Anomaly Detection based on 1D-CNN-LSTM Auto-Encoder for Bearing Data

DAEHEE LEE[1], HYUNSEUNG CHOO[1], JONGPIL JEONG[2]
[1]Department of Electrical and Computer Engineering, Sungkyunkwan University,
2066, Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do, 16419
REPUBLIC OF KOREA

[2]Department of Smart Factory Convergence, Sungkyunkwan University,
2066, Seobu-ro, Jangan-gu, Suwon-si, Gyeonggi-do, 16419
REPUBLIC OF KOREA

Abstract: - The manufacturing industry is developing rapidly due to the Fourth Industrial Revolution. If a piece of bearing equipment, which is one of the essential parts of the manufacturing industry, fails, it will hinder the production of the manufacturing industry, which will lead to huge losses for the company. To prevent this, this paper implements a 1 Dimension-Convolution Neural Networks-Long Short-Term Memory (1D-CNN-LSTM) Auto-Encoder model for fault diagnosis of bearing data. The 1D-CNN-LSTM Auto-Encoder model showed high accuracy of 58 to 100 percent for eccentric bearing data that are difficult to visually diagnose as faults. In the future, we would like to extend this to a real-time failure diagnosis system that can remotely monitor the condition of the bearing equipment through real-time communication with the cloud server and test bed.

Key-Words: - 1D-CNN, LSTM, Auto-Encoder, Unsupervised Anomaly Detection, Bearing Data, Smart Factory

Received: March 29, 2022. Revised: October 25, 2022. Accepted: November 27, 2022. Published: January 9, 2023.

1 Introduction

The manufacturing industry is developing at a high speed due to the 4th industrial revolution. As the manufacturing industry develops, the value and importance of bearing equipment are also increasing. If the bearing device fails, the product manufacturing process will be disrupted, which can lead to huge losses for the company. If the bearing equipment is predicted to fail and the equipment is replaced in advance, it will be a great advantage in industrial and economic aspects by increasing the production efficiency of the company.

In most companies, skilled technicians decide whether to replace the product by listening to the sound of the equipment, or by judging whether the bearing equipment fails based on their standards and know-how, such as their replacement period. This approach does not use 100 percent of the efficiency of the equipment, and another problem arises when a skilled technician retires and another technician comes. Recently, COVID-19 has led to rapid changes in industrial sites, including an increase in non-face-to-face work and a decrease in field technicians. It is time for an AI algorithm model to diagnose failures with objective criteria, not failure diagnosis of equipment using the know-how of skilled technicians.

To initially diagnose the failure of bearing equipment, various studies using artificial neural networks, [1], and genetic algorithms, [2], are being conducted. In particular, fault diagnosis research, [3], based on unsupervised learning is being intensively conducted.

This paper proposes a method for diagnosing failures of rotors using a more advanced 1 Dimension Convolution Neural Networks-Long Short-Term Memory (1D-CNN-LSTM) Auto-Encoder model from the LSTM Auto-Encoder, [4], model, one of the unsupervised learning-based models. Unsupervised learning proceeds with only normal data and judges normal and abnormalities for new data, which is suitable for fault diagnosis in industrial sites where failures do not easily occur. In smart factory sites, diagnosing failures is one of the most important factors, with the 1D-CNN-LSTM Auto-Encoder showing 58 to 100 percent accuracy for eccentric data that is difficult to diagnose failures.

This paper consists of the following. Section 2 briefly describes the models used to diagnose faults in bearing equipment: 1D-CNN, LSTM, Auto-Encoder, and Unsupervised Anomaly Detection. Section 3 describes the structure and hyperparameter for the 1D-CNN-LSTM Auto-

Encoder model. Section 4 details the test bed used, the bearing data experimental environment, data extraction and processing methods, and experimental results. In Section 5, the conclusion is made based on the results obtained in Section 4.

