A Deep Learning-based Approach for Fault Diagnosis of Roller Element Bearings

Mohammakazem Sadoughi[1], Austin Downey[2], Garrett Bunge[3], Aditya Ranawat[4], Chao Hu[5], and Simon Laflamme[6]

[1,2,3,4,5] *Department of Mechanical Engineering, Iowa State University, Ames, IA 50011 USA*

[2,6] *Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA 50011 USA*

[5,6] *Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA*

ABSTRACT

Condition monitoring and fault detection of roller element bearings is of vital importance to ensuring safe and reliable operation of rotating machinery systems. Over the past few years, convolutional neural network (CNN) has been recognized as a useful tool for fault detection of roller element bearings. Unlike the traditional fault diagnosis approaches, CNN does not require manually extracting the fault-related features from the raw sensor data and most CNN-based fault diagnosis approaches feed the raw or shallowly pre-processed data as the training/testing inputs to a CNN model, thereby avoiding the need for manual feature extraction. As such, these approaches can be considered as purely data-driven. However, it has been proven that some well-established signal pre-processing techniques such as spectral kurtosis and envelope analysis can effectively clean and pre-process a raw signal to be a better representative of the health condition of a bearing without losing critical diagnostic information. This study proposes a new approach to bearing fault diagnosis, termed the SK-based multi-channel CNN (SCNN), that combines signal pre-processing techniques with a modified 1D CNN. The proposed SCNN approach involves two main steps: in the first step, each raw sensor signal acquired from a bearing is pre-processed to maximize the signal-to-noise ratio without losing critical diagnostic information carried by the signal; and in the second step, all pre-processed signals are fed into a 1D multi-channel CNN that classifies the health condition of the bearing. An experimental case study was carried out to evaluate the performance of the proposed approach. In this case study, a machinery fault simulator was used to validate the performance of SCNN in the presence of faults unrelated to bearings such as shaft misalignment and rotor unbalance.

1. INTRODUCTION

The monitoring and condition-based maintenance of rotating machine elements are essential to reducing operation and maintenance (O&M) costs in industry. In particular, roller element bearings experience some of the highest failure rates in both commercial and industrial applications (Liu, Bo, & Peng, 2013). An early fault in a roller element bearing can grow over time and eventually lead to an unanticipated machine failure. Thus, early fault detection and appropriate maintenance policies can greatly reduce a machine's O&M costs while also increasing machine availability through reductions in both planned and unplanned downtime (Tian, Morillo, Azarian, & Pecht, 2016).

For accurate fault detection of roller element bearings, it is generally not adequate to directly feed raw sensor signals to a fault classification algorithm; therefore, signal pre-processing is often performed to extract the desired diagnostic information from the raw signals. Numerous signal pre-processing techniques based on spectral kurtosis (SK) (Tian et al., 2016) and envelope analysis (Abboud, Antoni, Sieg-Zieba, & Eltabach, 2017) have been introduced and implemented for the detection of faults in roller element bearings.

In the traditional fault diagnosis approaches, after extracting the fault-related features from the raw signals, a machine learning technique such as support vector machine (SVM) or artificial neural network (ANN) (Konar & Chattopadhyay, 2011) is often utilized to classify the health condition of an operating bearing. In recent years, a new branch of machine learning called deep learning has been recognized as a powerful tool in bearing fault detection. In comparison with the traditional machine learning approaches, deep learning approaches are shown to be advantageous in detecting and classifying incipient bearing faults. The deep learning approaches do not require manually extracting features from the input signals. Instead, they directly use multi-channel sensor signals as the input of a deep learning-based fault clarification

model and automatically detect and learn the desired diagnostic information from the signals (Tian et al., 2016).

Convolutional neural networks (CNNs) have been recently applied to machinery fault diagnosis. Janssens et al. (Janssens et al., 2016) built a CNN model with a shallow structure, in which one convolutional layer with a wide kernel is followed by a fully connected layer, to monitor the bearing health condition using vibration signals. Zhang et al. (Zhang, Li, Peng, Chen, & Zhang, 2018) proposed a 5-layer CNN architecture with wide kernels in the first convolutional layer and narrow kernels in the following layers to detect faults in roller element bearings.

