputs [29][30]. These architectures have been successfully employed for unsupervised feature learning during greedy layer-wise training under deep-learning framework.

However, Convolution Neural Network are specifically interesting due to their unique ability to maintain initial information regardless of shift and distortion in the input.. The CNN-models are optimized using an error-gradient algorithm [31]. CNNs are widely used in image classification [32] speech-processing tasks [33] and various other applications [34-36]. Feature learning through CNNs have several advantages over other deep-architectures. First, hierarichal multiple Convolutional-layers can autonomously learn complex feature representations on raw input data. Second, CNNs can effectively exploit the spatial structure in the data through local receptive fields, shared filter weights and spatial sub-sampling. In case of a frequency spectrum of a vibration signal, the spatial structure is defined as the ordered sequence of frequencies. The convolution operation across frequency provides the CNN-network with immunity to small spectral shifts, such as those introduced by slipping artifiacts of the rolling-bearings. Similarly, convolution across time can be useful in capturing temporal artifacts introduced by non-stationary vibration signals. Hence, making an effective use of convolution both in time and frequency domain might be helpful in extracting robust features on vibration data and could improve the fault detection performance.

Similarly, classical autoencoders (AEs), that are trained to denoise an artificially corrupted version of their input, were found good at learning robust features on input-data.Vincent et.al.[37] further extended the classical denoising autoencoder with greedy layer-wise training procedure of deep learning algorithm that allowed stacking of multiple DA's to construct a deep-model. Building a deep-model by stacking greedily-trained classical DA's is a concise and efficient method that can extract features that are robust and invariant to noises. The architecture is interesting for learning robust features from noisy vibration signals. It can improve classification performance for the cases in which fault signature may get distorted by secondary vibration sources from other components in the transmission path.

Considering the advantages of CNN and SDAE architectures, we will investigate a hybrid deep-model for efficient fault diagnosis by extracting robust features on underlying raw vibration signal.

4. FEATURE LEARNING WITH CONVOLUTION NETWORKS

A Convolution-Neural-Network (CNN) is the variant of a standard neural-network. Contrary to the traditional neural architectures where the receptive field for input-layer neurons spans the complete input, the CNNs define a local receptive field on the input. The layers of a CNN are referred as convolution layers. The input field is logically divided into small windows which forms the localized receptive fields for subsequent convolution-layer. Units /operators/kernels in the convolution layers operate on windowed input and computes features of the local region. These convolution-units generate global representations by computing and learning local features. A CNN-layer extracts features from the input signal by convolving the input signal with the filter (or kernel) learnt by the convolution-layer. The activation of a unit in the CNN-layer represents the result of the convolution operation. The convolutional operation detects patterns captured by the kernels, regardless of where the pattern occurs, by computing the activation of a unit on different regions of the same input. In CNNs, the activation levels of kernels corresponding to subsets of classes are optimised as part of the supervised training process. A feature map is an array of units that shares the same kernel-parameterization (weight vector and bias). Their activation yields the result of the convolution of the kernel across the entire input data. The application of the convolution operator to a one-dimensional temporal sequence can be viewed as a filter, capable of removing outliers, filtering the data or acting as a feature detector that respond maximally to specific temporal sequences within the time-span of the kernel.

5. HYBRID DEEP-MODEL ARCHITECTURE

The proposed hybrid-model consists of a multichannel CNN followed by Stacked DAE architecture at the top. A CNN model is trained on individual channel to extract abstract-level Vibration-features which characterize the normal and faulty conditions in individual channels. First unsupervised pre-training is commenced through convolution auto-encoder (CAE). CNN-model parameters are initialized with convolution filters pre-trained by the CAE-model. Finally, supervised training of the CNN-model is done with labeled data. After CNN-model training, in the next step the feature-representations at top-layer of CNN-model of each channel (generated via forward pass in the CNN) are fused by learning a SDAE-model on CNN-based feature representation. The SDAE helps model the collective vibration behavior of rotary system by fusing vibrations-features extracted from multiple channels, monitoring different components or proximity in the system. A greedy layer-wise training strategy is employed to stack multiple DAE's. Each layer extracts more abstract-level representation to model non-linearity in the underlying vibration system. Finally, a fully-connected neuron-layer is trained for fault classification on the top-encoder-layer of the SDAE-model. Figure 1 depicts the proposed model architecture and corresponding processing pipeline. The proposed method is able to mine vibration-features that are robust to noise, consequently could achieve high classification performance compared to shallow architectures.