2 Related Work

2.1 1D-CNN

CNN (Convolution Neural Network) is a type of deep learning algorithm, specialized in processing data arranged in a grid shape, and is an effective neural network for identifying patterns of data. Therefore, CNN utilizes several filters that can be used as shared parameters to maintain spatial information of images in two dimensions and effectively extract and learn features with adjacent images. CNN has the advantage of enabling simpler learning through minimal parameters and preprocessing. Among them, 1D-CNN (1 Dimension Convolution Neural Networks), [5], is often used for time-series analysis or text analysis rather than images. One-dimensional means that the kernel for the synthetic product and the sequence of data to be applied have a one-dimensional shape.

2.2 LSTM

LSTM, [6], stands for Long Short-Term Memory and is a model generated to address the long-term dependence of Recurrent Neural Networks (RNNs) used for learning time series data. LSTM has the advantage of storing and utilizing information on all input data, so it is widely applied to time series data processing. LSTM is used to solve the vanishing gradient problem and is advantageous for long time preprocessing. The simultaneous use of 1D-CNN and LSTM allows for large extraction of time series properties, which can be expected to improve fault diagnosis accuracy of bearing equipment.

2.3 Auto-Encoder

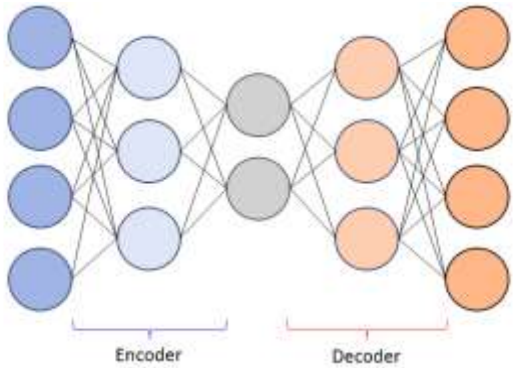

Fig. 1: Undercomplete Auto-Encoder model

Auto-Encoder, [7], reconstructs and outputs input and output values through Encoder and Decoder, and the loss function is calculated with the difference between input and output values. Figure 1 is a representative Auto-Encoder picture. Since the node (unit) of the hidden layer is smaller than the input layer, the input is expressed in low dimensions, such a model is called Undercomplete Auto-Encoder. Since Undercomplete Auto-Encoder cannot copy an input directly to the output by a hidden layer with low dimensions, it learns the most important feature from the input data to output something like the input. With Auto-Encoder, you can learn the features of the normal area, which is the main component of the data, without labeling the data. At this time if normal data is put into the learned Auto-Encoder, the difference between Input and Output hardly occurs because if abnormal data is put in, the difference between Input and Output is noticeable in the process of calculating the difference between Input and Output, so abnormal data can be detected.

2.4 Unsupervised Anomaly Detection

Anomaly Detection can be divided into three parts: Supervised Anomaly Detection, Semi-supervised (OneClass) Anomaly Detection, and Unsupervised Anomaly Detection. In this paper, we use Unsupervised Anomaly Detection [8]. Unsupervised Anomaly Detection requires a process of securing a label for a normal sample to know which data is a normal sample among numerous data, and most of them are learned without acquiring a separate label under the assumption that data is a normal sample. Unsupervised Anomaly Detection uses label-free data and allows users to perform more complex processing tasks. It can also be used to discover the underlying structure of the data and is advantageous for real-time data processing.

3 1D-CNN-LSTM Auto-Encoder Model

In this paper, we propose a 1D-CNN-LSTM Auto-Encoder model for bearing data. Auto-Encoder is widely used for data generation or restoration, and after model learning with normal data, test data is inserted to calculate the difference between normal data and decoded test data. Through this difference, it can be determined as normal and abnormal. In this paper, this Anomaly Detection method was applied to time series data. The model of this paper is shown in Figure 2. The layer of Auto-Encoder was configured as LSTM to enable sequence learning. In addition, by applying the 1D-CNN layer, learning was configured to proceed while moving the timestamp and feature information in detail. The structure of the model is designed so that the encoder and decoder are symmetrical with the 1D-CNN - Dense layer - LSTM - Dense layer. The maximum value of Train loss was set to the threshold value and compared with the loss value of the test data, and then the data was determined whether it was normal or abnormal. Experiments were conducted while adjusting the filters and kernel size of 1D-CNN and the unit values of Dense and LSTM.