Most of the existing deep learning approaches focus on how to use less training samples to learn more diagnostic information so that these deep learning approaches can achieve higher accuracy in fault detection than the traditional machine learning approaches (Zhang et al., 2018). However, with recent advances in sensing and data acquisition, collecting large amounts of training data, often with multiple sensors, is no longer a challenging problem. It has been shown in a case study that with high volumes of training data and under controlled conditions of lab testing with low levels of noise, almost all of the traditional machine learning and newly introduced deep learning approaches can achieve satisfactory accuracy in bearing fault detection (Zhang et al., 2018).

In order to address the mentioned problems, the authors propose a new deep learning approach, namely the SK-based multi-channel CNN (SCNN), that integrates two signal pre-processing techniques with a modified 1D CNN. The modified 1D CNN takes multiple channels of sensor data as the input, which allows considering multiple sensor signals data in fault detection. Also, signal pre-processing can help maximize the signal-to-noise ratio of the input signals. The proposed SCNN approach has the following unique features:

- Considers multiple sources of information (i.e. multiple sensor signals) as the input. These multiple data streams, once pre-processed, are then fed into the CNN model simultaneously and directly, without any manual extraction of features.

- Recognizes and separates the bearing fault from other types of malfunction such as rotor unbalance and shaft misalignment.

- Filters the input signals using signal pre-processing techniques to reduce the influence of noise and improve the diagnostic accuracy and robustness.

In this study, multiple accelerometers, AE sensors, and consumer microphones were considered as the sources of diagnostic information. An experimental study that simulates the real world behavior of rotating machinery systems was carried out.

Section 2 provides a brief background review on the CNN model used in this study. The proposed SCNN approach is described in Section 3. Section 4 presents and discusses the validation results. The paper is concluded in Section 5.

2. CNN

As a multi-stage neural network, CNN consists of multiple convolutional layers, batch normalization (BN) layers, activation layers, pooling layers, and classification layers (Zhang et al., 2018). The convolutional layers convolve the inputs with a set of unknown filters called kernels and then the activation layers generate the output features from the convoluted inputs. Each kernel is a matrix with a fixed length and dimension, whose elements (or hyperparameters) are learned during the training process. The kernels are used to extract the local features of the local input region. Since the same kernels are used to convolve the input units at each layer, the number of hyperparameters in CNN is often much smaller than that in ANN and thus CNN reduces the risk of over-fitting in the training process.

Another layer of CNN is the BN layer which is designed to reduce the shift of internal covariance and accelerate the training process of CNN (Zhang et al., 2018). A BN layer is usually added right after the convolutional layer or fully-connected layer and before the activation layer. All the aforementioned operations in CNN are of a linear form. In order to enhance the ability of the network in representing the non-linearity of the input-output relation, it is essential to add an activation layer after each convolutional layer (Zhang et al., 2018). In recent years, different activation functions have been introduced. In this study, we implement the Rectified Linear Unit (ReLU) activation function since it can accelerate the convergence of CNN. Another important layer of CNN is pooling layer which acts as a down sampling operation to reduce the number of hyperparameters of the network. In this study, we use max-pooling which performs the local max operation over the input features. Finally, the classification layers (fully connected layers with softmax activation) are similar to those used in artificial neural networks.

3. PROPOSED SCNN APPROACH

As shown in Fig. 1, the proposed SCNN approach involves two main steps. The first step, called signal pre-processing, filters the raw input signals and converts them into the frequency domain. The second step feeds the pre-processed input signals into CNN. These two steps are explained in more detail as follows.

