Journal of Vibration and Control
XX(X):1–8
DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

FAULT DIAGNOSIS OF ROTARY MACHINES USING DEEP CONVOLUTIONAL NEURAL NETWORK WITH RAW THREE AXIS SIGNAL INPUT

Davor Kolar[1], Dragutin Lisjak[1], Michał Pająk[2] and Danijel Pavkovic[1]

Abstract

Recent trends focusing on Industry 4.0 concept and smart manufacturing arise a data-driven fault diagnosis as key topic in condition-based maintenance. Fault diagnosis is considered as an essential task in rotary machinery since possibility of an early detection and diagnosis of the faulty condition can save both time and money. Traditional data-driven techniques of fault diagnosis require signal processing for feature extraction, as they are unable to work with raw signal data, consequently leading to need for expert knowledge and human work. The emergence of deep learning architectures in condition-based maintenance promises to ensure high performance fault diagnosis while lowering necessity for expert knowledge and human work. This paper presents developed technique for deep learning-based data-driven fault diagnosis of rotary machinery. The proposed technique input raw three axis accelerometer signal as high-definition image into deep learning layers which automatically extract signal features, enabling high classification accuracy.

Keywords

Maintenance, Rotary Machinery, Fault Diagnosis, Convolutional Neural Networks, Classification

Introduction

Rotating machines in general consist of three major parts—a rotor, rolling or journal bearings (fluid or anti-friction bearings) and a foundation. Since rotary machinery usually operates under a tough working environment, it makes it more vulnerable to various types of faults and increases the complexity of fault diagnosis. A failure in rotating machinery results in not only the loss of productivity but also in the delayed delivery of goods and services to customers and may even lead to safety, economic and environmental problems. Both studies and experience show that faults develop and occur in rotating machines during normal operation. This results in a variety of failures, finally ending up in reduced availability of equipment and higher operating costs. It can be concluded that early fault detection is important, which emphasizes the necessity of maintenance in manufacturing operations. In general, maintenance function is considered as necessary cost in industry. Alternatively, by looking at rotating machinery as a profit center that produces profit only when it is running, it can be concluded that by using modern condition based maintenance strategy additional net revenue can be generated (Subramanian 2012) . Nowadays, by using predictive maintenance (i.e. Condition Based Maintenance) as a maintenance program that recommends maintenance actions based on the processed data collected through condition monitoring (Jardine et al. 2006), maintenance cost and failures can be reduced. Likewise, vibration monitoring is recognized as a leading technique for equipment condition detection and diagnostics. Vibration in any rotating machinery is caused by faults like imbalance, misalignment, crack, etc. Analyzing vibration signature is considered the most powerful predictive maintenance technique (Khanam et al. 2014; Shen et al. 2013; Elbhbah 2013), capable of capturing vibrations of the rotating machinery and presenting it in the form of the simple harmonic motion in terms of variation in the amplitude of the vibration signal.

While data acquisition as a first step (see Figure 1) of the process is well covered in both research and practice by means of still constantly improving hardware components (sensors and data acquisition systems), the second and third step are currently intensely research-active.

In recent years, many techniques for signal processing and extraction of information in fault diagnosis was titled in research, primarily focusing in improving the currently available (traditional) or developing new techniques (Walker et al. 2013) . Research in prognostic and diagnostic support for decision-making is concerned with the identification of failures and forecasting the remaining useful life

[1]Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Ivana Lucica 1, Zagreb, 10000, Croatia. Tel.: +385 1 6168 308
[2]Department of Thermal Technology, University of Technology and Humanities in Radom, Stasieckiego Street 54, Radom, 26-600, Poland. Tel.: +48 48 3617149
Emails: m.pajak@uthrad.pl (Michał Pająk), dpavkovic@fsb.hr (Danijel Pavkovic)

Corresponding author:
Davor Kolar, Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Ivana Lucica 1, Zagreb, 10000, Croatia. Tel.: +385 1 6168 308; Dragutin Lisjak, Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Ivana Lucica 1, Zagreb, 10000, Croatia. Tel.: +385 1 6168 308
Email: dkolar@fsb.hr, dlisjak@fsb.hr

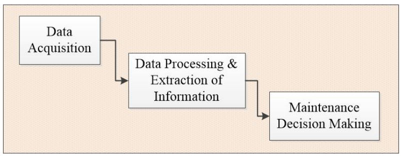

Figure 1. Fig. 1. CBM process (adopted by Lee (2004))

of equipment. Though the methods for prognostic and diagnostic support may be similar, the way of their implementation is different: prognosis is based on early failure detection, while diagnosis places greater emphasis on the determination of parameters and failure mode of the already occurring failure. Advancement in technology of measurement equipment and computing together with the increase in the number of data collected reinforces the importance of applying adequate techniques for processing collected data and thus supporting the decision-making process. Various sources categorize decision support approaches in predictive maintenance in a different way, but they can generally be divided into approaches based on physical models and approaches based on historical data. Physical model-based approach presupposes existence or creation of the digital twins of the real system, capable of simulating real behavior of the machines. Creating such models can be a demanding task, considering the possible complexity of the equipment. The development of sensor industry, communication protocols and *Industrial Internet of Things* leads to a lower price and greater availability of sensors and data acquisition and processing systems, consequently leading to the greater ability to extract knowledge from these available data. With the increase in the amount of condition data collected, it is possible to create another type of model that describes the system in operation and can provide accurate diagnosis result – data-driven models. They are becoming suitable even for the complex systems and are receiving more and more attention from the researchers and engineers. One of the most used method for data-driven fault diagnosis is machine learning, within which SVM (Shen et al. 2013; Widodo 2009), fuzzy logic (Cao et al. 2015) and finally artificial neural networks algorithms (Hwang et al. 2009; Pan et al. 2014; Zurita-Millán 2016) have been used.

