Frequency Domain Feature Extraction and Long Short-Term Memory for Rolling Bearing Fault Diagnosis

Tianhao Wang
Department of Electronic and Electrical Engineering
Brunel University London
London, UK
1815203@brunel.ac.uk

Rui Qin
Department of Computing and Mathematics
Manchester Metropolitan University
Manchester, UK
r.qin@mmu.ac.uk

Hongying Meng
Department of Electronic and Electrical Engineering
Brunel University London
London, UK
hongying.meng@brunel.ac.uk

MaoZhen Li
Department of Electronic and Electrical Engineering
Brunel University London
London, UK
Maozhen.Li@brunel.ac.uk

Meng Cheng
Department of Electronic and Electrical Engineering
Brunel University London
London, UK
2104413@brunel.ac.uk

Yi Liu
Department of New Media
Beijing Institute of Graphic Communication
Beijing, China
yiliu61easy@163.com

***Abstract*—With the rapid development of the high-speed railways, the speed of trains is getting faster and faster, and the dynamic load between the wheels and rails of the vehicle increases accordingly. The rolling bearing is a key part of the high-speed train transmission system. The train is subjected to high-frequency vibration for a long time during operation, and the bearing is prone to fatigue damage, which affects the safe operation of the train. Nowadays, many methods have been applied in fault diagnosis like reinforcement learning, convolutional neural networks and autoencoders. One of the typical methods is the reinforcement neural architecture research method. It makes neural network design automatic and eliminates the bottleneck associated with choosing network architectural parameters. However, this method focuses on the time domain signal, and a time domain signal cannot capture the particular properties of a frequency domain signal. In order to solve these problems, we propose a new method containing two Steps: Use FFT to convert the time domain signal to the frequency domain and use Bi-LSTM neural network model to recognize different faults. For each fault, the time series signal has some correlation with some specific frequencies. The frequency domain is more intuitive than the time domain and describes different states of faulty types. For recognition, LSTM is better at classifying sequence data than other methods, and Bi-LSTM can predict the sequence from both directions, achieving higher accuracy. Experiments on public data sets demonstrate the efficiency of the proposed method.**

***Index Terms*—Bi-LSTM, FFT, Bearing fault**

I. Introduction

A few essential components ensure that the train operates safely and efficiently, and the development of high-speed rail has considerably improved traffic conditions in many countries. The transmission system is one of them and is crucial to the efficient running of high-speed trains. In the majority of high-speed train transmission systems, the traction motor serves as the power source and is a critical component. Therefore, it is important to recognise traction motor breakdowns. The most frequent type of motor failure is motor bearing failure [1]. For instance, according to the research, 40% of motor failures are caused by motor bearing failures. Fatigue spalling, which includes partial damage or fall-off on the inner ring, outer ring, rolling element, and other bearing surfaces, is one of the most common defects in motor bearings. Fatigue stress from alternating loads is the main factor in bearing fatigue spalling [2]. When a bearing experiences a fatigue spalling failure, a particular frequency of shock pulse will manifest. As a result, rolling bearing maintenance is quite expensive and very important for every country. For instance, the US spends hundreds of billions of dollars every year on maintaining machinery and routinely replacing vital components [3].

However, if the crucial components of the equipment are not updated in a timely manner, catastrophic tragedies could happen. For instance, on June 3, 1998, a high-speed train in Germany's elastic wheel burst due to prolonged use, resulting in 101 fatalities and 194 serious injuries [4]. 72 people lost their lives and 416 were hurt when 9 carriages of the NO.T195 train from Beijing to Qingdao derailed and crashed with the

NO.5034 train from Yantai to Xuzhou on April 28, 2008, in China. [5].

Consequently, rolling bearing detection and recognition for health monitoring has emerged as one of the key research fields in order to lower the cost of rolling bearing maintenance and maintain the safety of operation for high-speed trains. In recent years, some laboratories used accelerometers located at the driving end of the motor housing to adopt a significant amount of vibration data. The sampling frequency is usually 12000 samples per second, 25600 samples per second and 48000 samples per second. These vibration data lengths are generally more than 120000, and many subcategories of vibration signals exist.