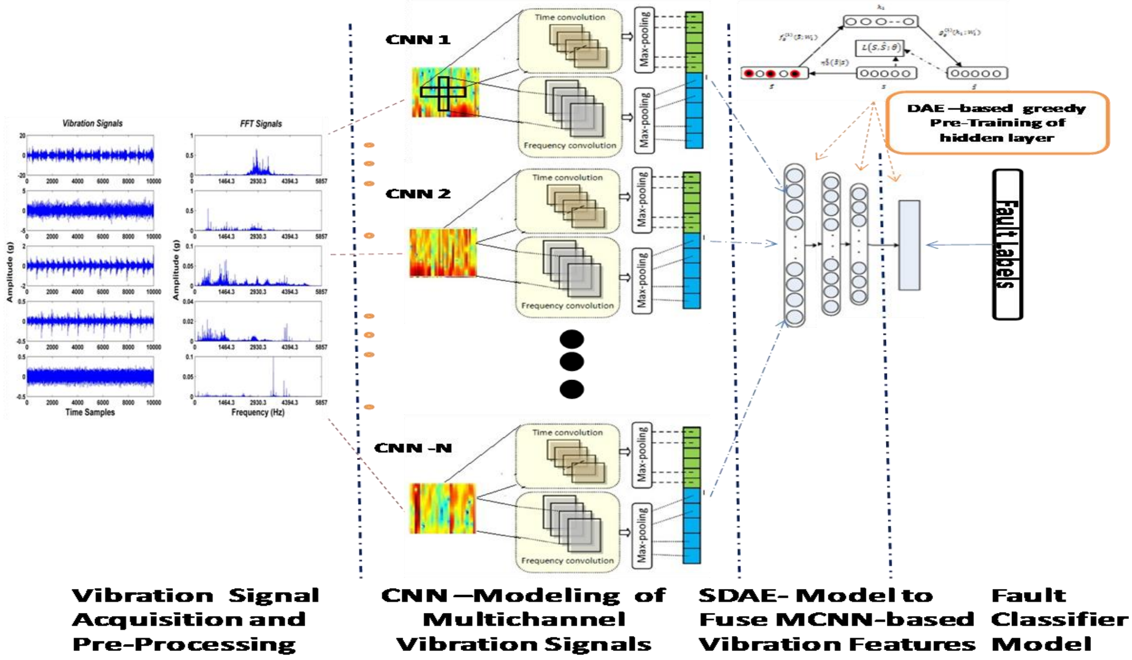


Figure 1: Hybrid MCNN-SDAE model architecture and corresponding processing pipeline.

6. MODEL SET-UP AND TRAINING

Input to the CNN-model is organized as input-maps. For the case of vibration signal , an input-map for first CNN-layer consists 10 consecutive time-frames of vibration signal to provide sufficient context in time. First the raw vibration signal is clipped into small frames with a 2 sec. time window. In the next step the log-energy is computed directly from the Short Time Fourier Spectral Coefficients that are calculated on each frame of raw vibration signal, which we will denote as STFS-features (Short Time Fourier Spectrogram features). In this way, an input-map corresponding to a vibration frame is represented as a spectrogram along with delta (first temporal derivative) and delta-delta (second temporal derivative) features. These three 2-D feature-maps are stacked to generated a single 3D input-mapthat represents input vibration signal's energy distribution along with delta and delta-delta change in energy along both frequency and time. In this case, a 2D convolution operation is defined in both frequency and time dimensions of the input map. The convolution and pooling layers apply their respective operations to generate feature representation corresponding to the input-maps of the vibration signal. Such pair of convolution and pooling layer is formally referred to as a single CNN-layer.