In the first step, a frequency/frequency resolution plane is built based on the fast kurtogram analysis (Antoni & Randall, 2006). Then, the subsignal with the highest kurtosis value is selected as the filtered signal. This process is essential for the proposed fault diagnosis approach since the raw in-
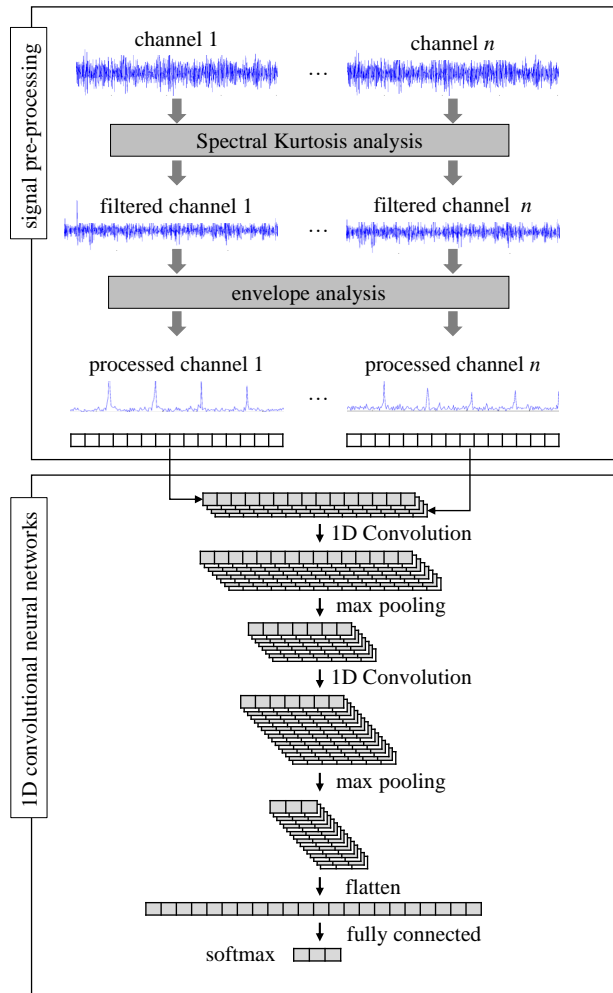
Figure 1. A typical architecture of the proposed SCNN approach.

put signals usually have high noise-to-signal ratios, which, in many cases, are impossible to control for. This signal pre-processing in the proposed SCNN approach helps minimize the noise components in the input signals. After that, the filtered signal is demodulated to remove the carrier frequencies that are unrelated to the bearing fault characteristic frequencies (Tian et al., 2016). Before being fed into the multi-channel CNN model, the pre-processed time-series signals are transformed into the frequency domain. This transformation is essential, given that SCNN considers multiple channels of sensor data as the input to the multi-channel CNN model. These different sensor signals are time series that are not perfectly synchronized (i.e. not collected at the exact same time). It means that there is an unknown time delay (in the order of milliseconds) between the measurement sample from one channel and that from another channel. Directly

feeding the raw time-series data into CNN may risk the loss of information due to the inherent non-synchronized characteristic of the time-series data.

In the second step, all the pre-processed signals (also named as processed channel 1 to channel n) are combined into a single set of data and fed into the CNN model. This step is shown in the bottom block of Fig. 1. In typical CNN models the kernel size of each convolutional layer is limited to less than 10. However, it has been shown that for the time-series data with a high sampling rate such as vibration, a wider kernel in the first convolutional layer can lead to better performance (Zhang et al., 2018). This better performance can be attributed to the fact that the convolution operation with a wider kernel can cover a longer length of inputs and thus the effect of noise on the convolution operation can be reduced. In this study, we found that the best-performing CNN model starts with wide kernels (i.e. around 10% of each input channel size) and long strides at the first stage and transitions to smaller-size kernels (i.e. down to three) and shorter stride (i.e. down to the size of 1 unit) at later stages.

4. CASE STUDY

An experimental study was performed to validate the performance of the proposed SCNN approach. This study was carried out on a machinery fault simulator to evaluate the performance of the proposed SCNN approach in detecting artificially seeded bearing faults in the presence of other types of malfunction including rotor unbalance and shaft misalignment.