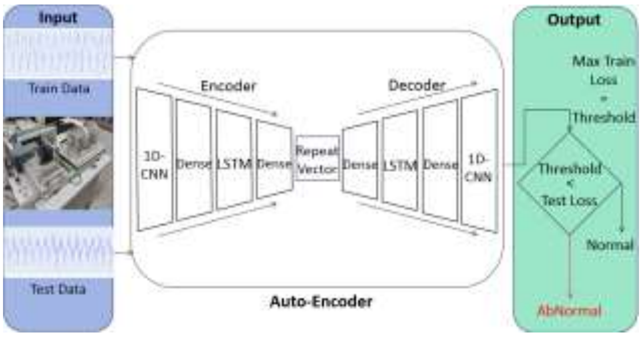

Fig. 2: 1D-CNN-LSTM Auto-Encoder Model Structure

The hyperparameter values of the 1D-CNN-LSTM Auto-Encoder model are shown in Table 1. Experiments were conducted with various hyperparameters, and the following cases showed the best performance. In the 1D-CNN layer, padding used the Relu function the same as the activation function. In addition, the dropout rate was set to 0.2 to prevent overfitting.

Table 1. 1D-CNN-LSTM Auto-Encoder Hyperparameter

Layers	Configurations
1D CNN layer	filters = 128, kernel-size = 32
Dense layer	filters = 64
LSTM layer	filters = 64
Dropout	rate = 0.2
Dense layer	filters = 32
Report Vector	Sequence Size = 31
Dense layer	filters = 32
Dropout	rate = 0.2
LSTM layer	filters = 64
Dense layer	filters = 64
1D CNN layer	filters = 128, kernel-size = 32
Time Distributed	filter = 1

We set up a model with a 1D-CNN-LSTM Auto-Encoder structure, which is expected to improve the fault diagnosis accuracy of rotors by extracting features that are advantages of Auto-Encoder and by extracting time series properties that are advantages of 1D-CNN and LSTM.

4 Experiment and Results

4.1 Experiment Environment

The experimental data of this paper used Sewoo Industrial System BLDC Motor (Figure 3) which combines one rotor motor and two rotors as a test bed. The rotor motor used a BLDC motor with specifications of Flange Size 90, Poles 12, Input 220V, and Output 220W made at Sewoo Industrial. Data extracted from the BLDC Motor was stored as a CSV-formatted file through oscilloscope equipment, and post-processing was applied secondarily. Train data and test data extracted values every 0.001 seconds and consisted of 9,982 data per experiment. The period of the data is 0.031 seconds, and if one data detects an outlier in this period, it was determined that a failure occurred in that period. Experiments in this paper were conducted at Google Colab, and the CPU used Intel (r) Xeon (R) 2.00g Hz (dual-core) CPU and NVIDIA tesla t4 (8 GB) GPU. Python Version is 3.7.15 and Cuda Version is 11.2.

Fig. 3: Rotating Motor Test Bed for Bearing Data

Looking at Figure 4, there is an acceleration sensor and two rotating plates (A and B), and rotating plate A is relatively closer to the acceleration sensor than rotating plate B. Since the experiment results differ depending on the position of the acceleration sensor and the rotating body, the experiment was divided into three parts, [9], when only rotating plate A is weighted, when rotating plate B is weighted, and when rotating plate A and B are simultaneously weighted.

