Bearing Fault Diagnosis of WDCNN-LSTM in Siamese Network

Daehwan Lee, Jongpil Jeong
dept. Smart Factory Convergence
Sungkyunkwan University
Seoul, Republic of Korea
dhlee0915@g.skku.edu
jpjeong@skku.edu

Chaegyu Lee
dept. Smart Factory Convergence
Sungkyunkwan University
Seoul, Republic of Korea
leechgyu@skku.edu

Hakjun Moon
dept. Computer Science and Engineering
Sungkyunkwan University
Seoul, Republic of Korea
gloriel621@g.skku.edu

Jaeuk Lee and Dongyoung Lee
dept. Mechanical Engineering
Sungkyunkwan University
Seoul, Republic of Korea
hg9430@g.skku.edu
ehddud6332@g.skku.edu

***Abstract*—In this paper, a Siamese network-based WDCNN + LSTM model was used to diagnose bearing faults using a few shot learning algorithm. Recently, deep learning-based fault diagnosis methods have achieved good results in equipment fault diagnosis. However, there are still limitations in the existing research. The biggest problem is that a large number of training samples are required to train a deep learning model. However, manufacturing sites are complex, and it is not easy to intentionally create equipment defects. Furthermore, it is impossible to obtain enough training samples for all failure types under all working conditions. Therefore, in this study, we propose a few-shot learning algorithm that can effectively learn with limited data. A Few shot learning algorithm and Siamese network based WDCNN + LSTM model bearing fault diagnosis, which can effectively learn with limited data, is proposed in this study.**

***Index Terms*—Few Shot Learning, Siamese Network, Fault Diagnosis, WDCNN, LSTM**

I. Introduction

Over time, industries have become more sophisticated and their products more complex. As industries and products become more complex, so do facilities, which consist of many devices with many different parts. In complex facilities, there is a high probability of equipment failure due to inter-component factors, such as defective parts, or environmental factors, such as rapid climate change, as products are produced on the manufacturing line on site. As the most critical component of rotating mechanical equipment, rolling bearings are the number one cause of rotating equipment failure and have a major impact on the entire facility and manufacturing line. Bearing defects gradually occur due to overloading, impact loading, heat generation due to creep, and the use of unsuitable lubricants. In addition, the types of bearing defects include flaking, peeling, scoring, smearing, fracture, crack, etc [1], [2], [3].

Jongpil Jeong, Chaegyu Lee

Previous studies include SVM(Support Vector Machine) [4], Bayesian classification [5], etc. Since then, in recent years, deep learning models have been increasingly used for bearing fault diagnosis due to their powerful data processing and feature learning capabilities. CNN(Convolution Neural Network) [6], RNN(Recurrent Neural Network) [7], GAN(Generative Adversarial Networks) [8], auto encoder, etc. have been studied. In the above studies, many data-driven and deep learning-based techniques have been applied to improve accuracy and reliability, but most deep learning models require a large number of data samples to learn all failure type classifications. However, in complex, fast-paced manufacturing sites, the data for all failure types is limited. Therefore, a few shot learning algorithms that can effectively proceed with less data is needed.

In this paper, we propose a Siamese network WDCNN(Wide First-layer Kernels Deep CNN)+LSTM model fault diagnosis method based on few shot learning. Along with the proposed few-shot learning algorithm, we develop a model that improves the existing WDCNN by adding an LSTM to the WDCNN backbone. It is hypothesized that the addition of an LSTM will improve the model's ability to capture the temporal dependencies present in the vibration signal, thereby improving its ability to accurately classify various fault conditions. The proposed approach is evaluated on the CWRU dataset and its performance is compared with SVM, Wide first-layer kernels Deep CNN (WDCNN), Five-shot (WDCNN), and Five-shot (New model) in terms of bearing fault diagnosis accuracy with different number of samples in each case.

The contributions of this paper are as follows. We developed a model that improves the existing WDCNN by adding LSTM to the WDCNN backbone along with a proposed few-shot learning algorithm. We demonstrate that the model developed in this paper outperforms the fault diagnosis performance of the existing model with several samples.

This paper is organized as follows. Section 2 describes Few shot learning, Siamese network, and LSTM. Section 3 describes the main idea of this paper, Few-shot learning based fault diagnosis. Section 4 describes the experimental procedure, dataset construction, and experimental results. Finally, Section 5 describes the conclusion and future work.

