Computers, Materials & Continua CMC, vol.64, no.1, pp.515-525, 2020

Acoustic Emission Recognition Based on a Two-Streams Convolutional Neural Network

Weibo Yang[1], Weidong Liu[2, *], Jinming Liu[3] and Mingyang Zhang[4]

Abstract: The Convolutional Neural Network (CNN) is a widely used deep neural network. Compared with the shallow neural network, the CNN network has better performance and faster computing in some image recognition tasks. It can effectively avoid the problem that network training falls into local extremes. At present, CNN has been applied in many different fields, including fault diagnosis, and it has improved the level and efficiency of fault diagnosis. In this paper, a two-streams convolutional neural network (TCNN) model is proposed. Based on the short-time Fourier transform (STFT) spectral and Mel Frequency Cepstrum Coefficient (MFCC) input characteristics of two-streams acoustic emission (AE) signals, an AE signal processing and classification system is constructed and compared with the traditional recognition methods of AE signals and traditional CNN networks. The experimental results illustrate the effectiveness of the proposed model. Compared with single-stream convolutional neural network and a simple Long Short-Term Memory (LSTM) network, the performance of TCNN which combines spatial and temporal features is greatly improved, and the accuracy rate can reach 100% on the current database, which is 12% higher than that of single-stream neural network.

Keywords: Convolutional neural network, acoustic emission, fault detection.

1 Introduction

The effective classification and identification of the AE signal of the rotor are of great significance for the early diagnosis of mechanical faults, the analysis of the degree of rubbing state and the warning of fault development trends. Many scholars have proposed a number of methods to extract the robust features of the rotor's rubbing acoustic emission signal. Modal Acoustic Emission (MAE) technology derived from traditional propagation theory is an effective method for representing AE signals. It uses multi-

[1] School of Information and Communication Engineering, Nanjing Institute of Technology, Nanjing, 211167, China.

[2] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, 221000, China.

[3] School of Information Science and Engineering, Southeast University, Nanjing, 211100, China.

[4] Department of Electrical and Computer Engineering, National University of Singapore, 119077, Singapore.

[*] Corresponding Author: Weidong Liu. Email: lwdcumt@163.com.

Received: 19 January 2020; Accepted: 05 April 2020.

modal suppression to decompose AE signals into basic modal acoustic waves, and then extracts the characteristic parameters of the AE signal. For example, in 2010, Deng et al. [Deng, Cao, Tong et al. (2014)] proposed a Gaussian mixture model (GMM) rubbing fault classification method, which uses the cepstral coefficient of AE signal as input feature. On the other hand, some traditional machine learning methods are also applied in this field such as K-NN algorithm [Wang (2016)], Bayesian classifier [Baraldi, Podofillini, Mkrtchyan et al. (2015)], and support vector machine (SVM) [Vapnik (2013)]. Meanwhile, with the rapid development of deep learning, CNN is also applied in fault diagnosis [Lei, Jia, Lin et al. (2016)].

The CNN was proposed by LeCun et al. of New York University in 1989 [LeCun, Bottou, Bengio et al. (1998)]. It is a neural network mainly used to process high-dimensional mesh data. In 2012, the AlexNet won the ImageNet classification competition with overwhelming advantages, reflecting the powerful ability in image recognition and the great potential of deep learning [Krizhevsky, Sutskever and Hinton (2012)].

In recent years, a large number of excellent CNN models have been applied in many fields. Not only do they perform well on visual tasks, they also perform well on some voice tasks, such as, sound source localization [Zhou, Wang, Chen et al. (2019)]. Similarly, we can also use the CNN model to deal with the task of AE.

At present, CNN has been initially applied in fault diagnosis and has improved the level and efficiency of fault diagnosis. For instance, Prosvirin et al. propose a method using CNN and AE signal to extract discriminative to detect bearing faults [Prosvirin, Kim and Kim (2017)]. At present, multi-streams neural networks are also frequently used, compared with a single-stream neural network, a multi-streams neural network can extract more features and has a better performance in aspects such as facial expression recognition [Khor, See, Phan et al. (2018)], fault detection [Li, Li, Qu et al. (2019)].

In this paper, an improved neural network is used to effectively identify the AE signal. Firstly, the time-frequency analysis of the AE signal is calculated, which reflects the frequency of AE signals as a function of time. Secondly, the STFT and MFCC are calculated to construct two-streams input data. Then, a two-streams neural network is proposed, in which one stream is used to extract spatial features through CNN and the other stream is used to extract temporal features through CNN-LSTM. Finally, the effectiveness of this TCNN network proposed in this paper is verified by experiments.