It can be noticed that artificial neural networks are often used as classifiers, but as such includes prior definition of the features that need to be extracted from the data collected. Condition-characteristic features definition and extraction has great impact on the result and requires wide expert knowledge of signal processing techniques. Furthermore, defined features are most often only applicable in that scenario.

In recent years, deep learning techniques have achieved huge success in image (Chan et al. 2015; He et al. 2016) and speech (Huang et al. 2014; Noda et al. 2015) recognition.

Most recently, researchers are beginning to exploit the potentials of deep learning in fault identification and diagnostics, with the aim of reducing or eliminating the shortcomings of shallow ANN architectures. Deep learning stands for class of machine learning techniques specific by its many layers of information processing stages in deep architectures that are exploited for pattern classification and other tasks (Schmidhuber 2015) .

Convolutional Neural Network

CNNs are biologically inspired feed-forward ANNs that present a simple model for the mammalian visual cortex, that are proposed by LeCun et al. (1998a) and now widely used and virtually have become the standard in many object recognition systems in an image or video. The main reason behind such superiority lies in the configuration of CNNs. Convolutional layers use linear kernels, whose parameters are optimized during the training process.

Figure 2 illustrates a 2D CNN model with an input layer accepting 28x28 pixel images, where image size represents input layer size. Unlike classic ANNs, each neuron of the first hidden layer is not connected with all input layer neurons, yet it is associated with local receptive fields defined by size of the initially defined weight matrix (kernel) and created by sliding the local receptive field over by one neuron for each first hidden layer neuron. This means that all the neurons in the first hidden layer detect exactly the same feature, just at different locations in the input layer. Each convolution layer after the input layer alternates with a sub-sampling layer, which decimates the propagated 2D maps from the neurons of the previous layer. Unlike hand-crafted and fixed parameters of the 2D filter kernels, in CNNs they are trained by the back-propagation (BP) algorithm. However, the kernel size and the sub-sampling factor, which are set to 5 and 2, respectively, for illustration purposes in Fig. 2, are the two major parameters of the CNN. The input layer is only a passive layer which accepts an input image and assigns its (R, G, B) color channels as the feature maps of its three neurons. With forward propagation over a sufficient number of sub-sampling layers, they are decimated to a scalar at the output of the last sub-sampling layer. The following layers are identical to the hidden layers of a MLP, fully-connected and feed-forward. These so called fully-connected layers end up with the output layer that produces the decision (classification) vector.

Convolutional neural networks stand for one of the most effective deep learning architecture and has been applied to fault diagnosis. In some researches, the rotating machinery data is transformed in 2-D image format which is afterwards used for training model. Chong (2011) suggest an approach to extract features from the signal by converting it to 2-D images. Similarly, Wen et al. (2018) investigated another signal-to-2D image conversion as a step to extracting features. Further on, Shaheryar et al. (2017) explored CNN in fault identification of spectrograms of vibration images previously converted using Short Time Fourier Transform. In contrast to classification of images, raw signal data can be described as 1D multivariate time series. Most recently 1D

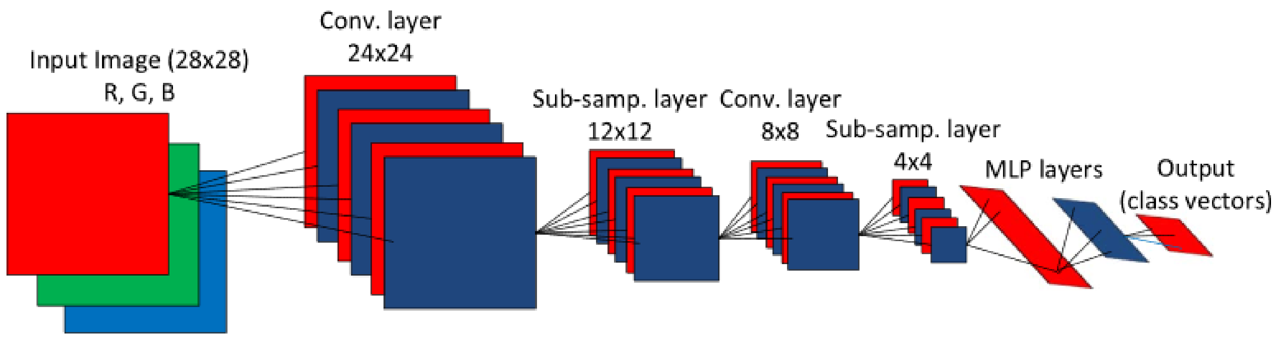

Figure 2. Fig. 1. Overview of a conventional 2D CNN (Ince et al. 2016)

CNNs are used for the classification of electrocardiogram (ECG) beats (Kiranyaz et al. 2016) achieving the state-of-the-art performance in terms of both accuracy and speed.