A multilayered CNN thus consists of two or more pairs of convolution and pooling layers. The pooling units from first CNN-layer are further organized as input-maps for next CNN-layer by following the same afore-mention input organization procedure. The equation 1 formulates the time and frequency convolution operation on input feature-map F. .

$$F_{r*s} = \begin{bmatrix} f_{1\times1}, f_{1\times2}, f_{1\times3}, \dots f_{1\times s} \\ f_{2\times1}, f_{2\times2}, f_{2\times3}, \dots f_{1\times s,} \\ . \\ . \\ . \\ f_{r\times1}, f_{r\times2}, f_{r\times3}, \dots f_{r\times s,} \end{bmatrix}$$

F_{r*s}Represent a 2D input feature-map in which each column represent a single input-frame corresponding to a particular time window. While f represents a receptive field constituting a time-frequency block of size $i \times j$. Assuming that the convolution layer has K filters then N activations corresponding to each filter generate an activation map with shared filter-parameters as follow.

$$h_{m=1,n} = \sigma\left(\sum_{n=1}^{r\times s} W_m * f_n + B_m\right) \quad (1)$$

$$A_{K\times N} = \sigma\left(\sum_{m=1}^{k} \sum_{n=1}^{r\times s} W_m * f_n + B_m\right) \quad (2)$$

Equation (1) represents a convolution operation $W_m * f_n$ by a single filter W_m over receptive-blocks f_n of the input-mapF_{r*s}. Similarly, activation maps corresponding to all filters in the filter bank can be generated via convolution in equation 2 followed by non-linear activation operator σ. Now, the pooling operation is independently applied on each of these convolution-based activation-maps. It is usually a simple function such as maximization or averaging and serves as generalizations over the features of the convolution map. The pooling size parameter determines the invariance of the convolution layer filters to small frequency shifts in the spectal representation of vibration signals. Hence, serve as small shift invariance over the local region that is determined by pooling size parameter. The max-pooling function is used as:

$$P_{i,m} = \max_{n=1}^{L} \left(h_{i,(m-1)\times s+n}\right) \quad (3)$$

Where L and s are the pooling and shift sizes, respectively. The shift size determines the overlap of the adjacent pooling block windows.

The output of a CNN-layer consists of a stack of feature-maps that are supplied to the next CNN- layer for learning higher level features on previous feature-maps. In the proposed hybrid-model an independent CNN is setup against each vibration-channel.

Further, DAE architecture is employed to model abstract level representations of a collective vibration-system by fusing CNN-layer representations corresponding to each vibration channel. It is trained to capture robust features under learning objective of clean input x reconstruction from partially corrupted or missing input $\tilde{x}$. Where input x are the top CNN-layer representations against each vibration channel. A three-layered DAE architecture in Figure 2 comprises an input, output and a hidden layer. It operates on partially corrupted input version$\tilde{x} \in [0,1]^{d}$, generated through a stochastic mapping $\tilde{x} \sim \eta_D(\tilde{x}|x)$ on clean input$x \in [0,1]^{d}$. During the learning process, the corrupted version $\tilde{x}^{d}$ is initially encoded to a hidden representation $h^{d'}$ through encoder function $f_\theta(\tilde{x})$ in Eq. (4) and then reconstructing the clean input $\hat{x}^{d}$ from hidden mapping $h^{d'}$ through the decoder function $g_{\theta'}(h)$ in Eq. (5). $\theta = \{\theta, \theta'\} = \{W, b, W', b'\}$ Parameterizes both encoder and decoder functions, where $\{W, W'\}$ and $\{b, b'\}$ are the corresponding weight and bias matrices. $\hat{x}$is an approximate reconstruction of clean input x,and $\{d, d'\}$ correspond to the input and hidden-layer dimensions, respectively.

$$h(\tilde{x}_i) = f_\theta(\tilde{x}) = \sigma(W\tilde{x}_i + b) \quad (4)$$

$$\hat{x}(\tilde{x}_i) = g_{\theta'}(h) = \sigma\left(W'h(\tilde{x}_i) + b'\right) \quad (5)$$