II. Related Work

A. Few Shot Learning

Few-shot learning was first addressed in the 1980s [9]. Deep learning has been very successful in many fields, but the performance of the model is hindered when the data set is small. To solve this problem, few shot learning was proposed. Few-shot learning can alleviate the burden of collecting large-scale supervised data. It uses a new machine learning paradigm called few shot learning to learn with less data.

Recently, Few-shot learning has made great progress in solving the data shortage problem. Few-shot learning is different from a typical CNN model. General deep learning is divided into training set and test set, but few shot learning is divided into training set, support set, query set, and performs N-way K -shot classification. N is the number of classes, K is the number of support data for each class, and N is inversely proportional and K is directly proportional, so the larger the N, the harder the problem to solve, and the larger the K, the easier the problem to solve [10], [11], [12].

B. Siamese Network

Siamese networks were first introduced by Bromley and Lekun in the early 1990s to solve signature verification as an image matching problem [9]. Fig. 1 shows the siamese network. A Siamese network is a type of neural network architecture used to learn the similarity or dissimilarity between two inputs and was first introduced by Bromley and Lekun in the early 1990s to solve signature verification as an image matching problem. The network consists of two identical subnetworks that share the same weights and architecture, hence the name "Siamese". The inputs are passed through the subnetworks and a similarity score is calculated by comparing the output representations of the two inputs [13], [14].

Subnetworks can be any type of neural network architecture, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), or a combination of several types of networks. Subnetworks are trained to learn representations of inputs so that similar inputs have similar representations and different inputs have different representations.

Siamese networks are commonly used for tasks such as image or text similarity, one-shot learning, and few-shot learning. They are also used in conjunction with triple loss functions, a method of training a network to produce similar outputs (e.g., 1 or positive pair) for similar inputs and different outputs (e.g., 0 or negative pair) for different inputs.

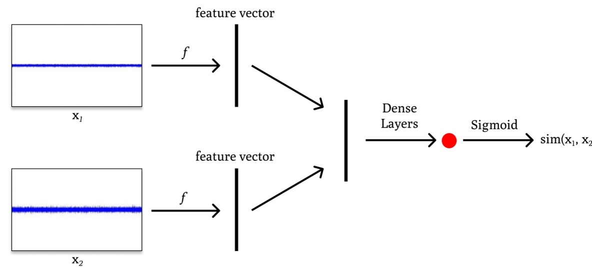

Fig. 1. Siamese Network Structure

C. LSTM

LSTM stands for Long Short-Term Memory and is a type of Recurrent Neural Network (RNN), a type of deep learning proposed by Hochreiter et al. LSTMs are models that can remember long-term information, solving one of the limitations of RNNs, the gradient blowup problem. It is widely used in fields such as natural language processing and speech recognition [15]. Fig. 2 shows the LSTM structure. An LSTM consists of four gates and one memory cell. It consists of four gates, Forget Gate, Input Gate, Output Gate, and Cell State, and an update memory cell that creates a new memory cell by updating the information in the previous memory cell. The roles of each gate are as follows.

1. Forget Gate: Determines which of the information in the previous memory cell is discarded.

2. Input Gate: Adds or modifies information about the current input value.

3. Output Gate: Determines which information to use to generate the final output value.

4. Cell State: The memory cell of the LSTM, which remembers and transmits information in the long term.

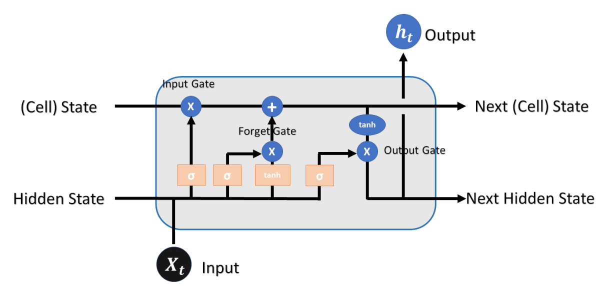

Fig. 2. LSTM Structure.

III. WDCNN-LSTM Based Bearing Fault Diagnosis

It is a Siamese network few-shot learning classification method based on our proposed WDCNN+LSTM model. Fig. 3 shows the system structure.