2 Main principle

2.1 CNN

CNN are improvements to traditional common neural networks. For neural networks that take images as input, each pixel of the image can be viewed as a feature value. There will be too many parameters when a fully connected network connection is used. The CNN implements the functions of local feature extraction and hierarchical feature extraction by introducing the convolutional layer and the pooling layer, and reduces the number of parameters of the entire network by using the network weight parameter sharing mechanism. The CNN is mainly composed of an input layer, a convolution layer, a pooling layer, a fully connected layer and an output layer, which are shown in Fig. 1.

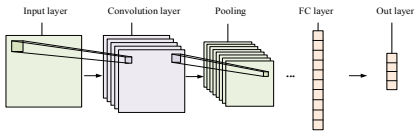


Figure 1: The structure of CNN

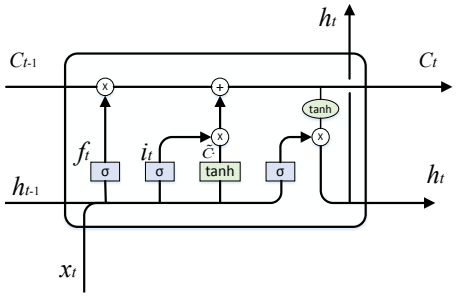
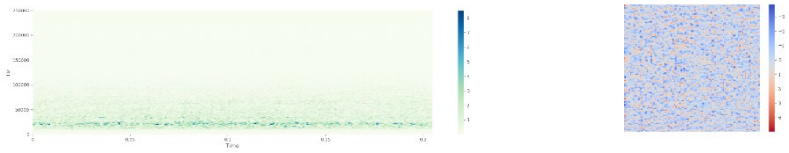
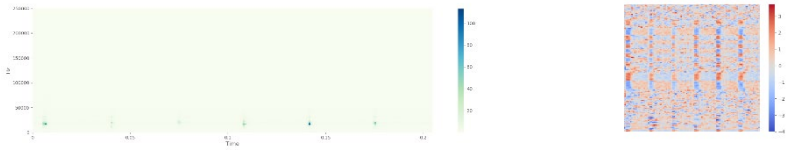


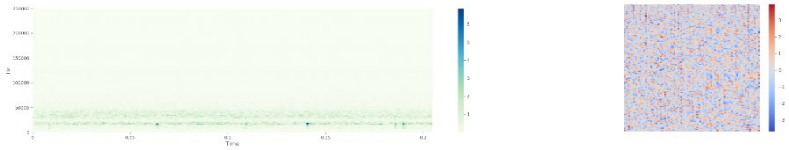
Figure 2: The structure of LSTM



(a)



(b)



(c)

Figure 3: STFT spectra and MFCC of rubbing, cracking, and normal AE signals. (a) rubbing. (b) Cracking. (c) Normal

2.2 LSTM

LSTM is a kind of time circulation neural network, which is specially designed to solve the long-term dependence problem existing in general RNN. Here, we use LSTM to extract time features. The structure of LSTM is shown in Fig. 2.

2.3 Two-streams input

Rotating machine rotor AE signal is a kind of acoustic signal, which acoustic signal characteristics is similar to natural speech [Kundu, Kishore and Sinha (2009)], so that it can be analyzed and identified with reference to the processing method of a speech signal. Here, the rubbing AE signal can be regarded as short-term stationery, that is, for any time t, the signal can be spectrally analyzed in a small range of time near the moment. A two-dimensional spectrogram of the AE signal can be obtained by performing a continuous spectrum analysis on a series of t values. Fig. 3 shows the STFT spectra and MFCC spectra of three AE signals for normal, cracking, and rubbing.

In STFT, each column represents a 512 points FFT of one frame of signal, each frame has a duration of 1.024 ms and a frame overlap rate of 0.5. Considering the relatively stable and unitary frequency distribution of the AE signal STFT and MFCC spectrum and CNN's powerful image learning classification ability, two-streams input is designed for the convolutional neural network to improve the model accuracy. Compared with single-stream input, two-streams inputs are sufficient to preserve more valid features of the output. The two-streams inputs are:

(1) STFT amplitude spectrum

(2) MFCC amplitude spectrum

Here, the STFT and MFCC amplitude spectrum are used to extract the variation characteristics of the AE signal spectrogram, such as some image edge features.