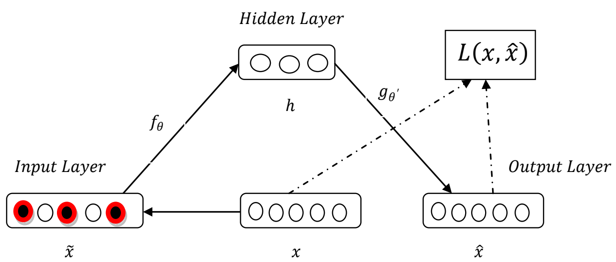
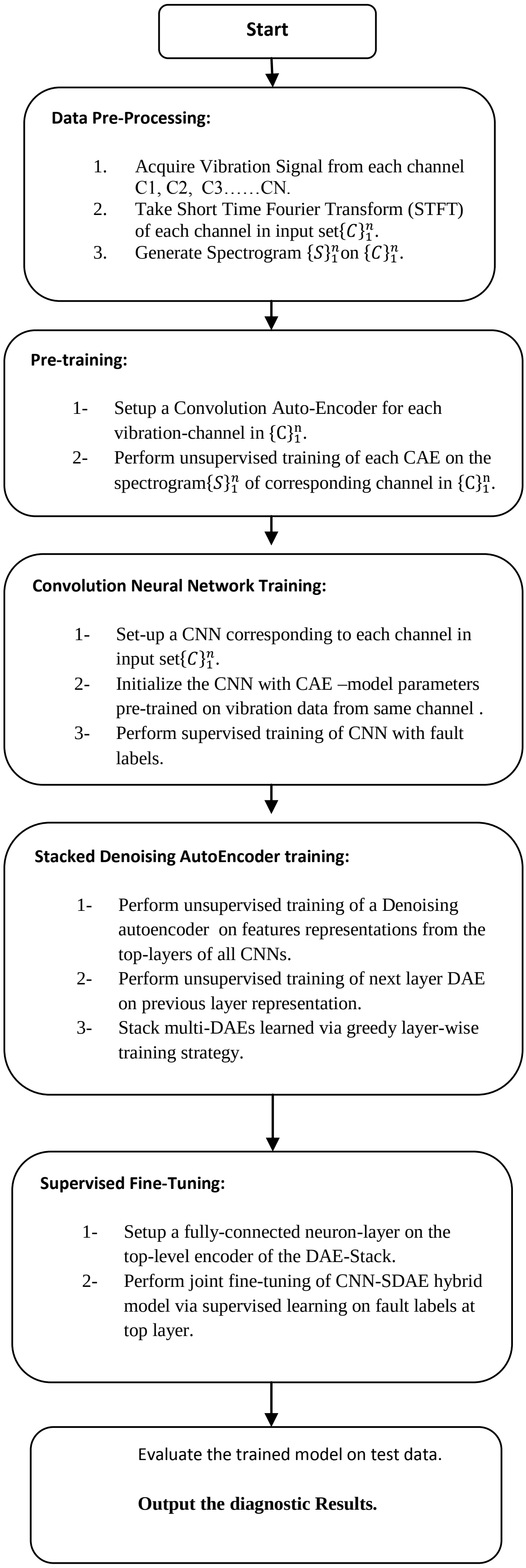


Figure 2 A typical Denoising-Auto-Encoder (DAE) setup.

Multiple DAE-layers are greedly trained on previous layer outputs/representations in an unsupervised way as depicted in Figure 2. Finally, the whole model is fine-tuned via supervised training through back propagation of the error on the final-layer classification labels at the top of SDAE (Stacked denoising Auto-encoder) structure. The Flow of hybrid-model's processing pipeline is charted as follow.



7. MODEL VALIDATION

A bearing-fault data-set provided by Case Western Reserve University [28] is used to validate the proposed model. The data-set consists of vibration signals that were collected from an experimental test-rig, as shown in Figure 3. The test-rig apparatus consist of a 2-hp motor, a torque transducer and a dynamometer. The motor shaft was supported by 6205-2RS JEM SKF bearings. The three bearing components under study are (1) the inner race (IR), (2) the outer race (OR) and (3) the rolling element, the ball (B). Single-point faults ranging in diameter from 0.007 to 0.028 inches were artificially seeded into each of the above-mentioned bearing element at both drive and fan end of the motor drive. For the faults localized to the IR, the B rolling element and the OR, the accelerometers are arranged in the dead-end position at 12 o'clock, 6 o'clock and 3 o'clock, respectively. Vibration signals from three channels were sampled at 12 kHz and in some cases at 48kHz. Figure 4 shows the vibration samples of different bearing health conditions. A dataset comprising vibration signals from healthy and three faulty bearing conditions is used for analysis. For each fault type, the vibration-data is collected against three different fault-sizes. A MCNN-SDAE model is trained on healthy and faulty training samples with parameters listed in Table 1. Different fault-types, fault-sizes and corresponding data-samples for training and testing are detailed in Table 2.