Abdeljaber et al. (2017) applied 1-D CNN in damage detection, while 1D CNNs in studies of (Ince et al. 2016; Shao et al. 2016) have achieved satisfactory accuracy for fault detection in induction motors. In their study, Zhang et al. (2018) introduced end-to-end solution for bearing degradation classification. Further on, Zheng et al. (2014) introduced a deep learning framework for multivariate time series classification named Multi-Channels Deep Convolutional Neural Network (MC-DCNN). However, most of the research investigated include one or more signal conversion that in the same time requires expert to perform.

Architecture of CNN for raw signal data input

Most of the data driven techniques cannot handle raw sensor data hence signal preprocessing in data driven fault identification and diagnosis is of crucial importance. Primarily, data processing aims to extract features of the raw sensor data, that can be used in model training. Extracting and identifying correct features can be difficult and requires expert knowledge. The idea of this study is to use raw time domain accelerometer signal in three axis as 3-channel image input of convolutional neural network.

In this study modified Multi-Channels Deep Convolutional Neural Network configuration is used to fuse feature extraction and learning phases of the raw accelerometer data, which can eliminate expert knowledge in vibration signal preprocessing. Multivariate raw signal data is divided into univariate in a way that each channel (signal axis) presents input in a feature learning stage. For each channel 2-stage feature learning is done and after that learned features are concatenated in a fully connected layer, as it can be seen in Figure 3. Multi-Channels Deep Convolutional Neural Network (MC-DCNN) consists of two main parts. One is a feature extractor, that is used for automatic learning features from raw data and the other is trainable fully connected MLP, which performs classification based on the features learned in the first stage. Feature extractor is composed of multiple similar stages made up of three cascading layers: filter layer, activation layer and pooling layer. The inputs and outputs of each layer are called feature maps.

Specifically, modified 2-stages MC-DCNN for failure classification is developed. Input signal consist of 3 channels and length of each input is 6400. The input (i.e., the univariate time series) is fed into a 2-stages feature extractor, which learns hierarchical features through filter (kernel), activation and pooling (sub-sampling) layers. The MC-DCNN contains two convolutional layers with alternating kernel number, each followed by max pooling layer, finally ending with fully connected layer, output unit activation function and classification layer.

The output unit activation function is the softmax function:

$$y_r(x) = \frac{\exp(a_r(x))}{\sum_{j=1}^{k} \exp(a_j(x))} \tag{1}$$

Mini-batch stochastic gradient descent with momentum based learning

The loss function for previously observed CNN is defined as cross entropy function:

$$E = -\sum_{i=1}^{n}\sum_{j=1}^{k} t_{ij} \ln y_j(x_i, \theta) \tag{2}$$

where$\theta$is the parameter vector, t_{ij} is the indicator that the *i*-th sample belongs to the *j*-th class and $y_j(x_i, \theta)$ is the output for sample *i*, respectively. The output $y_j(x_i, \theta)$ can be interpreted as as the probability that the network associates *i*-th input with class *j*, that is$P(t_j = 1|x_i)$. A full cycle of parameter updating procedure includes three cascaded phases : feedforward pass, backpropagation pass and the gradient applied (Bouvrie 2006). Widely used mini-batch gradient-based backpropagation with momentum developed by LeCun et al. (1998b) is used to minimize the loss function. (Keskar et al. 2016) found that models respond better during testing when trained on smaller batches and then update parameters. The weight w_{ij}^l is updated like stated in (3) and (4):

$$w_{ij}^l = w_{ij}^l + \triangle w_{ij}^l \tag{3}$$

$$\triangle w_{ij}^l = momentum \cdot \triangle w_{ij}^l - \varepsilon \cdot w_{ij}^l - \varepsilon \cdot \frac{\partial E}{\partial w_{ij}^l} \tag{4}$$

where w_{ij}^l represents the weight between $x_i^{l-1}$and x_i^l, $\triangle w_{ij}^l$denotes the gradient of $w_{ij}^l$and εdenotes the learning rate. The kernel weights and biases are updated in similar way as The kernel weights and biases are updated in similar way as w_{ij}^l. We set *momentum* value to 0.9 and learning