4.1. Experimental setup

Two bearings were mounted on a simulator which was driven by an electric motor via a shaft (see Fig. 2 (a)). Eight sensors including four vibrations sensors (12 kHz ICP Accelerometer manufactured by PCB Piezotronics Inc), two AE sensors (100-1000 kHZ WSA sensor manufactured by Physical Acoustics), and two consumer microphones (MAX9814 manufactured by Maxim Integrated) were used to monitor the health condition of the bearings. The characteristics of the sensors are shown in table 1. The locations of bearings and sensors on the simulator are shown in Fig. 2 (b).

Four different types of bearing defect (or fault) were considered in this experimental study. These include inner race defect, outer race defect, ball defect, and combination of these defects. The machinery fault simulator ran under nine different rotation speeds from 10 Hz (600 rpm) to 30 Hz (1800 rpm) to cover a wide range of machinery operating conditions. To simulate the true behavior of the machinery systems in real world applications, two levels of rotor unbalance and three levels of shaft misalignment were also considered in the test plan. The parameters and their values used in the design of experiments are summarized in Table 2. Under each ex-
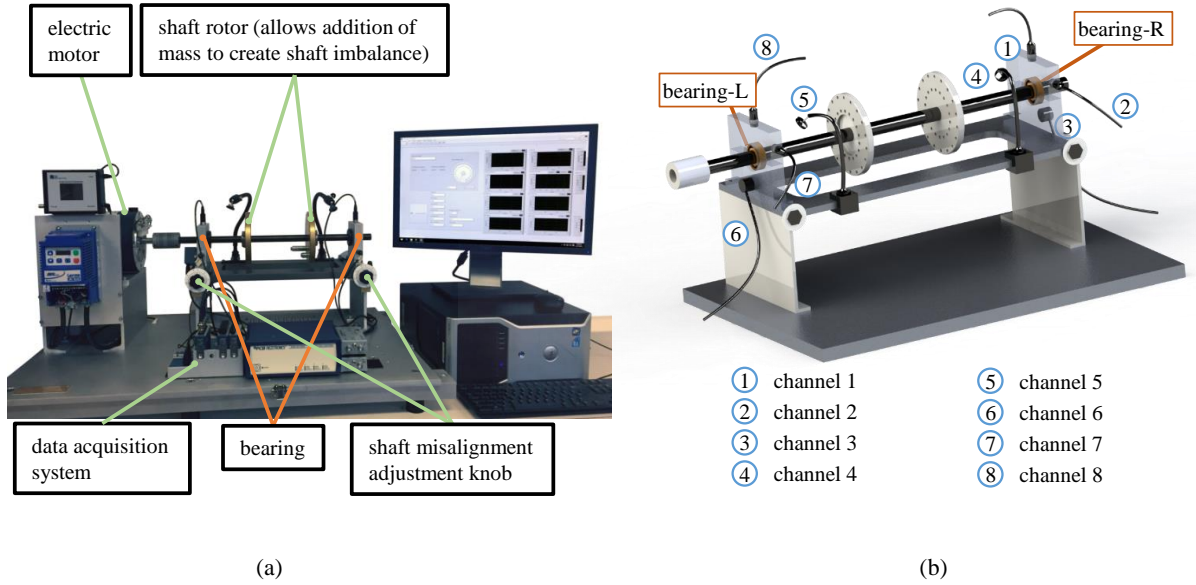
Figure 2. Machinery fault simulator: (a) experimental setup, and (b) sensors' location.

Table 1. Characteristics of sensors.

channel #	type	max frequency (kHz)	sampling rate (kHz)
1	vibration	12	100
2	vibration	12	100
3	AE	900	500
4	microphone	20	100
5	microphone	20	100
6	AE	900	500
7	vibration	12	100
8	vibration	12	100

perimental condition (i.e., unique combination of the experimental parameters), the experiment was repeated for 10 times to account for the run-to-run variability of the parameters. In total, the simulator was tested under 2,340 experimental conditions, resulting in a total of 2,340 measurement samples.