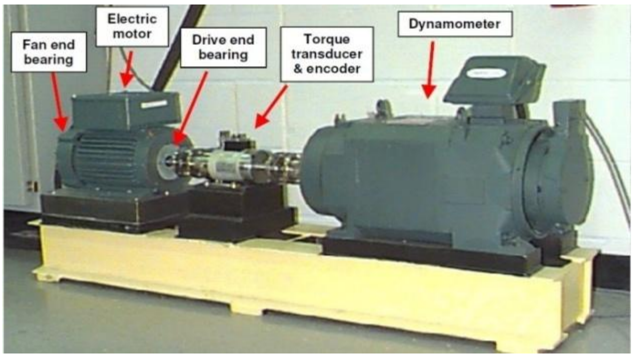


Figure 3 Experimental test-rig to collect benchmark vibration data corresponding to various bearing faults. various bearing faults.

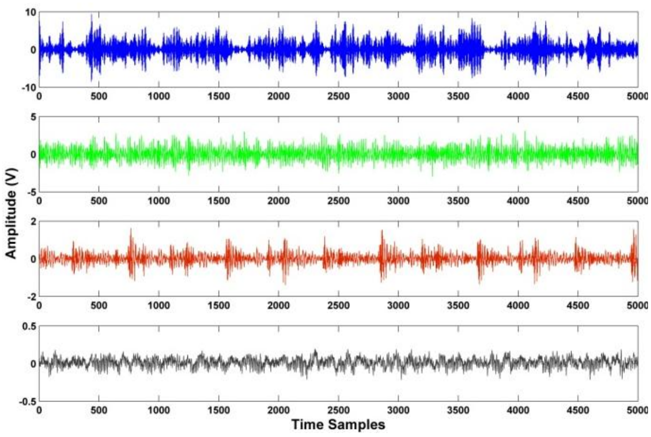


Figure 4 Vibration signals depicting different bearing fault-types.

Table 1 Hybrid-model parameters.

MCNN Parameters	
No. of CNN-layers (convolution and pooling pairs)	2
Input-map size F_{r*s} (No. of vibration-signal frames)	s=10
Block size ($i \times j$) of input receiptive field f	$f_{i \times j}$=[100x5]
Receptive Window overlapping	[50x2]
No. of Feature-Kernels/filters (W)	K=20 for layer 1 K=30 for layer 2
SDAE Parameters	
No. of layers in the Stack	3
Configuration of layers (No. of neurons)	700-500-300
Corruption-type for denoising.	Gaussian Noise (% of nominal value)= [20%]
Noise level (Corrupted Input fraction)	[15-30]%

8. CONCLUSION

We compared the bearing fault classification accuracy of the trained MCNN-SDAE with WPT-DBN[50] and TDSF-DBN[52] based deep-models as well as WPT-ANN[47], WPT-SVM[41] and TDSF-SVM based shallow models. Five-fold cross-validation procedure is used to evaluate the test accuracies of all models as reported in table 3. The classification accuracies in table 3 shows that the proposed hybrid outperformed the competitive methods by achieving an average accuracy of 99.81% and individual accuracy of 100% for inner race and ball-element faults and 99.4% in case of out-race faults, respectively.

We further validated hybrid model robustness by evaluating classification accuracies at varying noise-levels and frequency-shifts in both healthy and faulty vibration signals. We introduced an artificial noise by increasing signal to noise ratio (SNR) from 15dB-10dB and evaluated each model's average test accuracies corresponding to those noise-level. In order to validate frequency shift-invariance, the following procedure is followed.