Hence, signal processing and status recognition have taken centre stage in rolling bearings research. For signal processing, Huang et al. [6] proposed a method of adaptively decomposing non-stationary signals into a series of zero-mean intrinsic modal functions (IMF), which was called empirical mode decomposition (EMD). Saidi et al. [7] used EMD to dissect the non-stationary signal into several IMFs according to the local characteristic time scale of the signal. Ali et al. [8] used the Intrinsic Mode Function (IMF) energy bribe generated by empirical mode decomposition to describe seven different bearing states. Dybała and Zimroz [9] proposed an early damage detection method for rolling bearings based on EMD. Wang et al. [10] proposed a new non-negative EMD manifold (NEM) bearing failure feature extraction method. Popular learning has been a more popular dimensionality reduction method in recent years. It has been used in a wide variety of fields of fault diagnosis. Among them, Arena et al. [11] proposed Laplacian Eigenmaps (LE), He and Partha [12] proposed Locality Preservation Projection (LPP), Roweis and Saul [13] proposed Locally Linear (LLE), Zhang et al. [14] proposed linear local tangent space alignment (LLTSA). He [15] used LLE to extract the popular features of wavelet packet energy and effectively distinguished bearing and gear failures with different failure degrees. Li and Zhang [16] used the supervised LLE algorithm to map the features from the high-dimensional space to the embedding space and performed bearing fault classification in the embedding space.

More importantly, status recognition achieves much success as well. For instance, Shao et al. [17] proposed a Deep wavelet auto-encoder (DWAE) with an extreme learning machine (ELM). They used the wavelet function to design a wavelet autoencoder, to get data features and improve the ability to study unsupervised features. ELM is a classifier. The result is 95.2%. Shao et al. [18] proposed ensemble deep auto-encoders (EDAEs). Use the Unsupervised feature learning from the raw vibration data and design a strategy to ensure accuracy and stability. The result is 97.18%. Tao et al. [19] proposed deep belief networks (DBN). DBN can reduce energy loss between the output and input vibration signals. The result is 96.67%. Gan and Wang [20] proposed hierarchical diagnosis network (HDN) can achieve 99.03%. Zhuang and Qin [21] proposed a multi-scale deep CNN (MS-DCNN) model that can reach 99.27%. Guo et al. [22] constructed a hierarchical adaptive deep convolutional neural network (ADCNN), the accuracy is 97.7%.

In these recognition methods, the ability to reinforce learning is the most similar to manually detection. Among them, Wang et al. [23] proposed a reinforcement neural architecture search method to achieve success. The article suggested and validated the neural network architecture automatic search method. The framework of the article includes two parts: the controller model and the child model.

The controller model has 2 Nascell layers, and the output of this model are convolutional kernel size and kernel number and a pooling kernel size of each layer. They formed the CNN [24]. The child models are CNNs. The model consists of an input layer. The two groups of the same convolutional layer, the pooling layer, take turns to each other. The complete connection layer.

However, the time domain is the main emphasis of these methods. The time domain analysis is unable to observe the frequency-dependent signal properties for the vibration signal. The frequency domain analysis is more succinct than the time domain. Following the signal in the frequency domain provides a deeper and more practical analysis of the issue.

This paper proposed feature extraction and recognition for rolling bearing fault diagnosis based on frequency domain and Long Short-Term Memory (LSTM) to overcome the shortcomings mentioned before. In this method, uses Fast Fourier Transform (FFT) to alter the bearing's time-domain signal before it is transmitted to the network. We only need to fine-tune the maxepochs and hidden units in the process.

II. Methodology

A. Overview of system

The structure of the proposed method is shown in Fig. 1, the vibration signal is collected by accelerometers which are located drive end of the motor housing, and the signal is measured at 1750 RPM in each working state. Then the vibration signal is divided into several overlapping samples, and each sample is window processed and transformed with FFT. And then input these data into networks, the neural network train these data, gets outputs and calculates the accuracy. Adjust the number of the hidden units and MaxEpochs until the result reaches the best.

The layers in the neural network system include the input layer, Bi-LSTM layer, fully connection layer, softmax Layer and classification Layer.