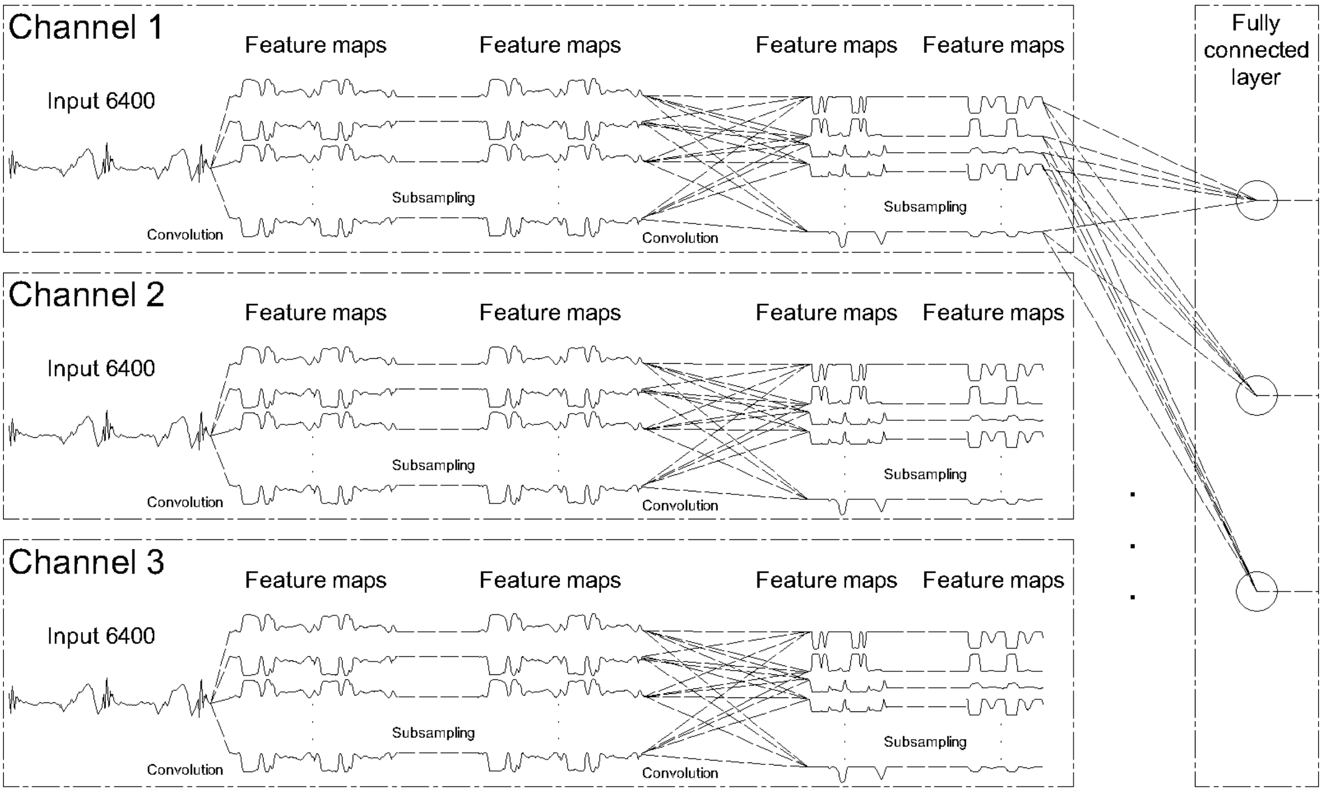

Figure 3. Two-staged modified MC-DCNN architecture with 3 channels input

rate to 0.01 with mini-batch size of 128, respectively. Additionally, learning rate is specified to multiply by factor 0.1 after each 10 epochs. Both initialization and the momentum play an important role in convolutional neural networks performance, hence future research in this field is necessary.

Experimental setup

In this study, the vibration signals acquired from a machine fault simulator are used. A SpectraQuest variable speed Machinery Fault Simulator (MFS) was used to generate both normal operation and faulty condition data. The system (illustrated in Figure 4) is comprised of a 1 HP variable speed motor driving a shaft-rotor component via coupling supported with two sets of ball bearings. The MFS is outfitted with three-axis accelerometer and a tachometer, that are connected to a National Instruments DAQ System.

Three-axis accelerometer is mounted on the bearing housing on the shaft side opposite of the motor position. The sampling frequency is set to 6,4 kHz, while revolving speed during the experiment is 1500 rpm. Vibration signals in three directions are acquired when the system operates under three different conditions. Each acquired sample of 6400 data points is stored as dataset representing state. Vibration signals under four different working conditions are used in this study, and they are divided into training and testing datasets separately, which are randomized before being used in training and testing the model. The descriptions of them are listed in Table 1.

Convolutional neural network training is done on GPU of our machine learning platform that consist of Intel i7-7700 CPU, 32GB of RAM and CUDA capable GeForce RTX 2070 graphics card with 2304 Cuda Cores and 1410 MHz base clock.

Table 1. Simulated fault conditions

No	Condition	Description
1	Normal state	Machine is running without simulated fault.
2	Debalanced rotor	Machine is running with simulated fault of imbalance on main shaft.
3	Eccentric rotor	Fault is simulated by adding eccentric rotor on main shaft.
4	Bearing fault	Machine is running with bearing outer race fault.

12 000 datasets have been collected to train and test the convolutional neural network data-driven model for failure classification.Table 2 illustrates the data composition of collected samples. From all the samples, 70% of the data is used for training and validation during training while rest of 30% is used for testing the model. 10 % of training data is used for validation during training. The samples for training, testing and validation during the experiment were selected randomly.