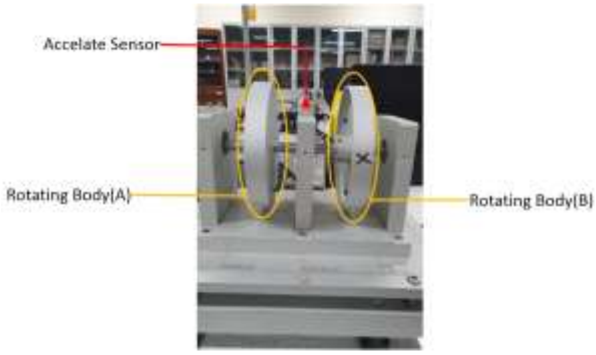

Fig. 4: Test Bed Accelerate Sensor and Rotating Body (A, B)

Eccentricity, [10], refers to a state in which the centers of an object are biased to one side and the centers are not aligned with each other. Looking at Figure 5, if there are 36 holes in the rotating plate that can insert screws, the interval between each hole is 10 degrees. After inserting two screws, a fault diagnosis experiment was conducted according to the change in the gap between each screw. If the two screws achieve 180 degrees, the sum of the weight vectors is zero, so there is no eccentricity. The absence of eccentricity means a stable state. However, if the angle of the screw is not 180 degrees, the vector sum of the two screws is not zero, so there is an eccentricity. As the angle between the two screws decreases, the eccentricity gradually increases. As the eccentricity increases, the sum of the weight vectors increases, making it easier to diagnose failures, and on the contrary, as the eccentricity decreases, the sum of the weight vectors decreases, making it difficult to diagnose failures. In this paper, experimental results were prepared for 160 degrees and 170 degrees for the angles of two screws, which are generally difficult to diagnose faults, because the eccentricity of the two screws is large and failure diagnosis is possible for all data.
This paper’s experiment, [11], was conducted by dividing the rotating bodies A, B, and (A and B) into three simultaneously. Train data has to proceed with normal data, so the angle of the two screws is 180 degrees, that is, the case where there is no eccentricity. The test data was set when the angles of the two screws were 160 degrees and 170 degrees.

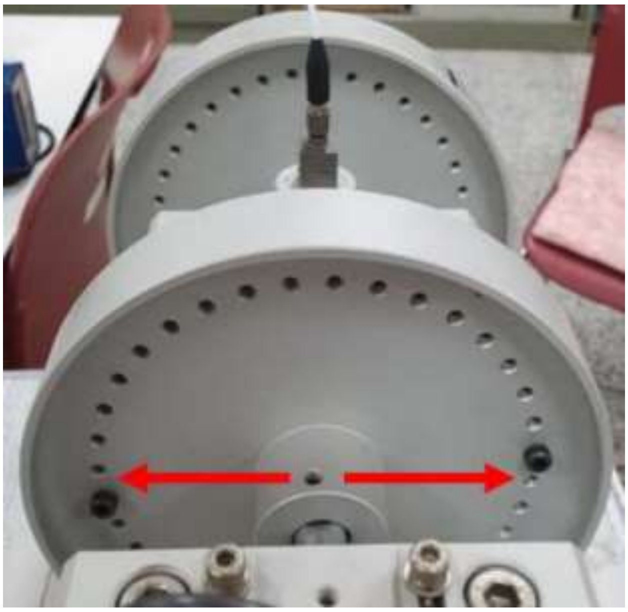

Fig. 5: Rotating Body 180 degrees

4.2 Performance Matrics

The experimental model evaluation confirmed the experimental results using the most commonly used Confusion Matrix in binary classification. The confusion Matrix has four evaluation methods True Positive (TP) means that true is classified as true. True Negative (TN) means that true is classified as false. False Positive (FP) means that false is classified as true. Finally, False Negative (FN) refers to a case where false appears as false. True Positive (TP) means that true is classified as true. True Negative (TN) means that true is classified as false. False Positive (FP) means that false is classified as true. Finally, False Negative (FN) refers to a case where false appears as false.

Accuracy (1) represents the ratio of the total number of samples to what the algorithm correctly predicted. For example, if the algorithm is 80 percent accurate, only 80 out of 100 samples are correctly classified.

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

Recall (2) is the ratio of true classes compared to what the model predicts. The parameters Recall and Precision have a trade-off relationship.