B. Frequency domain analysis

The existence of FFT makes Discrete Fourier Transform play a central role in algorithms in digital signal processing. The calculation formula of Discrete Fourier Transform is [25]:

$$X(k) = \sum_{n=0}^{N-1} x(n)W^{nk},(0\leq k \leq N-1)$$

$$X(k) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)W^{-nk},(0\leq k \leq N-1)$$

Where x stands a limited long sequence, X stands data after Discrete Fourier transformation, N is sampled n points in a sinusoidal cycle, and $W = e^{\frac{-j2p}{N}}$ is the Fourier factor. For a

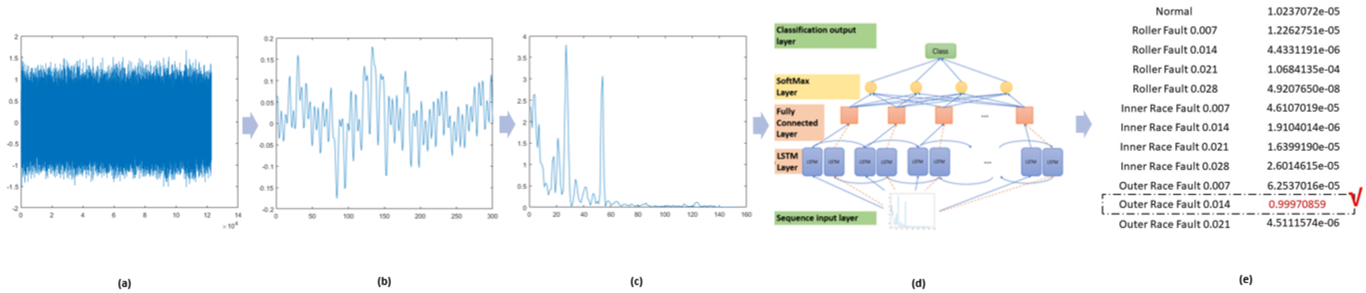

Fig. 1: Overview of the system, (a) is raw data example from CWRU dataset [29], (b) is egmented data with 300 points, (c) is segmented data using FFT, (d) is training data sent to the network and (e) is score of 12 classes.

discrete signal x, FFT will transform the signal in frequency domain. Fig. 2 shows that the signal changes before and after FFT.

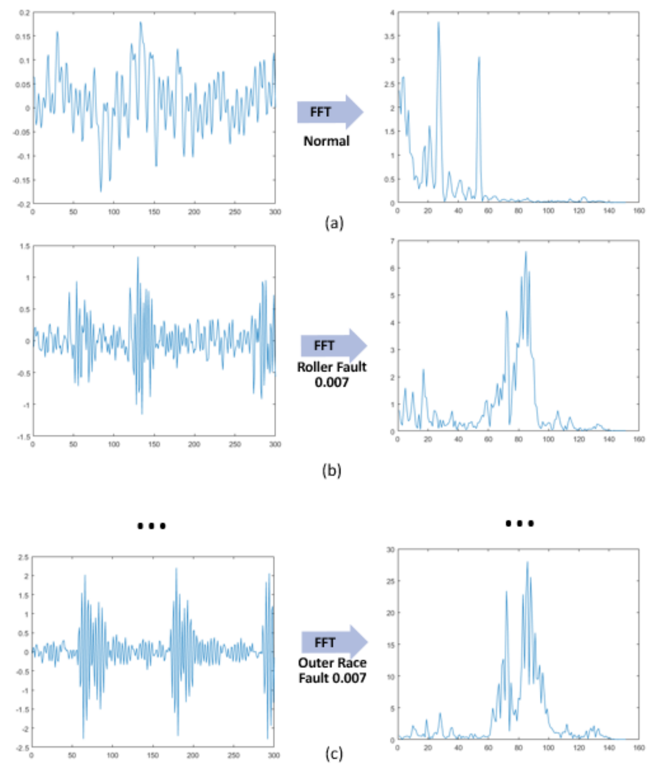

Fig. 2: Signal transformed by FFT. Left part is raw data, each of them have 300 points, right part is data with FFT, each of them have 150 points. And (a) is an example of Normal signal, (b) is an example of Roller Fault size 0.007 inches signal, (c) is an example of Outer Fault size 0.007 inches signal.