Results

In this section, we will discuss the diagnosis accuracy of the proposed technique for fault classification. The CNN

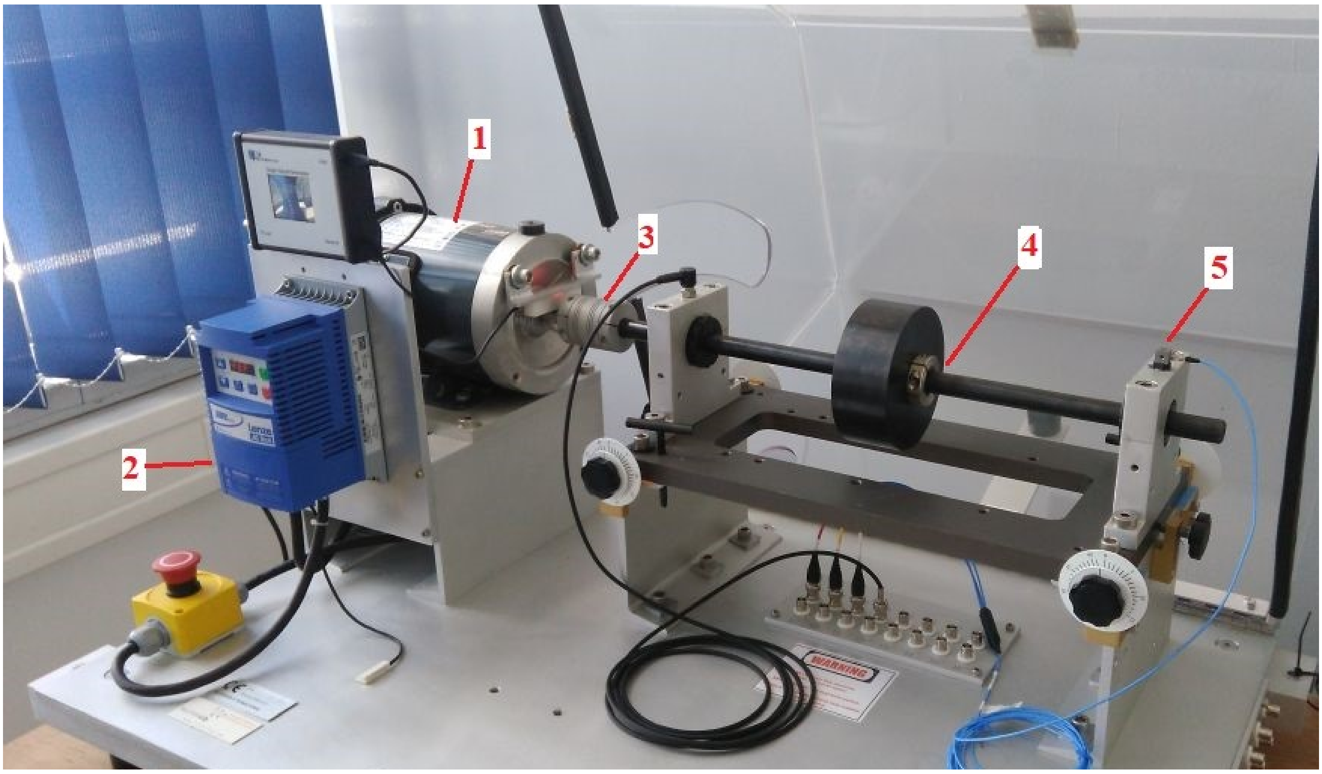

Figure 4. Fault simulator. 1 — Three-phase induction motor, 2 — Variable speed motor drive, 3 –clutch, 4 — main shaft with load 5 — 3-axis accelerometer

Table 2. Composition of collected samples for fault classification

12 000 datasets collected (76 800 000 data points)	3000 samples collected in normal working condition	2100 samples for training and validation during training (stochastic)
		900 samples for testing (stochastic)
	3000 samples collected in failure type 1: Imbalance on the main shaft	2100 samples for training and validation during training (stochastic)
		900 samples for testing (stochastic)
	3000 samples collected in failure type 2: Eccentric rotor	2100 samples for training and validation during training (stochastic)
		900 samples for testing (stochastic)
	3000 samples collected in failure type 2: Bearing fault	2100 samples for training and validation during training (stochastic)
		900 samples for testing (stochastic)

structure on this study contains two alternating convolutional and pooling layers with one fully connected layer followed by softmax and classification layer. First convolution layer uses wide kernel (31), while second kernel size is smaller (4). By using such a combination of kernels, all 6400 univariate time series samples spread across 3 channels are used for feature learning. The parameters of each layer are presented in Table 3 . First convolutional layer output consists of k_1 feature maps calculated using k_1 number of kernels, that are translated into second layer inputs. Further on, by using k_2 number of kernels, k_2 feature maps of second convolution layer are calculated. Subsampled feature maps of second convolution layer are used as fully connected layer input.

CNN_k_1-k_2 denotes that there are k_1 number of kernels in Convolutional layer 1 and k_2 number of kernels in Convolutional layer 2. There are nine models with alternating number of kernels in first and second convolutional layer. Each CNN model training run 10 times, and mean, minimum, maximum and standard deviation of the classification accuracy are the results measure terms presented in Table 4.

From the results, best average accuracy achieves CNN_24-48 with average accuracy of 99.93% and maximum achieved accuracy of 99.97% with standard deviation of 0.0506%. The best maximum accuracy is 100% achieved by CNN_24-16, while in the same time this network has also produced lowest accuracy of 99.64%, respectively. Lowest mean accuracy achieved CNN_8-48. Overall, all networks have mean accuracy equal or greater of 99.80%. Additionally, networks with higher number of kernels in first layer gained slightly better performance.