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

Precision (3) refers to the ratio of true classes to what the model classifies as true.

$$(\Pr e\,cision) = \frac{TP}{(TP+FP)} \tag{3}$$

F1-Score (4) is called harmonic mean and accurately evaluates the performance of the model when the data labels are unbalanced.

$$(F1 - Score) = \frac{2 * Precision * Recall}{(Precision + Recall)} \qquad (4)$$

All experiments used the Accuracy mentioned in the evaluation index, and F1-Score was used as an evaluation index to compensate for the shortcomings of accuracy.

4.3 Results

The 1D-CNN-LSTM Auto-Encoder model was applied to bearing data extracted from the Sewoo Industrial System test bed, and the accuracy shown in Table 2 was derived. Rotating plate A diagnosed the failure for most of the situations except for 180 degrees data of the angles of the two screws. In particular, when there are simultaneous eccentric data of rotating body B and rotating body (A and B) at the same time, failure was diagnosed in all situations.

Table 2. Results of Bearing Data

Bearing Data	Data A	Data B	Data (A, B)
170 degree	0	1	1
160 degree	0.58	1	1
150 degree	1	1	1

Figure 6 shows the confusion matrix when the angle of the two screws on the rotating body A is set to 180 degrees data as train and 160 degrees data of the two screws as a test on the rotating body A. The result of rotating body A is not as good as that of rotating body B or rotating body (A and B) at the same time because the distance between the acceleration sensor and rotating body A is relatively too close to that of rotating body B to detect a perfect failure.

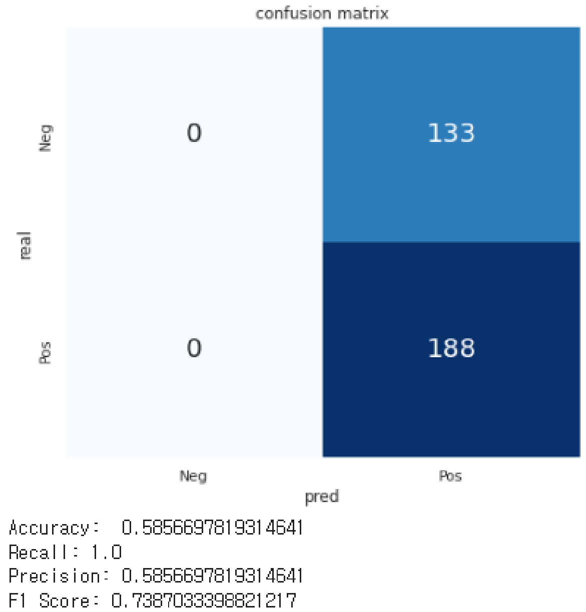

Fig. 6: Result of Rotating Body A 160 degrees data

Figure 7 shows the target confusion matrix for bearing data. In the case of a model using both rotating body B and rotating body (A and B), the target values were obtained for all cases, but only 150 degrees were obtained for rotating body A, and different results were obtained for 160 degrees and 170 degrees.

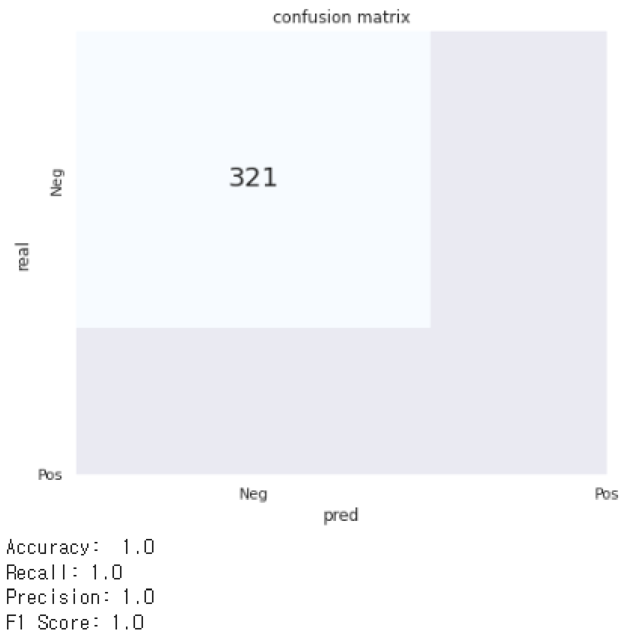

Fig. 7: Goal for Anomaly Detection

5 Conclusion

This paper proposes an artificial neural network using 1D-CNN-LSTM Auto-Encoder using actual measured bearing data. Learning on fine eccentric data that is generally difficult to distinguish, the 1D-