C. Bi-LSTM

Long Short Term Memory (LSTM) is a special Recurrent Neural Network (RNN) that can learn long-term dependencies [26]. Vanishing and exploding gradient problems are hard to avoid in traditional RNNs. LSTM learned the long-term dependence on the network with passed these problems. The hidden layer of traditional RNN is usually a $tanh$ function or ReLU. A typical LSTM unit will conclude 3 $sigmoid$ layer and 1 $tanh$ layer.

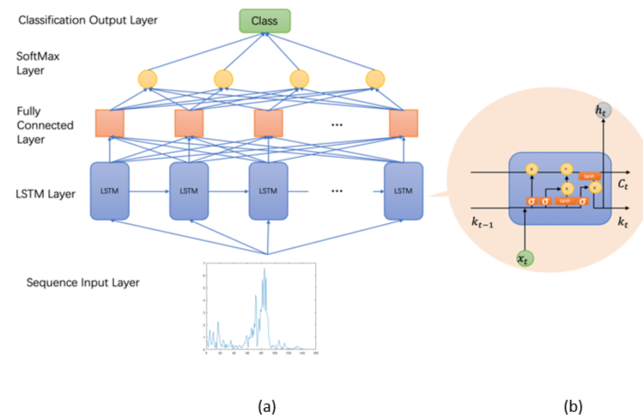

Fig. 3: (a) is notwork structure of LSTM, (b) is LSTM unit. The repetitive module in LSTM has four interaction layers, three sigmoid and one $tanh$, and they interact in a unique way.

LSTM consists of three gate variables: Input gate, Forgetting Gate and Output gate.

A cell state C is applied in LSTM with only a few linear operating on it, which could retain information easily. The first gate in LSTM is forget gate, which decides what information should be discarded. x_t will be send to a sigmoid function with h_{t-1} and get a value between 0 and 1 which multiplied with the cell state C_{t-1}. The output of the sigmoid function will decide how much information remains. Part of the information in the last layer $t-1$ has been forgotten in the cell state C_{t-1}, and the new information in the current layer will be added by a $tanh$ function and a sigmoid function. This sigmoid function is called input gate and the output of it will multiply by a $tanh$ function. When the value of it is 0, the cell state doesn't need to update.

Then the last cell state C_{t-1} multiply with forget gate f_t to discard part of information and update the information from $i_t \times C_t$. The output gate concludes the information in updated cell state C_t and the output after a $tanh$ function and a sigmoid function. A brief figure of LSTM is shown in Fig. 3.

TABLE I: Description of 12 states, one normal state and eleven failure states

Data no.	Fault type	Fault size/inches	Motor speed (r/min)	Sampling frequency (kHz)	Size of training / testing samples
1	Normal	0	1750	12	350/50
2	Roller Fault	0.007	1750	12	350/50
3	Roller Fault	0.014	1750	12	350/50
4	Roller Fault	0.021	1750	12	350/50
5	Roller Fault	0.028	1750	12	350/50
6	Inner Race Fault	0.007	1750	12	350/50
7	Inner Race Fault	0.014	1750	12	350/50
8	Inner Race Fault	0.021	1750	12	350/50
9	Inner Race Fault	0.028	1750	12	350/50
10	Outer Race Fault-Center@6:00	0.007	1750	12	350/50
11	Outer Race Fault-Center@6:00	0.014	1750	12	350/50
12	Outer Race Fault-Center@6:00	0.021	1750	12	350/50

LSTM has the advantages of long-term trajectory memory and short memory unification, simulation of selective brain forgetting, and more accurate trajectory modelling. Therefore, the multi-layer structure can be mixed to solve the efficiency and stability problems of massive data training.

Bi-LSTM is an RNN with LSTM unit and will predict the sequence from both directions [27]. It could perform better in a sequence without directionality [28]. Actually the sequences are sent to two LSTM unit indenpently with different directions. The structure of Bi-LSTM is shown in (a) of Fig. 4. In this paper, Bi-LSTM model was selected instead of LSTM model because the input data is the FFT of the vibration signals. In frequency domain, there is no strong dependent relationship between current and previous components. Bi-LSTM can model both relationship of the one from low frequency to high frequency as well as the one from high frequency to low frequency.