Although CNN-s are widely presented as black-box solutions and it is somewhat hard to understand inner

Table 3. Convolutional neural network parameters

Layer	Size and parameters
Input layer	Input signal: [6400 x 1 x 3]
Convolutional layer 1	k1 kernels: [31 x 1 x 3] Layer output size: 6370 x 1 x k1
Activation Layer 1	ReLU
Pooling layer 1	Max pooling [2 x 1] Layer size: 3185 x 1 x k1 Stride = 2
Convolutional layer 2	k2 kernels [4 x 1 x 16] Layer size: 3182 x 1 x k2
Activation Layer 2	ReLU
Pooling layer 2	Max pooling [2 x 1] Layer size: 1591 x 1 x k2 Stride = 2
Fully connected layer	Size: 4
Softmax	
Output layer	Classes

Table 4. Results of CNN models with different number of kernels

CNN	Mean	StDev	Max	Min
CNN_8-16	99.86%	0.07%	99.97%	99.78%
CNN_8-32	99.81%	0.0666%	99.89%	99.70%
CNN_8-48	99.80%	0.0448%	99.89%	99.72%
CNN_16-16	99.86%	0.0275%	99.89%	99.81%
CNN_16-32	99.84%	0.0492%	99.89%	99.75%
CNN_16-48	99.87%	0.0637%	99.97%	99.81%
CNN_24-16	99.86%	0.1036%	100.00%	99.64%
CNN_24-32	99.86%	0.0369%	99.92%	99.81%
CNN_24-48	99.93%	0.0506%	99.97%	99.83%

operating mechanisms, activations can be visualized. For the CNN_8-16, we plot kernels of the first and second convolutional layer for all three axes. Both Fig. 5 and Fig. 6 gives us better insights of features that are learned in first convolutional layer. Although time-domain kernels (Figure 5) are physically understandable, better visualization can be done by implementing Fast Fourier Transformation. Figure 6 presents first convolutional layer kernels learned for each axis of input signal. It is noticeable that most of the features learned for X-axis takes place in middle frequency range, Y-axis in low and medium range, while Z-axis features are extracted from medium to high frequencies. If compared to signal processing techniques, it can be concluded that first convolutional layer features present efficient frequency cut-off filters.

Further on, distribution of all test samples extracted from input signal, each convolution layer and fully connected layer for the CNN_24-48 is given Figure 7 . Visualization is done by t-SNE (Maaten and Hinton 2012) . By looking from the input layer through convolutions, it can be clearly seen features become extracted and divided as we are going to the fully connected layer, enabling high classification accuracy.

Finally, test samples feature visualization by t-SNE for networks with different number of kernels in first convolutional layer has been done to investigate how the number of kernels in first convolutional layer influence on final features division. Figure 8 presents features of fully connected layer for Conv_8-48, Conv_16-48 and Conv_24-48, respectively. Although all these three types of network achieve mean accuracy of over 99.8%, it can be seen that networks with higher number of kernels in first convolutional layer extract features in a way that they are more clustered and easily divided in fully connected layer.

Conclusion and future work

This study proposes a new CNN-based fault diagnosis technique. The main contribution of this study is developing an algorithm that input raw three-axis accelerometer signal as 1D matrix into features extractor part of convolutional neural network, that consequently automatically extract features and enable classification. When compared to traditional data-driven fault diagnosis, the omission of the need for manual extraction of features can be highlighted as the main advantage, while retaining high classification performance.

The developed CNN technique is tested on experimental data collected in Laboratory for Maintenance of University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture.

Different combinations of number of kernels in first and second convolutional layer has been investigated in order to find optimal parameters. Results shows potential of the proposed CNN technique in the data-driven fault diagnosis field, especially since vibration signals from three axis accelerometer enters model without any time consuming manual feature extraction.

Limitations of developed technique can be considered in the form of applications on real rotary machinery. Common faulty conditions must be detected and labeled for training purposes, as previously not learned faults could be misclassified. Likewise, additional testing of proposed technique on different types of failures and on known datasets is essential for performance comparison. Further on, selecting optimal hyperparameters is still a challenge. Finally, training process of developed MC-DCNN is time demanding and using GPU hardware is highly advisable. Taking that into account, future work will be based on additional testing of the technique, as well as on doing research about hyperparameter optimization.

References

Abdeljaber O, Avci O, Kiranyaz S, Gabbouj M and Inman DJ (2017) Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks,. *J. Sound Vib* 388: 154–170.

Bouvrie J (2006) Notes on Convolutional Neural Networks,.