2.4 Two-streams CNN

Due to the poor performance of the small database on the deep network, compared with the deep network like ResNets [He, Zhang, Ren et al. (2016)] and VGGNet [Szegedy, Liu, Jia et al. (2015)], we chose the CNN network with few layers. Meanwhile, in order to prevent over-fitting, we added dropout behind each max-pooling layer. The CNN structure used in this paper is shown in Fig. 4. And the details of this structure are shown in Tab. 1. At the same time, in order to obtain the time features, we add LSTM [Hochreiter and Schmidhuber (1997)] structure after the CNN of the stream which input is MFCC.

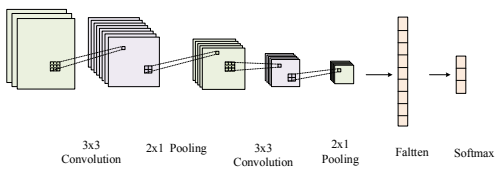


Figure 4: The CNN structure

Table 1: The model parameter setting of the CNN

No.	Layer	Dropout	Kernel	Channel	Activation
1	Convolution	-	(3, 3)	32	Rectified Linear Unit (ReLu)
2	MaxPooling	-	(2, 2)	-	-
3	Dropout	0.5	-	-	-
4	Convolution	-	(3, 3)	32	ReLu
5	MaxPooling	-	(2, 2)	-	-
6	Dropout	0.5	-	-	-
7	Dense	-	-	-	Softmax

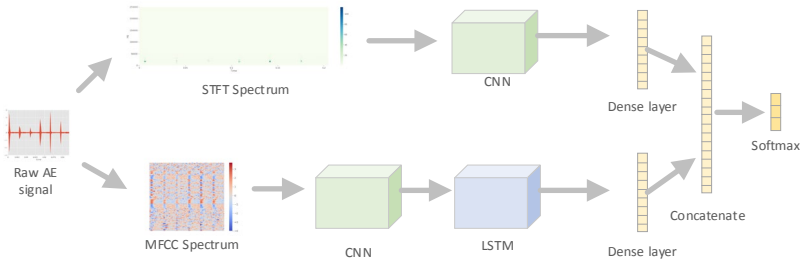


Figure 5: The overall framework

The input data of the network is the images of a rubbing AE signal STFT spectrum and MFCC spectrum with a frame overlap ratio of 50%, the size of the STFT spectrum is 257×200×1, while the size of the MFCC spectrum is 128×100×1.

The whole TCNN framework is shown in Fig. 5.

The STFT spectrum and MFCC spectrum of the raw AE signal are obtained by preprocessing. These spectrums are used to represent the feature of AE signals. Then the two spectrums were input into two CNN networks of the same structure. The next step is to input the output of the MFCC stream into the LSTM model. After that, we concatenate the last dense layer of the two streams, and connect the result with a softmax layer.

3 Experiments

3.1 Database

The AE signal database adopted in this paper is a database of normal signals, rotor cracks and rotor rubbing AE signals composed of the AE signals of rotating machinery independently collected by our research group of the laboratory in the past two years. In total, the database contains AE signals under three different rotational speed conditions (600 rad/s, 700 rad/s, 800 rad/s).

Tab. 2 shows the whole picture of the AE database used in this paper. The sampling rates used here are all 500 KHz. Here, the sampling time of each piece of data lasts 102.294 ms. Under the condition that the rotation speed is 600 rad/s, the total point length of each AE discrete signal is 51147, and about 9.8 rotor cycles are continuously collected. Fig. 6 shows the image of these three signals in time domain.

Table 2: The quantity distribution of three AE signals at different rotational speeds (unit: Sample)

		The machine speed		
		600 rad/s	700 rad/s	800 rad/s
State of the rotor	Normal	118	121	123
	Cracking	494	728	270
	Rubbing	592	518	396