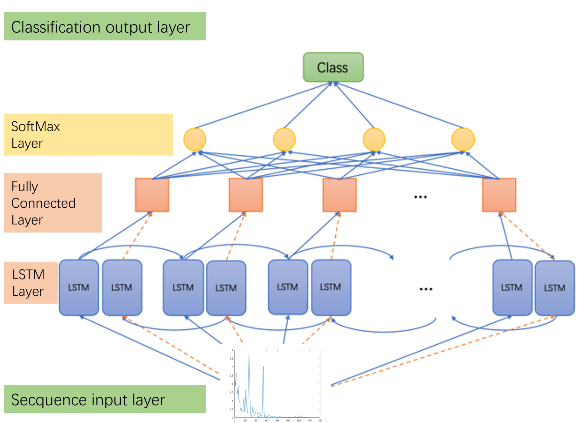

Fig. 4: Bi-LSTM networks. The LSTM framework is used to merge the input sequence's front and backward directions. The two LSTM layers' vectors can be added to, averaged out, or connected.

III. Experiment

A. Dataset

The data are collected from the Electrical Engineering Laboratory at Case Western Reserve University [29]. The data was collected at 12,000 samples/second and has four different fault inches: 0.007, 0.014, 0.021 and 0.028 inches.

B. Parameters

The parameter information is shown in Table 1. Twelve data in different states include one normal state and eleven faults states. Each of these has 400 samples. Among them, the first 350 samples are training samples, and the left 50 samples are test samples, a totally of 4800 samples, and the length of each sample is 300. The last vision of network Parameters in this paper is in the following Table II.

TABLE II: Parameters of network

Name of Parameters	Size of Parameters
Mini Batch-Size	**20**
Input Size	**1**
Number Hidden Units	**100**
Number Classes	**12**
Max Epochs	**100**

C. Experimental results

In this experiment, the experimental setting follows the same one in paper [23] for a fair comparison. In this setting, the subset data related to 12 faulty types are used, as shown in Table 1. In [23], the result of the recognition accuracy is 98.47%, and the standard deviation obtained by reinforcement neural architecture after 10 experiments are 0.61.

The average classification accuracy rate of proposed method on raw data can reach 94.88%, and the accuracy standard deviation after 10 experiments is 1.90 as shown in Table 3. This paper's average classification accuracy rate for data with FFT can reach 99.70±0.23%, improved by 4.82%. Compared with neural reinforcement architecture, enhanced by 1.23%, the classification accuracy of test samples is summarized in Table 3, which is better than reinforcement neural architecture. Therefore, it can be considered that FFT data in Bi-LSTM has better performance.

TABLE III: Results of comparison paper, Bi-LSTM for raw data and FFT data

Method	Average accuracy (%) ± standard deviation (%)
Reinforcement neural architecture	98.47 ± 0.61 [23]
Raw data/Bi-LSTM	94.88±1.90
FFT/Bi-LSTM	**99.70±0.23**

And the 10 times of results show in table 4, the best result for raw data is 97.83% and the worst is 92%. the best result for data with FFT approached 100% and the worst is 99.33%, which improved 2.17% and 7.33% respectively.

TABLE IV: Test results of Raw data and Data with FFT

NO. Times	Raw data/Bi-LSTM %	FFT/Bi-LSTM %
1	97.83	100
2	96.67	100
3	96	99.83
4	96.5	99.83
5	95.67	99.83
6	94	99.67
7	93.67	99.5
8	93.67	99.5
9	92.83	99.5
10	92	99.33

It is significantly better than the comparison network structures, it can be seen that the proposed method has better performance. the diagnostic results are summarized in Fig. 5.

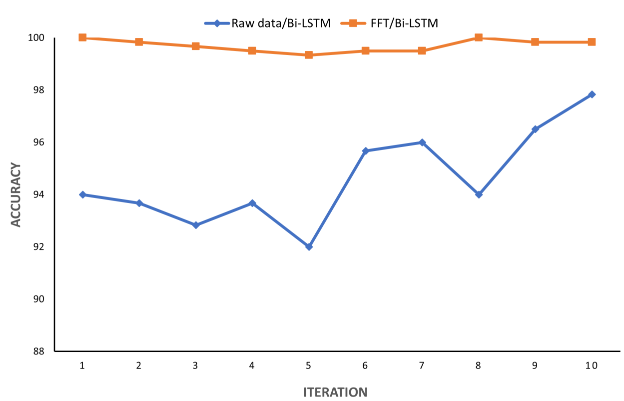

Fig. 5: Accuracy of the 10 times.The horizontal axis is the number of iterations and Vertical axis is accuracy. The blue folding line is the result of the raw signal combined with the Bi-LSTM network, the orange folding line is the result of the signal after FFT combined with the BI-LSTM network.