It consists of data preparation step (first), training and testing process with few shot learning (second), and finally Siamese network structure based on WDCNN+LSTM model (third). To verify the model performance, 12k drive end

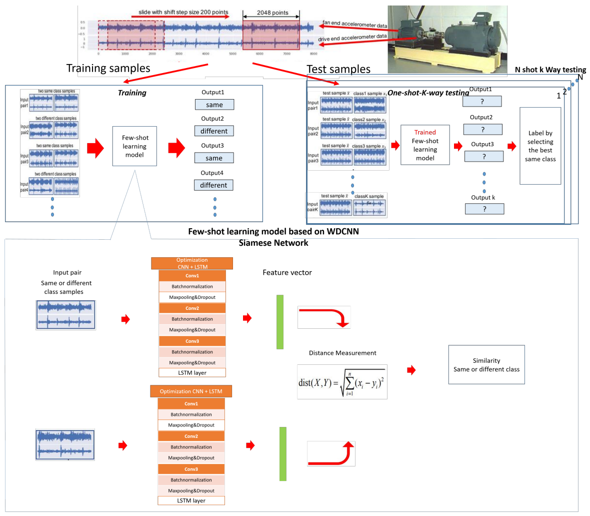

Fig. 3. System Structure.

bearing failure data from the bearing data set of Case Western Reserve University (CWRU) is selected as the experimental data.

The first step is data preparation. In the experiment, each sample is extracted from two vibration signals (fan end, drive end). Half of the vibration signals are used to generate training samples and the other half are used to generate test samples. The training samples were generated with a window size of 2048 points and 80 shift steps. The test samples are also generated with the same window size and non-overlapping.

The second is the training and testing phase. During training, the model is trained on a set of sample pairs of the same or different categories. The inputs are pairs of samples with the same or different classes. The WDCNN+LSTM model takes as input the two vibration signals prepared above. Each model outputs two feature vectors extracted from the two input vibration signals. After the output, find the difference (distance) between the two feature vectors. Then use a dense layer to find the difference between the vectors. Apply a sigmoidal activation function to get a number between 0 and 1. Measure the similarity between the two, outputting a value close to 1 if the two vibration signals are in the same class of normal or faulty type, and close to 0 otherwise, and measure the difference between the target value and the predicted scalar using the loss function.

The test is performed using multiple one-shot k-way tests. In an N-shot K-way test, the model is given a support set of K different classes, each with N samples. Determine which support set class the test sample belongs to. In this paper, we used 5 shots, so each time a support set is randomly selected from the training data, we repeat the one-shot K-way test 5 times. After 5 attempts, 5 probability factors (P1, P2, P3, P4, P5) are calculated and then their sum is calculated to get the largest value.

The third is the structure of WDCNN+LSTM model in the Siamese network. Table I shows the WDCNN+LSTM structure. This model consists of three convolution layers and an LSTM layer. We also added Batchnomalization, maxpooling, and dropout between convolutional layers.

TABLE I
WDCNN+LSTM STRUCTURE

no	Layer type
1	Conv 1 Filters 16 Kernel Size 64 Strides 16
2	Batchnormalization
3	MaxPooling(pool size = 2, strides = 2)
4	Dropout
5	Conv 2 Filters 32 Kernel Size 3 Strides 1
6	Batchnormalization
7	MaxPooling(pool size = 2, strides = 2)
8	Dropuout
9	Conv 3 Filters 64 Kernel Size 3 Strides 1
10	Batchnormalization
11	MaxPooling(pool size = 2, strides = 2)
12	Dropuout
13	LSTM

IV. EXPERIMENT AND RESULTS

A. Experiment Environments

Table II shows the experimental environment. The hardware used in this study consisted of an Intel Core i5- 13600KF processor and Geforce RTX 4080. The software uses Window 10, Tensorflow 2.10 and Python 3.9.

TABLE II
SYSTEM SPECIFICATION

Hardware Environment	Software Environment
CPU: Intel Core i5-13600KF CPU@ 3.50GHZ	window10
GPU : NVIDIA Geforce RTX 4080	Python 3.9, Tensorflow 2.10

The Case Western Reserve University (CWRU) bearing dataset was used to validate the performance of the model in this paper. The CWRU dataset consists of three types of data: healthy and defective. The defect data types are inner race, outer race, and ball. The defect sizes for each type are 0.007, 0.014, and 0.021 inches. The data was collected from the fan end and drive end, measured at 12k (12000 vibrations per second) and 48k (48000 vibrations per second), respectively. For each fault size, we configured 0 to 3 horsepower, and the outer race fault measured faults at the 3, 6, and 12 o'clock positions.