4.2. Implementation of SCNN

First the eight input channels were filtered by fast kurtogram analysis and then the power spectrum of the enveloped signals was considered as the input to the multi-channel CNN model. Each input channel had 3,000 sequenced data points corresponding to the power spectrum of the enveloped signal in the range of 0 to 500 Hz. This range was chosen to ensure that it covers the largest fault characteristic frequency of the simulator bearing.

The architecture of the multi-channel CNN used in this case study consists of four convolutional, batch normalization, and max pooling layers, followed by a fully connected layer and a softmax layer (see table 3). The kernel sizes for the first and second convolutional layers are 48 and 9, respectively, and the kernel sizes for the later two layers are both 3. The activation function is ReLU. The multi-channel CNN was implemented using Tensorflow (Abadi et al., 2016) toolbox in Python 3.6.

To evaluate the performance of the proposed SCNN approach, two classification scenarios were considered. As shown in table 4, the first scenario evaluates the performance of the propossed approach in detecting the fault in the system. The second scenario aims to test the performance in localizing the fault (i.e. detecting which bearing is defective). The number of measurement samples in each class under either classification scenario is presented in table 4.

4.3. Results

To perform a fair estimation of the accuracy of the proposed SCNN approach, we left out a set of measurement samples corresponding to one speed for use as a testing data set. The remaining samples were first shuffled and then 70% of those were used for training the CNN model and the rest were used for validating a trained model. This process was performed by treating each of the nine speeds as the testing speed. Figure 3 shows the accuracy of SCNN in fault detection and fault localization when the testing speed was varied from 10 Hz to 30 Hz. The results suggest that at all speeds, the approach is able to detect and localize the fault with high accuracy under the noisy environment and in the presence of other types of malfunction including rotor unbalance and shaft misalignment.

Figure 4 compares the diagnostic accuracy between the proposed SCNN approach and two well-known fault diagnosis approaches based on SVM and ANN. When implementing
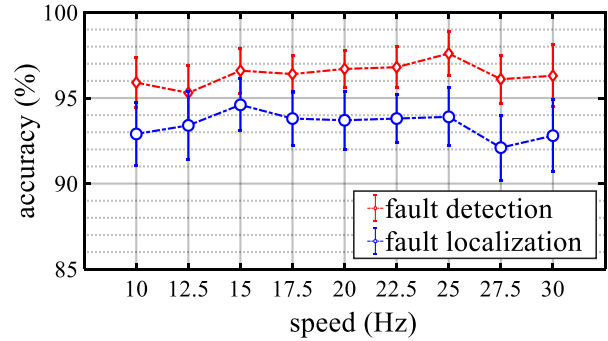

Figure 3. The accuracy of the proposed SCNN approach in predicting fault detection and localization under the varying testing speed.
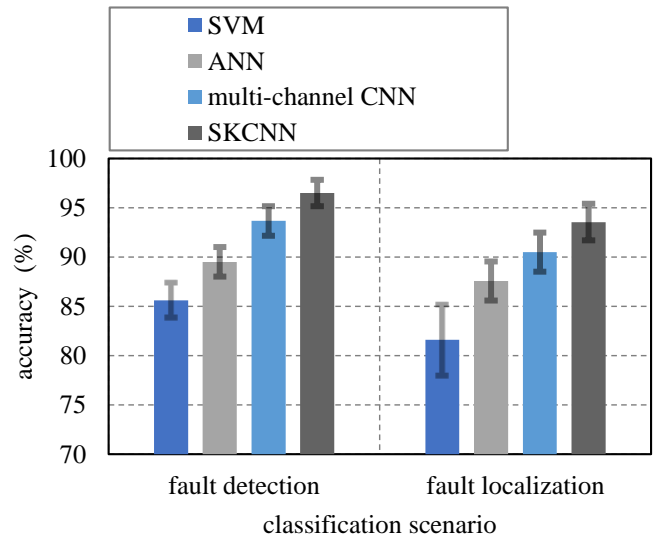

Figure 4. Accuracy of different methods in fault detection and localization.

the SVM and ANN-based approaches, we closely followed the procedure described in (Konar & Chattopadhyay, 2011). First, the signal was filtered using SK analysis and then the RMS, crest factor, and kurtosis of the enveloped signals for all eight channels were considered as the input features to the SVM and ANN models. The results are also compared with those produced by the average single-channel CNN and by the multi-channel CNN without the signal pre-processing step. For both classification scenarios, SCNN shows higher accuracy than the other four approaches.