Conclusion

Even if the data from the time sequence is transferred to the frequency domain after being processed by FFT, it remains a continuous signal. It is equally applicable to engineering practice and produces positive outcomes. In the paper [30], [31], [32], combination method of FFT and LSTM has achieved good results.

As a result of comparisons between paper [23], the experimental procedures are completely consistent. Paper [23] only uses training and testing sets, and there is no verification set. Therefore, there is no verification set in the experiments described in this paper.

Finally, deep learning methods, in particular, have become increasingly popular in recent years. As a result, deep learning-based automated recognition is now utilised in numerous new sectors. The goal of increasingly more automatic recognition systems is to simplify our lives. In this paper, the signal is transformed from the time domain to the frequency domain by fully using of the FFT. The best result, when combined with BI-LSTM, is 100%, while the mean value over ten times is 99.70±0.23 %. Future studies on the existing research issue can include a variety of methods, including EMD. Practical engineering challenges will be solved using some methodologies that will be enhanced and employed in this research.

Acknowledgment

The authors would like to thank the Natural Science Foundation of China (NSFC) on the International Collaboration and Exchange Project under Grant 6196020601 for the support.

References

[1] S. Nandi, H. A. Toliyat and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review", IEEE transactions on energy conversion, 2005, 20(4): 719-729.

[2] L. Rosado, N. H. Forster, K. L. Thompson and J. W. Cooke, et al. "Rolling contact fatigue life and spall propagation of AISI M50, M50NiL, and AISI 52100, Part I: experimental results", Tribology Transactions, 2009, 53(1): 29-41.

[3] C. Wu, J. Liu, F. Peng and R. Li. "Gearbox fault diagnosis using adaptive zero phase time-varying filter based on multi-scale chirplet sparse signal decomposition", Chinese Journal of Mechanical Engineering, 2013, 26(4): 831-838.

[4] X. Li, S. Liu, S. Cheng, J. Lin, R. Liu, L. Wang and Z. Zhou, "Research on Ultrasonic Quantitative Evaluation Technology of Complex Defects Based on Neural Network", Journal of Physics: Conference Series. IOP Publishing, 2022, 2196(1): 012022.

[5] Y. Cao, Y. An, S. S, G. X and Y. S, "A statistical study of railway safety in China and Japan 1990–2020." Accident Analysis Prevention 175 (2022): 106764.

[6] N. E Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, Q. Zheng, NC. Yen, C. C. Tung and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis", Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, vol. 454, no. 1971, pp. 903-995, 1998. Available: 10.1098/rspa.1998.0193.

[7] L. Saidi, J. Ali and F. Fnaiech, "Bi-spectrum based-EMD applied to the non-stationary vibration signals for bearing faults diagnosis", ISA Transactions, vol. 53, no. 5, pp. 1650-1660, 2014. Available: 10.1016/j.isatra.2014.06.002.

[8] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals", Applied Acoustics, vol. 89, pp. 16-27, 2015. Available: 10.1016/j.apacoust.2014.08.016.

[9] J. Dybała and R. Zimroz, "Rolling bearing diagnosing method based on Empirical Mode Decomposition of machine vibration signal", Applied Acoustics, vol. 77, pp. 195-203, 2014. Available: 10.1016/j.apacoust.2013.09.001.

[10] C. Wang, M. Gan and C. Zhu, "Non-negative EMD manifold for feature extraction in machinery fault diagnosis", Measurement, vol. 70, pp. 188-202, 2015. Available: 10.1016/j.measurement.2015.04.006.