1- Calculate ball pass frequency for inner and outer race faults (BPFO, BPFI) and ball spin frequency (BSF) for rolling-element assuming no-slip condition.
2- Take FFT of the samples from all three fault-types.
3- Identify the corresponding fault-frequency (i.e. BPFO, BPFI,BSF) in the frequency spectrum of related fault-type (e.g. inner-race, outer-race or rolling element faults).

4- Take a frequency band of 100Hz centered at corresponding fault-frequency (i.e. BPFO, BPFI, and BSF) in the frequency spectrum and shift it by offsetting it by gap of 5, 10 or 15Hz.
5- Take the inverse FFT and used the new vibration signal to calculate features corresponding to each model.
6- Calculate average classification-accuracy.

The classification-accuracies of different feature-classifier models against varying noise-levels are reported in table 4. A general trend of decrease in classification-accuracies with increasing noise-level is observed for all models. A minimum classification accuracy of 94.6% is achieved by hybrid-model against 10db SNR which is the highest accuracy among all compared models. Similarly in table 5, a maximum 1% decrease in classification accuracy of hybrid MCNN-SDAE model is observed against a shift in fault-frequency with an offset of 15Hz. However, the classification accuracy of other models decreased by 2-3% against frequency shift with 15Hz offset. The results in table 4 and table 5 show that the hybrid MCNN-SDAE model is robust to noise in vibration signals and small frequency-shifts caused by slipping of roll-bearing.

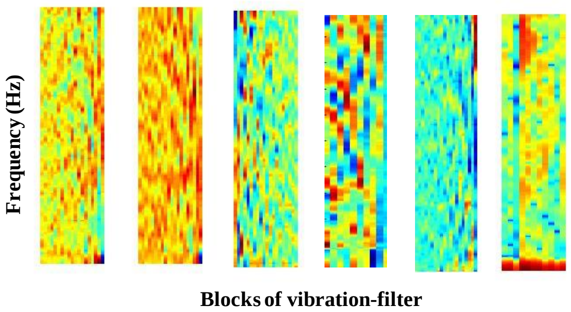


Figure 5 Each block depicts a spectral representation of first-layer CNN-filters learnt by the model. A single block represents spectogram corresponding to 20 base-filters learnt by the CNN. Each filter represents a salient vibration-pattern learnt from training-data. A particular bearing-fault is modeled as a combination of these base vibration-features via higher-order layers in MCNN-SDAE model.

9. ACKNOWLEDGMENTS

Authors acknowledges the support of China Scholarship Counsel (CSC) and china Nature Science Foundation (NSF) for supporting this research . We are thankful to Case Western Reserve University (CWRU) for providing the benchmark vibration data-set to conduct this research.

10. REFERENCES

[1] Verma, A. and Srivastava, S. 2014. Review on Condition monitoring techniques oil analysis, thermography and vibration analysis. Int. J. Enhanc. Res. Sci. Technol. Eng. **3** (2014), 18–25.

[2] Tandon, N. and Parey, A. 2006. Condition monitoring of rotary machines. Ser. Adv. Manuf. 5 (2006), 151–157.

[3] El-Thalji, I. and Jantunen, E. 2015. A summary of fault modeling and predictive health monitoring of rolling element bearings. Mechanical Systems and Signal Processing. 60-61(2015), 252–272.

[4] Lei, Y.G., Lin J., Zuo, M.J. and He, Z.J. 2014. Condition monitoring and fault diagnosis of planetary gearboxes: a review. Measurement. 48 (2014), 292–305.

[5] Boudiaf, A., Moussaoui, A., Dahane, A. et al. 2016. A Comparative Study of Various Methods of Bearing Faults Diagnosis Using the Case Western Reserve University Data. J. Fail. Anal. and Preven. 16-2 (2016), 271–284.

[6] Kankar, P., Sharma, S.C. and Harsha, S. 2011. Fault diagnosis of ball bearings using machine learning methods. Expert Systems with Applications, 38 (3) (2011), 1876–1886

[7] Weston, J. and Watkins, C. 1998. Multi-class support vector machines. Citeseer

[8] Widodo, A. and Yang, BS. 2007. Support vector machine in machine condition monitoring and fault diagnosis Mechanical Systems and Signal Processing. 21 (6) (2007), 2560–2574.