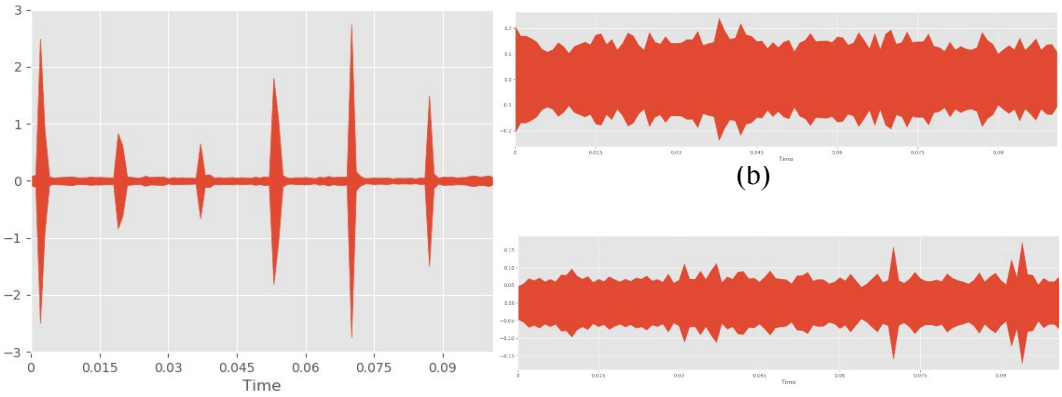


Figure 6: The samples of our database. (a) Cracking. (b) Rubbing. (c) Normal

3.2 Experiment setup

In this chapter, the AE signal with a rotation speed of 600 rad/s is used for the main experiment. The AE signal with the rotation speed of 700 rad/s and 800 rad/s is used as the reference experiment. The distribution of various types of data is shown in Tab. 2. The Hanning window is used to window framing the discrete AE signal. The choice of frame length depends mainly on the validity of the FFT point representation. Experiments were performed on the 256 points, 512 points, 1024 points and 2048 points FFTs respectively, and finally, a 512 points FFT was taken. This experiment uses the TensorFlow deep learning framework to build and train the network. And the ratio of the training set to the test set is 10:1.

The network uses the ReLu [Nair and Hinton (2010)] activation function, which represents a 'corrected linear unit' that can avoid gradient disappearance to some extent. The function expression of ReLu is:

$$ReLu(x) = max(0, x) \tag{1}$$

This experiment uses the Adam optimization algorithm for training. More details of the training are shown in Tab. 3.

Table 3: The details of the training

No.	Network	Input	Epoch	Batch Size	Loss
1	CNN	STFT	5	64	Categorical cross-entropy
2	CNN-2	MFCC	50	64	Categorical cross-entropy
3	LSTM	The output of CNN-2	10	64	-
4	Combine net	Concatenated result	10	64	Categorical cross-entropy

3.3 Result analysis

3.3.1 Recognition result

Fig. 7 shows the comparison before and after two streams plus the LSTM module. It can be seen from the figure that the accuracy rate of streams with MFCC as input is improved, so our final model chooses to add the LSTM module after this stream. Fig. 8 shows the confusion matrix proposed by the two single-stream convolutional neural networks for the classification of 600 rad/s rotating rubbing AE signals. It can be seen from the confusion matrix that CNN whose input is STFT spectrums, its accuracy is 94.12%, while the accuracy of the other single-stream CNN-LSTM is 88.03%. However, according to Fig. 9, we can see that the accuracy of our proposed TCNN can reach 100%, which is improved by 12% and 6% respectively compared with the two single-stream networks.