Table III shows the composition of the data. Label from 1 to 10 for normal and for the size of each fault type. dataset A, B, and C have training data and test data corresponding to angular loads 1, 2, and 3. dataset D has training data and test data. dataset d is the combined dataset of datasets A, B, and C. There are 1980 training and 75 test data for each type of defect, corresponding to load 1, 2, and 3.

Fig. 4 shows the bearing simulator of CWRU. CWRU The CWRU simulator is composed of the dynamometer, Electric motor, Drive end bearing, Fan end bearing and Torque transducer and encoder.

TABLE III
DESCRIPTION OF ROLLING BEARING DATASETS

Fault Location		None	Ball			Inner Race			Outer Race			Load
Fault Diameter(inch)		0	0.007	0.014	0.021	0.007	0.014	0.021	0.007	0.014	0.021	
Fault Label		1	2	3	4	5	6	7	8	9	10	
Dataset A	Train	660	660	660	660	660	660	660	660	660	660	1
	Test	25	25	25	25	25	25	25	25	25	25	
Dataset B	Train	660	660	660	660	660	660	660	660	660	660	2
	Test	25	25	25	25	25	25	25	25	25	25	
Dataset C	Train	660	660	660	660	660	660	660	660	660	660	3
	Test	25	25	25	25	25	25	25	25	25	25	
Dataset D	Train	1980	1980	1980	1980	1980	1980	1980	1980	1980	1980	1,2,3
	Test	75	75	75	75	75	75	75	75	75	75	

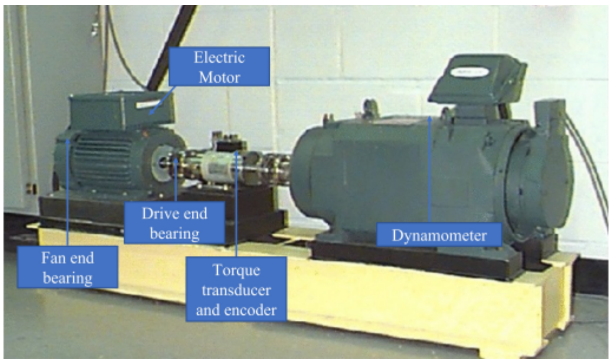

Fig. 4. CWRU Bearing Simulator.

B. Evaluation Metrics

Accuracy is the most intuitive indicator. The problem, however, is that unbalanced data labels can skew performance. The equation for this parameter is:

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |FP| + |FN| + |TN|} \tag{1}$$

The recall is the ratio of a class to what the model predicts as true among those that are actually true. The recall can be expressed by the following equation:

$$Recall(Sensitivity) = \frac{|TP|}{|TP| + |FN|} \tag{2}$$

Precision is the proportion of what the model classifies as true that is actually true. Precision can be expressed by the following equation:

$$Precision = \frac{|TP|}{|TP| + |FP|} \tag{3}$$

The f1-score is the harmonic average of precision and recall. When the data labels are unbalanced, the performance of the model can be accurately evaluated. The f1 score can be expressed in the following equation:

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

C. Results

Fig 5 shows accuracy comparison between the models. These accuracy figures are averaged over 20 iterations of 5-shot. It can be seen that the accuracy of the model presented in this paper is high in all samples. This is especially true for samples 90, 120, 200, and 300. However, as the number of samples increases, the accuracy of the existing model is also high, so it can be seen that there is not much difference with the model presented in this paper.

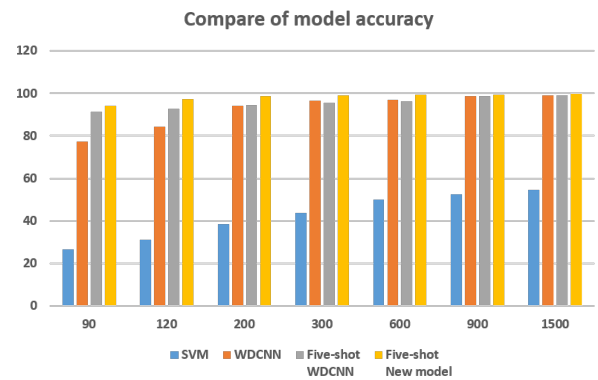

Fig. 5. Compare of Model Accuracy

Table IV shows the accuracy comparison for each model.