[11] P. Arena, L. Patanè and A. Spinosa, "Data-based analysis of Laplacian Eigenmaps for manifold reduction in supervised Liquid State classifiers", Information Sciences, vol. 478, pp. 28-39, 2019. Available: 10.1016/j.ins.2018.11.017.

[12] X. He and N. Partha, "Locality preserving projections." Advances in neural information processing systems 16 (2003).

[13] S. Roweis and L. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding", Science, vol. 290, no. 5500, pp. 2323-2326, 2000. Available: 10.1126/science.290.5500.2323.

[14] T. Zhang, J. Yang, D Zhao and X. Ge. "Linear local tangent space alignment and application to face recognition." Neurocomputing 70.7-9 (2007): 1547-1553.

[15] Q. He, “Vibration signal classification by wavelet packet energy flow manifold learning”, Journal of Sound and Vibration, vol. 332, no. 7, pp. 1881-1894, 2013. Available: 10.1016/j.jsv.2012.11.006.

[16] B. Li and Y. Zhang, “Supervised locally linear embedding projection (SLLEP) for machinery fault diagnosis”, Mechanical Systems and Signal Processing, vol. 25, no. 8, pp. 3125-3134, 2011. Available: 10.1016/j.ymssp.2011.05.001.

[17] H. Shao, H. Jiang, X. Li and S. Wu, “Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine“, Knowledge-Based Systems, 2018, 140: 1-14.

[18] H. Shao, H. Jiang, Y. Lin and X. Li, “A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders“, Mechanical Systems and Signal Processing, 2018, 102: 278-297.

[19] J. Tao, Y. Liu, D. Yang, F. Tang and C. Liu, “Fault diagnosis of rolling bearing using deep belief networks“, 2015 International Symposium on Material, Energy and Environment Engineering. Atlantis Press, 2015: 566-569.

[20] M. Gan and C. Wang, “Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings“, Mechanical Systems and Signal Processing, 2016, 72: 92-104.

[21] Z. Zhuang and W. Qin, “Intelligent fault diagnosis of rolling bearing using one-dimensional multi-scale deep convolutional neural network based health state classification“, 2018 IEEE 15th International conference on networking, sensing and control (ICNSC). IEEE, 2018: 1-6.

[22] X. Guo, L. Chen, C. Shen, “Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis”, Measurement, 2016, 93: 490-502.

[23] R. Wang, H. Jiang, X. Li and S. Liu, “A reinforcement neural architecture search method for rolling bearing fault diagnosis”, Measurement, vol. 154, p. 107417, 2020. Available: 10.1016/j.measurement.2019.107417.

[24] B. Zoph and Q.V. Le, “Neural architecture search with reinforcement learning”, arXiv preprint arXiv:1611.01578, 2016.

[25] P. Heckbert, “Fourier transforms and the fast Fourier transform (FFT) algorithm”, Computer Graphics, 1995, 2: 15-463.

[26] J. Schmidhuber, and S. Hochreiter. “Long short-term memory”, Neural Comput 9.8 (1997): 1735-1780.

[27] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks”, in IEEE Transactions on Signal Processing, vol. 45, no. 11, pp. 2673-2681, Nov. 1997, doi: 10.1109/78.650093

[28] A. Graves, S. Fernández and J. Schmidhuber, “Bidirectional LSTM networks for improved phoneme classification and recognition”, International conference on artificial neural networks. Springer, Berlin, Heidelberg, 2005: 799-804.

[29] CWRU, Case Western Reserve University, “Bearing data”, [Online]. Available:https://engineering.case.edu/bearingdatacenter/download-data-file

[30] J. Wan, L. Shih, Y. Hong and H. Che, ”Evaluation of deep learning neural networks for surface roughness prediction using vibration signal analysis.” Applied Sciences 9.7 (2019): 1462.

[31] M. Hachemi, A. Ghomari, Y. Hadjadj-Aoul and G. Rubino, ”Mobile traffic forecasting using a combined FFT/LSTM strategy in SDN networks.” 2021 IEEE 22nd International Conference on High Performance Switching and Routing (HPSR). IEEE, 2021.

[32] H. Park, E. Kwon, S. Byon, W. Shin, E. Jung and Y. Lee, ”Punch Analysis with FFT and LSTM of Accelerometer and Gyroscope Data.” 2020 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, 2020.