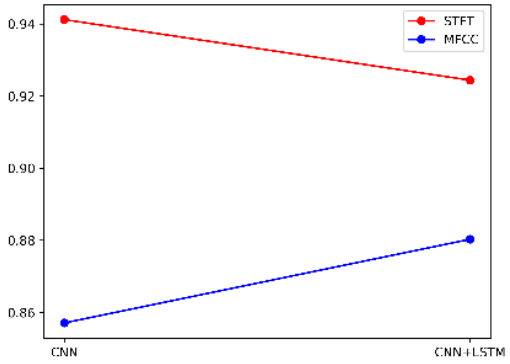


Figure 7: The comparison before and after two streams plus LSTM module

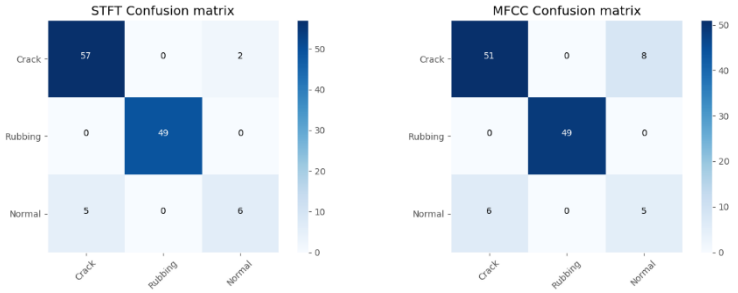


Figure 8: The confusion matrix of the single-stream CNN and CNN-LSTM

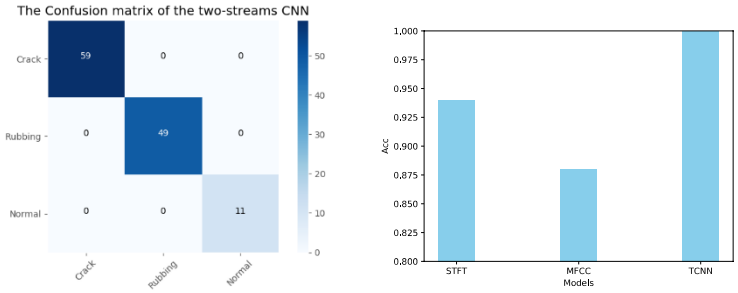


Figure 9: The confusion matrix of the TCNN and the accuracy of different networks

3.3.2 Model comparison experiment

In addition, in order to further explore the effectiveness of the proposed model, the conventional methods for the recognition of AE signals are compared. Tab. 4 shows the performance of different classifiers on the AE signal classification. The results are also shown in Fig. 10.

Table 4: The recognition rates of AE signals by different models at 600 rad/s (%)

Algorithm	**AE signal**			**Average**
	Normal	Cracking	Rubbing	
KNN	60.11	56.78	55.04	57.23
DNN	69.16	72.68	67.23	69.70
Sparse representation	59.97	66.35	60.04	62.18
SVM	69.98	76.49	70.54	72.40
CNN - Inception	78.29	84.53	81.24	81.44
Single-stream CNN	100.00	96.60	54.54	94.12
TCNN	100.00	100.00	100.00	100.00

It can be seen from the table that the recognition accuracy of the improved CNN is the highest for all of normal, cracking and rubbing AE signals, reaching 100% respectively, and the overall average recognition rate is 18% higher than SVM. The performance of SVM in the traditional AE signal identification method is the best, reaching 72.40%, which is 15.17% and 10.22% higher than KNN and sparse representation respectively. The classifier KNN has a poor recognition effect in the rubbing AE signal, and the recognition rate is only 57.23%. Although the recognition rate of sparse representation has increased to 62.18%, the classification effect is not satisfactory and the training time is longer.

Compared with DNN, which is also a neural network, since TCNN's inputs are STFT and MFCC, the more features of AE signals can be effectively captured, so the overall recognition effect is better, and average recognition rate has reached 100%. In addition, compared with the deep CNN, the shallow CNN proposed by us is more suitable for the training of small data sets.

3.4 Speed comparison experiment

Finally, this paper investigates the effect of machine speed on the recognition performance of AE signals by comparing the AE signals at different speeds.

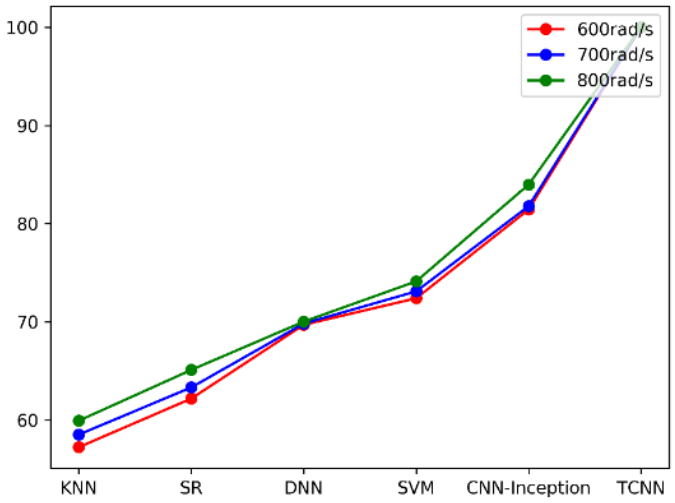


Figure 10: Recognition rates of different models at four speeds

The data shows that the improved CNN network proposed in this paper achieves the best recognition performance under different speed conditions. Compared with the traditional KNN, sparse representation and SVM, the effect of TCNN is obviously improved. The performance of each classifier is relatively stable on the AE signal at different speeds. The recognition rate of CNN under the three speed conditions is maintained 100%. The SVM has achieved the best recognition effect in the traditional classifier, and it has reached more than 70% under the three kinds of speed conditions. The sparse representation algorithm is inferior, and the effect of KNN is the worst, which recognition accuracy under the three speed conditions is maintained between 54.96% and 59.72%.