[9] Kankar, P.K., Sharma, S.C. and Harsha,S.P. 2011. Rolling element bearing fault diagnosis using wavelet transform Neurocomputing, 74 (10) (2011), 1638–1645.

[10] Konar, P. and Chattopadhyay, P. 2011. Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs) Appl. Soft Comput., 11 (6) (2011), 4203–4211.

[11] Keskes, H., Braham, A. and Lachiri, Z. 2013. Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM. Electric Power Syst. Res. 97(2013), 151–157.

[12] Liangpei, Huang, Chaowei, et al. 2015. Fault pattern recognition of rolling bearing based on wavelet packet decomposition and BP network Sci. J. Inform. Eng., 67 (1) (2015), 7–13.

[13] Saravanan, N. and Ramachandran, K.I. 2010. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). Expert Syst. Appl. 37(2010), 4168–4181.

[14] Bin, G.F., Gao, J.J., Li, X.J., et al. 2012. Early fault diagnosis of rotating machinery based on wavelet packets—empirical mode decomposition feature extraction and neural network Mech. Syst. Signal Process., 27 (1) (2012), 696–711

[15] Yang, B.S., Hwang W.W., Kim D.J., et al. 2005. Condition classification of small reciprocating compressor for refrigerators using artificial neural networks and support vector machines Mechanical Systems and Signal Processing, 19 (2005), 371–390

[16] Ding, Y., Ma, J. and Tian, Y. 2015. Health assessment and fault classification for hydraulic pump based on LR and softmax regression J. Vibroeng., 17 (2015), 1805–1816.

[17] Schmidhuber, J. 2015. Deep learning in neural networks: an overview. Neural Networks. 61(2015), 85–117.

[18] Yu, D., Deng, L., Jang, I., et al. 2011. Deep learning and its applications to signal and information processing.

IEEE Signal Processing Magazine, 28(1)(2011), 145–154.

[19] Jia, F., Lei, Y., J. Lin, et al. 2015. Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech. Syst. Signal Process. 72–73 (2015), 303–315.

[20] Gan, M., Wang, C. and Zhu, C. 2015. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. Mech. Syst. Signal Process., 72–73 (2015), 92–104.

[21] Chen, Z., Li, C. and Sánchez, R.V. 2015. Multi-layer neural network with deep belief network for gearbox fault diagnosis. J. Vibroeng. 17(2015), 2379–2392.

[22] Tamilselvan, P. and Wang, P. 2013. Failure diagnosis using deep belief learning based health state classification. Reliab. Eng. Syst. Saf. 115(2013), 124–135.

[23] Shao, H., Jiang, H., Zhang, X., et al. 2015. Rolling bearing fault diagnosis using an optimization deep belief network. Measurement Science and Technology, 26(11) (2015), Article ID 115002.

[24] Wong, P.K.., Yang, Z., Vong, C.M., et al. 2014. Real-time fault diagnosis for gas turbine generator systems using extreme learning machine. Neurocomputing . 128(2014), 249–257.

[25] Lei, Y., Jia, F., Zhou, X., et al. 2015. A deep learning-based method for machinery health monitoring with big data. Journal of Mechanical Engineering, 51(21)(2015), 49–56.

[26] Guo, X., Chen, L. and Shen, C. 2016. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement. 93 (Nov 2016), 490-502.

[27] Tao, J., Liu, Y. and Yang, D. 2016. Bearing Fault Diagnosis Based on Deep Belief Network and Multisensor Information Fusion. Shock and Vibration. Article ID 9306205.

[28] Loparo, K. A. Case Western Reserve University Bearing Data Center. Available online: http://csegroups.case.edu/bearingdatacenter/home.

[29] Larochelle, H., Bengio, Y., Lourador, J. et al. 2009. Exploring Strategies for Training Deep Neural Networks. J. Mach. Learn. Res. 10(2009), 1–40.

[30] Vincent, P., Larochelle, H., Bengio, Y., et al. June 2008 Extracting and composing robust features with denoising autoencoders. In Proceedings of the ACM International Conference, Helsinki, Finland.

[31] Lecun, Y.L., Bottou, L., Bengio, Y., et al. 1998. Gradient-based learning applied to document recognition. Proc. IEEE. 86 (11) (1998), 2278–2324.