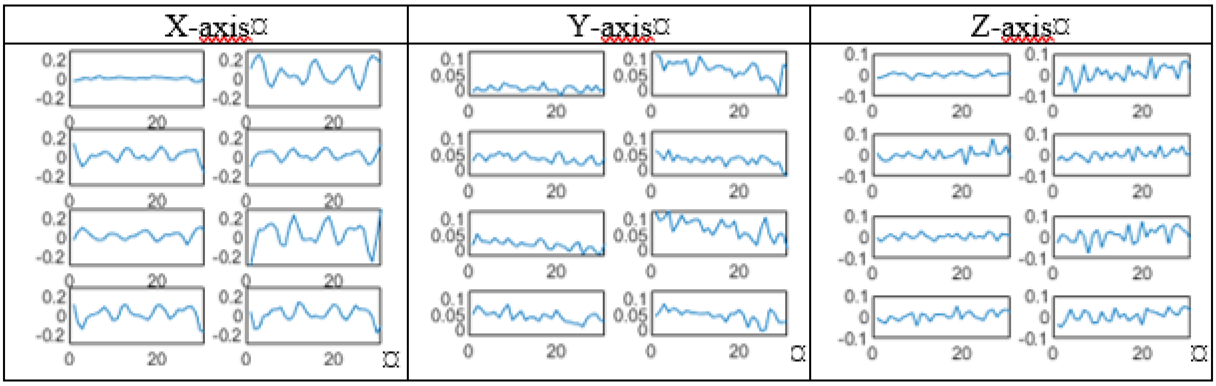

Figure 5. Kernels of first convolutional layer

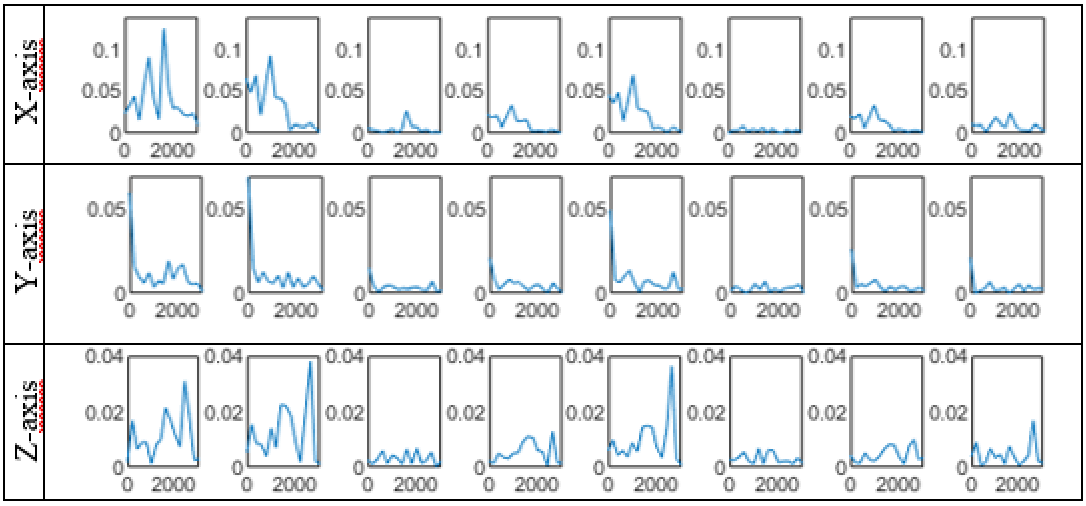

Figure 6. . 1. Frequency domain representations of kernels of first convolutional layer

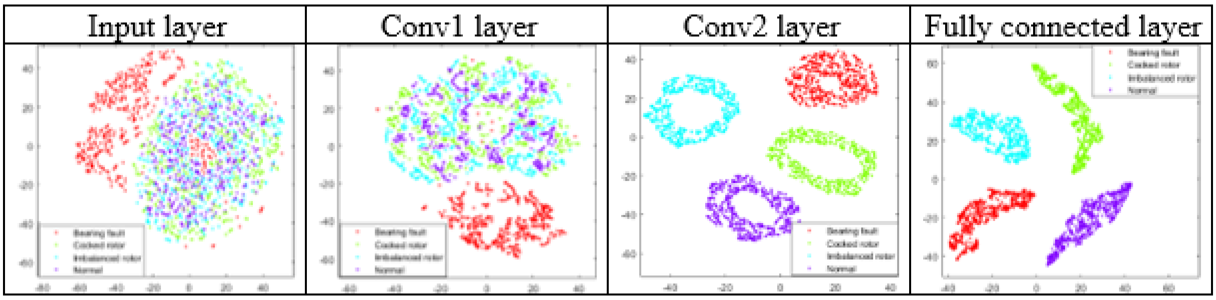

Figure 7. t-SNE feature representations

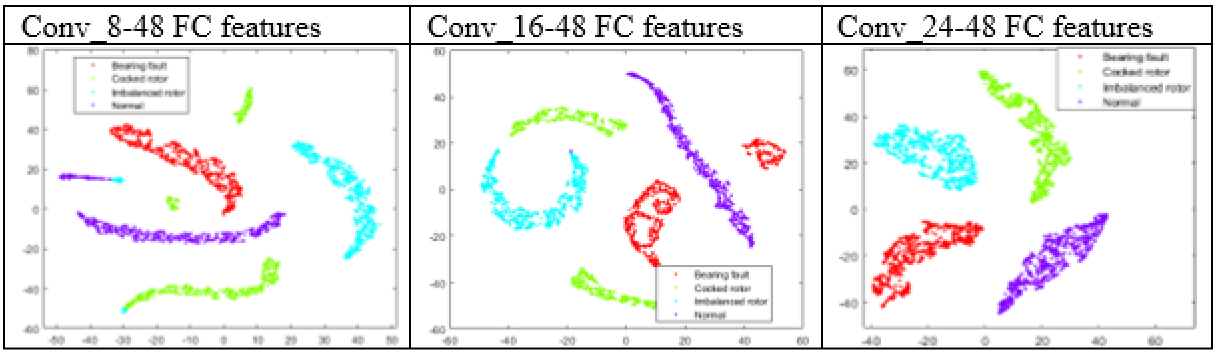

Figure 8. Fully connected layer feature visualization for three different networks

Cao SQ, Ma X, Zhang Y, Luo L and Yi F (2015) “A Fault Diagnosis Method Based on Semi -. *Supervised Fuzzy C - Means Cluster Analysis,” Int. J. Cybern. Inform* 4(2): 281–289.