5. CONCLUSION

In this study, a new approach to bearing fault diagnosis, termed the SK-based multichannel CNN (SCNN), is proposed which combines two signal pre-processing techniques with a modified 1D CNN that takes sensor signals from mul-

Table 2. Design of experiments.

parameter	value
shaft speed (Hz)	10, 12.5, 15, 17.5, 20, 22.5, 25, 27.5, 30
misalignment level (in)	0, 0.01
rotor unbalance (gr)	0, 5
bearing-L condition	no defect, inner race defect, outer race defect, ball defect, combination of defects
bearing-R condition	no defect, inner race defect, outer race defect, ball defect, combination of defects
trials	1, 2, 3, 4, 5, 6, 7, 8, 9, 10

Table 3. The CNN architecture adopted by the proposed SCNN approach.

	layer (type)	kernel size	number of kernels	kernel stride	output shape (width, depth)
1	convolution 1	48	8	3	151, 8
2	max pooling 1	2	8	2	75, 8
3	convolution 2	9	16	3	20, 16
4	max pooling 2	2	16	2	10, 16
5	convolution 3	3	32	1	8, 32
6	max pooling 3	2	32	2	4, 32
7	convolution 4	3	32	1	2, 32
8	global max pooling	2	32	2	64
9	fully connecter	100	1	1	100
11	softmax	2	1	-	2

Table 4. Classification scenarios.

	Class name	Number of data
Fault detection	Class 1: both bearings are healthy	540
	Class 2: at least one bearing is faulty	1800
Fault localization	Class 1: both bearings are healthy	540
	Class 2: only Bearing R is faulty	360
	Class 3: only Bearing L is faulty	360
	Class 4: both bearings are faulty	1080

tiple channels as the input. The signal pre-processing techniques allow maximizing the signal-to-noise ratio of each input signal without losing diagnostic information carried by the signal, while the multi-channel CNN enables the use of multiple types of sensor data. An experimental case study was carried out to examine the performance of SCNN in monitoring the health of multiple bearings in the real world applications. Compared to the conventional machine learning- and CNN-based approaches, SCNN is able to detect and localize the faults with higher accuracy.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., & Citro, M. (2016). Tensorflow: large-scale machine learning on heterogeneous distributed systems. arxiv preprint (2016). *arXiv preprint arXiv:1603.04467*.

Abboud, D., Antoni, J., Sieg-Zieba, S., & Eltabach, M. (2017). Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment. *Mechanical Systems and Signal Processing*, *84*, 200–226. doi: 10.1016/j.ymssp.2016.06.033

Antoni, J., & Randall, R. (2006). The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, *20*(2), 308–331. doi: 10.1016/j.ymssp.2004.09.002

Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., . . . Hoecke, S. V. (2016). Convolutional neural network based fault detection for rotating machinery. *Journal of Sound and Vibration*, *377*, 331–345. doi: 10.1016/j.jsv.2016.05.027

Konar, P., & Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and support vector machines (SVMs). *Applied Soft Computing*, *11*(6), 4203–4211. doi: 10.1016/j.asoc.2011.03.014

Liu, X., Bo, L., & Peng, C. (2013). Application of order cyclostationary demodulation to damage detection in a direct-driven wind turbine bearing. *Measurement Science and Technology*, *25*(2), 025004. doi: 10.1088/0957-0233/25/2/025004

Tian, J., Morillo, C., Azarian, M. H., & Pecht, M. (2016). Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with k-nearest neighbor distance analysis. *IEEE Transactions on Industrial Electronics*, *63*(3), 1793–1803. doi: 10.1109/tie.2015.2509913

Zhang, W., Li, C., Peng, G., Chen, Y., & Zhang, Z. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing*, *100*, 439–453. doi: 10.1016/j.ymssp.2017.06.022