[32] Krizhevsky, A., Sutskever, I. and Hinton G.E. 2012. ImageNet classification with deep convolutional neural networks Adv. Neural Inform. Process. Syst. 25 (2) (2012), 1106–1114.

[33] Sainath, T.N., Kingsbury, B., Saon, G., et al. 2015. Deep convolutional neural networks for large-scale speech tasks. Neural Networks. 64 (2015), 39–48.

[34] Leng, B., Guo, S., Zhang, X., et al. 2015. 3D object retrieval with stacked local convolutional autoencoder. Signal Process.112(2015), 119–128.

[35] Lee, H., Pham, P.T., Yan, L., et al. 2009. Unsupervised feature learning for audio classification using convolutional deep belief networks. Adv. Neural Inform. Process. Syst. (2009), 1096–1104

[36] Chen, X., Xiang, S., Liu, C.L., et al. 2014. Vehicle detection in satellite images by hybrid deep convolutional neural networks. IEEE Geosci. Remote Sens. Lett., 11 (2014).

[37] Vincent, P., Larochelle, H., Lajoie, I., et al. 2010. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. J. Mach. Learn. Res.11(3)(2010), 3371–3408.

11. APPENDIX

Table 2 Description of different bearing fault-types included in the benchmark data-set

Items	**Health**	**Fault 1**			**Fault 2**			**Fault 3**		
Fault Location	*None*	*Outer Race*	*Outer Race*	*Outer Race*	*Inner Race*	*Inner Race*	*Inner Race*	*Bearing*	*Bearing*	*Bearing*
Motor Speed (RPM)	*1730,1750 1772, 1797*	*1730,1750, 1772, 1797*			*1730,1750, 1772, 1797*			*1730,1750, 1772, 1797*		
Fault Size	0	0.007”	0.014”	0.024”	0.007”	0.014”	0.021”	0.007”	0.014”	0.021”
Testing Samples	50	50	50	50	50	50	50	50	50	100
Training Samples	50	50	50	50	50	50	50	50	1000	1000

Table 3 Bearing-fault classification accuracies of different models.

Health Condition	**(5-fold Cross Validated)**					
	MCNN-SDAE	***TDSF-DBN***	***WPT-DBN***	***WPT-ANN***	***WPT-SVM***	***TDSF-SVM***
Inner-Race fault	**100%**	96.2%	98.1%	95.3%	95.7%	95.6%
Outer-Race fault	**99.4%**	98.7%	97.4%	96.5%	97.6%	92.5%
Ball-fault	**100%**	99%	99.1%	98.9%	98.3%	97.4%
Average Accuracy	**99.81%**	97.96%	98.2%	96.9%	97.2%	95.1%

Table 4 Accuracies of Classifier-models tested against increasing noise-levels in vibration signal.

(5-fold Cross Validated)						
SNR (Signal to Noise Ratio)	***MCNN-SDAE***	***TDSF-DBN***	***WPT-DBN***	***WPT-ANN***	***WPT-SVM***	***TDSF-SVM***
15dB	**99.1%**	96.4%	97.1%	95%	96.2%	93.6%
14dB	**98.6%**	95.7%	96.6%	93.7%	94.9%	91.3%
13dB	**97.7%**	95%	95.8%	92.8%	93.5%	89.5%
12dB	**96.2%**	93.8%	94.1%	92%	92.2%	87.6%
11dB	**95.3%**	92.7%	93.4%	90.2%	91.1%	85.2%
10dB	**94.6%**	91.6%	92.2%	88.3%	89.4%	82.6%

Table 5 Accuracies of classifier models tested against small shifts in corresponding bearing-fault related frequencies.

(Five-fold Cross Validated)						
Spectrum Shift Offset *(BPFO,BPFI, BSF)* **+Offset**	***MCNN-SDAE***	***TDSF-DBN***	***WPT-DBN***	***WPT-ANN***	***WPT-SVM***	***TDSF-SVM***
5Hz	**99.53%**	96.83%	97.6%	96%	96.5%	95%
10Hz	**99.23%**	96.33%	96.8%	95.1%	95.3%	94.3%
15Hz	**98.7%**	95.7%	96.35%	94.3%	94.6%	92.7%