Chan TH, Jia K, Gao S, Lu J, Zeng Z and Ma Y (2015) PCANet: A Simple Deep Learning Baseline for Image Classification?,. *IEEE Trans. Image Process* 24(12): 5017–5032.

Chong VTDUP (2011) Signal Model-Based Fault Detection and Diagnosis for Induction Motors Using Features of Vibration Signal in Two-Dimension Domain,. *Stroj. Vestn. – J. Mech. Eng* 57(9): 655–666.

Elbhbah JKSK (2013) A future possibility of vibration based condition monitoring of rotating machines,. *Mech. Syst. Signal Process* 34: 1–2.

He K, Zhang X, Ren S and Sun J (2016) *Deep Residual Learning for Image Recognition,*. NV, USA: IEEE.

Huang PS, Kim M, Hasegawa-Johnson M and Smaragdis P (2014) *Deep learning for monaural speech separation,*. Florence, Italy: IEEE.

Hwang YR, Jen KK and Shen YT (2009) Application of cepstrum and neural network to bearing fault detection,. *J. Mech. Sci. Technol* 23(10): 2730–2737.

Ince T, Kiranyaz S, Eren L, Askar M and Gabbouj M (2016) Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks,. *IEEE Trans. Ind. Electron* 63(11): 7067–7075.

Jardine AKS, Lin D and Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance,. *Mech. Syst. Signal Process* 20(7): 1483–1510.

Keskar NS, Mudigere D, Nocedal J, Smelyanskiy M and Tang PTP (2016) On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima,. *ArXiv160904836 Cs Math* .

Khanam S, Tandon N and Dutt JK (2014) Fault Size Estimation in the Outer Race of Ball Bearing Using Discrete Wavelet Transform of the Vibration Signal,. *Procedia Technol* 14: 12–19.

Kiranyaz S, Ince T and Gabbouj M (2016) Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks,. *IEEE Trans. Biomed. Eng* 63(3): 664–675.

LeCun Y, Bottou L, Bengio Y and Haffner P (1998a) Gradient-based learning applied to document recognition,. *Proc. IEEE* 86(11): 2278–2324.

LeCun Y, Bottou L, Orr GB and Müller KR (1998b) “Efficient BackProp,” in Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop.

Lee J (2004) *“An integrated platform for diagnostics, prognostics and maintenance optimization,” Proceeding Intell*. Maint. Syst.

Maaten and Hinton LG (2012) Visualizing non-metric similarities in multiple maps,. *Mach. Learn* 87(1): 33–55.

Noda K, Yamaguchi Y, Nakadai K, Okuno HG and Ogata T (2015) Audio-visual speech recognition using deep learning,. *Appl. Intell* 42(4): 722–737.

Pan Y, Er MJ, Li X, Yu H and Gouriveau R (2014) Machine health condition prediction via online dynamic fuzzy neural networks,. *Eng. Appl. Artif. Intell* 35: 105–113.

Schmidhuber J (2015) Deep learning in neural networks: An overview,. *Neural Netw* 61: 85–117.

Shaheryar A, Yin XC and Yousuf W (2017) Robust Feature Extraction on Vibration Data under Deep-Learning Framework: An Application for Fault Identification in Rotary Machines,. *Int. J. Comput. Appl* 167(4): 37–45.

Shao S, Sun W, Wang P, Gao RX and Yan R (2016) Learning features from vibration signals for induction motor fault diagnosis,.

Shen C, Wang D, Kong F and Tse PW (2013) Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier,. *Measurement* 46(4): 1551–1564.

Subramanian NKVTSS (2012) Cost benefit analysis of intelligent condition based maintenance of rotating machinery,.

Walker R, Perinpanayagam S and Jennions IK (2013) Rotordynamic Faults: Recent Advances in Diagnosis and Prognosis,. *Int. J. Rotating Mach* 2013: 1–12.

Wen L, Li X, Gao L and Zhang Y (2018) A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method,. *IEEE Trans. Ind. Electron* 65(7): 5990–5998.

Widodo A (2009) Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine,. *Expert Syst. Appl* 36(3): 7252–7261.

Zhang W, Li C, Peng G, Chen Y and Zhang Z (2018) A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load,. *Mech. Syst. Signal Process* 100: 439–453.

Zheng Y, Liu Q, Chen E, Ge Y and Zhao JL (2014) “Time Series Classification Using Multi-Channels Deep Convolutional Neural Networks,” in Web-Age. *Information Management* 8485: 298–310.

Zurita-Millán D (2016) Vibration Signal Forecasting on Rotating Machinery by means of Signal Decomposition and Neurofuzzy Modeling,. *Shock Vib* 2016: 1–13.