In addition, it can be found that as the AE signal speed increases, the recognition performance of each classifier is improved, indicating that the higher speed provides sufficient load for triggering the AE signal, which indicates that increasing the speed of the machine within a reasonable range facilitates the classification of the AE signal.

4 Conclusion

This paper proposes a TCNN and applies it to the detection of rotor rubbing AE faults. Firstly, the STFT and MFCC spectrum of the rubbing AE signals are constructed as network inputs. Secondly, in the network structure, the shallow CNN and CNN-LSTM structures are introduced to extract temporal and spatial features, which improve the generalization ability of the network while they don't require a lot of computing power. Finally, the performance of the TCNN network is verified by comparison experiments. The experimental results show that the TCNN network has a much higher recognition accuracy for AE signals than the traditional identification method, and good results can be obtained even in the case of small data set training.

Funding Statement: The author(s) received no specific funding for this study.

Acknowledgments: This research was funded by the National Natural Science Foundation of China [Nos. 51908285, 61673108 and 61571106], School-level Research Fund Project of Nanjing Institute of Technology [YKJ201975] and China Postdoctoral Science Foundation [2018M630559].

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

Reference

Baraldi, P.; Podofillini, L.; Mkrtchyan, L.; Zio, E.; Dang, V. N. (2015): Comparing the treatment of uncertainty in Bayesian networks and fuzzy expert systems used for a human reliability analysis application. *Reliability Engineering and System Safety*, vol. 138, pp. 176-193.

Deng, A.; Cao, H.; Tong, H.; Zhao, L.; Qin, K. et al. (2014): Recognition of acoustic emission signal based on the algorithms of TDNN and GMM. *Applied Mathematics and Information Sciences*, vol. 8, no. 2, pp. 907-907.

He, K.; Zhang, X.; Ren, S.; Sun, J. (2016): Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778.

Hochreiter, S.; Schmidhuber, J. (1997): Long short-term memory. *Neural Computation*, vol. 9, no. 8, pp. 1735-1780.

Khor, H. Q.; See, J.; Phan, R. C. W.; Lin, W. (2018): Enriched long-term recurrent convolutional network for facial micro-expression recognition. *13th IEEE International Conference on Automatic Face & Gesture Recognition*, pp. 667-674.

Krizhevsky, A.; Sutskever, I.; Hinton, G. E. (2012): Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, pp. 1097-1105.

Kundu, P.; Kishore, N. K.; Sinha, A. K. (2009): A non-iterative partial discharge source location method for transformers employing acoustic emission techniques. *Applied Acoustics*, vol. 70, no. 11-12, pp. 1378-1383.

LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. (1998): Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324.

Lei, Y.; Jia, F.; Lin, J.; Xing, S.; Ding, S. X. (2016): An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137-3147.

Li, X.; Li, J.; Qu, Y.; He, D. (2019): Gear pitting fault diagnosis using integrated CNN and GRU Network with both vibration and acoustic emission signals. *Applied Sciences*, vol. 9, no. 4, pp. 768-768.

Nair, V.; Hinton, G. E. (2010): Rectified linear units improve restricted Bboltzmann machines. *Proceedings of the 27th International Conference on Machine Learning*, pp. 807-814.

Prosvirin, A.; Kim, J.; Kim, J. M. (2017): Bearing fault diagnosis based on convolutional neural networks with kurtogram representation of acoustic emission signals. *Advances in Computer Science and Ubiquitous Computing*, pp. 21-26.

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S. et al. (2015): Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9.

Vapnik, V. (2013): The nature of statistical learning theory. Springer Science and Business Media.

Wang, D. (2016): K-nearest neighbors based methods for identification of different gear crack levels under different motor speeds and loads: revisited. *Mechanical Systems and Signal Processing*, vol. 70, pp. 201-208.

Zhou, L.; Wang, L.; Chen, Y.; Tang, Y. (2019): Binaural sound source localization based on convolutional neural network. *Computers, Materials & Continua*, vol. 60, no. 2, pp. 545-557.