Diagnostic results of the proposed model training with different number of training samples, SVM, WDCNN, 5-shot(WDCNN) [16], and comparison graph. In all samples, the accuracy of the model proposed in this paper is the highest. However, it can be seen that the general WDCNN model is higher than 5-shot in 300, 600, and 900 samples.

TABLE IV
ACCURACY OF MODEL

	90	120	200	300	600	900	1500
SVM	26.56	31.2	38.67	43.89	50.05	52.35	54.53
WDCNN	77.39	84.19	93.97	96.59	97.03	98.69	98.87
Five-shot WDCNN	91.37	92.66	94.32	95.65	96.14	98.55	99.13
Five-shot New model	94.06	97.22	98.5	99.13	99.38	99.34	99.58

Table V shows the F1-score as a function of the number of samples. F1-score is a machine learning metric used in classification models. In all samples, the F1-score is better than 94% indicating that the data is not imbalanced.

TABLE V
F1-SCORE OF MODEL

	90	120	200	300	600	900	1500
F1-score	94.06	97.27	98.51	99.11	99.37	99.32	99.59

Table VI shows the accuracy comparison for each model. The model of WDCNN block3 is model in [16] accuracy.

Ref. [16] the block3 model is less accurate than the block 5 [16] model. However, in the model presented in this paper, although the number of blocks is 3 (convolution layer+maxpooling), the accuracy is sharply improved by adding batchnormalization, dropout, and LSTM to the existing block3 model, and the accuracy is higher than the existing block3 model and block5 model. Also, Table VII we can see that the F1-score is high.

Confusion matrices are used to train a model using a training set and then evaluate its performance using a test set when

TABLE VI
COMPARE OF MODEL ACCURACY

Block3	90	120	200
WDCNN	74.47	78.46	86.16
WDCNN+LSTM	94.06	97.22	98.5

TABLE VII
COMPARE OF MODEL F1-SCORE

Block3	90	120	200
WDCNN	74.47	78.46	86.16
WDCNN+LSTM	94.06	97.22	98.5

training and testing. The confusion matrix shows the predicted result of a sample on the horizontal axis and the actual label of the sample on the vertical axis. In this paper, the confusion matrix represents the value when testing once for 5-shot. on the horizontal axis and the sample's actual label on the vertical axis. In this paper, the confusion matrix represents the values when testing once for 5-shot.

Fig 6, Fig 7, Fig 8 shows the confusion matrix results for samples 120, 200, and 3000. The higher the number of samples, the better the performance.

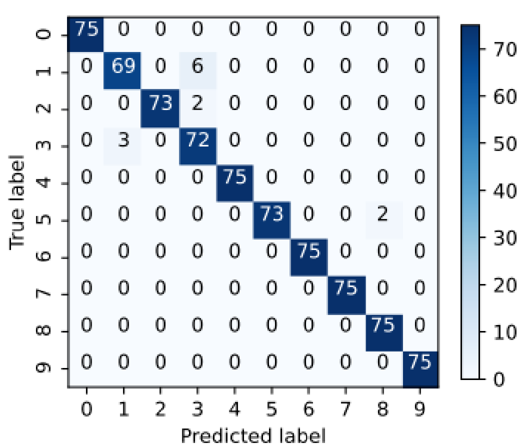

Fig. 6. Samples 120 Confusion Matrix.

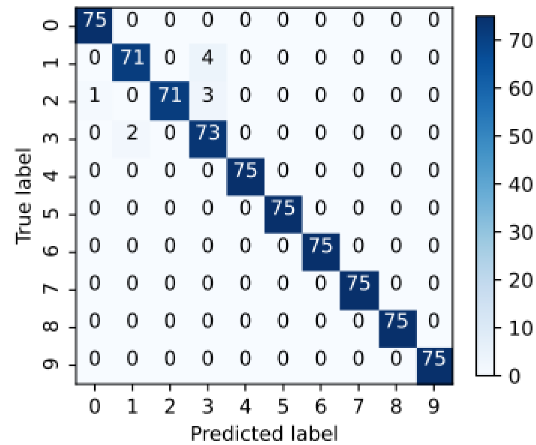

Fig. 7. Samples 200 Confusion Matrix.

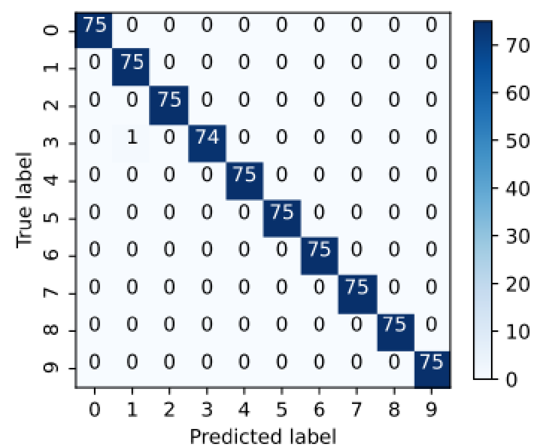

Fig. 8. Samples 3000 Confusion Matrix.

CONCLUSION

In this paper, a Siamese network-based WDCNN+LSTM model is proposed, and the accuracy of bearing fault diagnosis is verified using a push-pull learning algorithm and CWRU dataset. The proposed algorithm and model improve the performance of the existing model by about 4%. In future research, we can validate the proposed model in this paper with other open datasets. In addition to open datasets, we can also utilize data from actual manufacturing sites with noise to measure the accuracy of bearing fault diagnosis. Second, we can consider combining not only LSTM but also new models such as attention module and transformer.

REFERENCES

[1] Zhu. Y, chen. H, Meng. W, Xiong. Q and Li. Y, "A wide kernel CNN-LSTM-based transfer learning method with domain adaptability for rolling bearing fault diagnosis with a small dataset," Advances in Mechanical Engineering, vol. 14, pp. 1-20, November 2022.

[2] Lee. S and Jeong. J, " SSA-SL Transformer for Bearing Fault Diagnosis under Noisy Factory Environments," Journal of electronics, vol. 11, pp. 1-21, May 2022.

[3] Widodo. A and Yang. B. S, "Support vector machine in machine condition monitoring and fault diagnosis," Mechanical systems and signal processing, vol. 21, pp. 2560-2574, August 2007.

[4] Zhou. D. H and Frank, P. M, "Fault diagnostics and fault tolerant control," IEEE Transactions on aerospace and electronic systems, vol. 34, pp. 420-427, April 1988.

[5] Zhao. B, Zhang. X, Li. H and Yang. Z, "Intelligent fault diagnosis of rolling bearings based on normalized CNN considering data imbalance and variable working conditions," Knowledge-Based Systems, vol. 199, July 2020.

[6] Liu. H, Zhou. J, Zheng. Y, Jiang. W and Zhang. Y, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," ISA transactions, vol. 77, pp. 167-178, June 2018.

[7] Mao. W, Liu. Y, Ding. L and Li. Y, "Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study," Ieee Access, vol. 7, pp. 9515-9530, January 2019.

[8] K. Yip and G. Sussman, "Sparse Representations for fast, One-Shot Learning," National Conference on Artificial Intelligence, pp. 1-29, July 1997.

[9] Wang. Y, Yaol. Q, Kwok. J. T and Ni. L. M, "Generalizing from a few examples: A survey on few-shot learning," ACM computing surveys, vol. 53, pp. 1-34, June 2020.

[10] L. Fei-Fei, R. Fergus and P. Perona, "One-shot learning of object

categories," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 594–611, February 2006.

[11] M. Fink, "Object classification from a single example utilizing class relevance metrics," In Advances in Neural Information Processing Systems, vol 17, 2004.

[12] Koch. G, Zemel. R and Salakhutdinov. R, "Siamese Neural Networks for One-shot Image Recognition", In ICML deep learning workshop, vol. 2, pp. 1-27, July 2015.

[13] Chopra. S, Hadsell. R and LeCun. Y, "Learning a similarity metric discriminatively, with application to face verification," In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 539-546, June 2005.

[14] Chen. H, Cen. J, Yang. Z, Si. W and Cheng. H, "Fault Diagnosis of the Dynamic Chemical Process Based on the Optimized CNN-LSTM Network," ACS omega, vol. 7, pp. 34389-34400, september 2022.

[15] Zhang. A, Li. S, Cui. Y, Yang. W, Dong. R and hu. J, " Limited Data Rolling Bearing Fault Diagnosis With Few-Shot Learning," IEEE Access, vol. 7, pp. 110895-110904, August 2019.

[16] Lee. D and Jeong, J. "Bearing Fault Detection based on Few-Shot Learning in Siamese Network," in WSEAS Transactions on Systems, vol. 21, pp. 276-282, December 2022.