CNN-LSTM Auto-Encoder model proposed in this paper showed 58 to 100 percent accuracy. It is not easy to obtain fault data in the actual field, and the model proposed in this paper is an Unsupervised model, which has the advantage of being able to learn only with a normal sample. Failure of bearing data may occur in a misalignment-like manner, except for eccentricity. It may be set as an additional diagnostic failure evaluation element for the misalignment. In addition, the current experimental data is extracted with an oscilloscope rather than real-time communication, and the CSV file is used through secondary processing in a PC environment. As a plan, failure detection of bearing data can be made in real-time, [12], by linking the data value of the rotating body with DB.

Acknowledgement:
"This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1F1A1060054), the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2018-0-01417) and the ITC Creative Consilience Program (IITP-2022-2020-0-01821) supervised by the IITP (Institute for Information Communications Technology Planning Evaluation) supervised by the IITP (Institute for Information Communications Technology Planning Evaluation)" Corresponding author: Professor Hyunseung Choo and Jongpil Jeong.

References:
[1] M. Dix, A. Chouhan, S. Ganguly, S. Pradhan, D. Saraswat, S. Agrawal, and A. Prabhune, "Anomaly detection in the time-series data of industrial plants using neural network architectures", *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*, 2021, pp.222-228.
[2] Wanjuan Song, Wenyong Dong, and Lanlan Kang, "Group anomaly detection based on Bayesian framework with genetic algorithm", *Information Sciences*, 2020, pp. 138-149.
[3] Subutai Ahmad, Alexander Lavin, Scott Purdy, and Zuha Agha, "Unsupervised real-time anomaly detection for streaming data", *Neurocomputing*, 2017, pp. 134-147.
[4] B. Hou, J. Yang, P. Wang, and R. Yan, "LSTM Based Auto-Encoder Model for ECG Arrhythmias Classification", *IEEE Transactions on Instrumentation and Measurement*, 2020, pp. 1232-1240.
[5] Eren, L., Ince, T, and Kiranyaz, S, "A Generic Intelligent Bearing Fault Diagnosis System Using Compact Adaptive 1D CNN Classifier", *Journal of SignalProcessing Systems*, 2019, pp. 179–189.
[6] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM Fully Convolutional Networks for Time Series Classification", *IEEE Access*, 2018, pp. 1662-1669.
[7] Yasi Wang, Hongxun Yao, and Sicheng Zhao, "Autoencoder based dimensionality reduction", *Neurocomputing*, 2016, pp. 232-242.
[8] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series", *IEEE Access*, 2019, pp. 1991-2005.
[9] H. Im, S. Kim, S. Jung, S. Hong, G. Oh and J. Park, "Analysis of Vibration Signal for Failure Diagnosis of Rotating Devices", *Journal of Korean Society for Precision Engineering*, 1995, pp. 301-307.
[10] X. Gu and P. Velex, "On the dynamic simulation of eccentricity errors in planetary gears", *Mechanism and Machine Theory*, 2013, pp. 14-29.
[11] Daehee Lee, Jaehoon Lee, Jinho Park, Jongin Choi, and Taeyoung Choe, "Anomaly Detection in Rotating Motor using Two-level LSTM", *Proceedings of KIIT Conference*, 2020, pp. 425-428.
[12] Mantere, M. Sailio, and M. Noponen, "Network Traffic Features for Anomaly Detection in Specific Industrial Control System Network", *Future Internet 2013*, 2013, pp. 460-473.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself
"This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1F1A1060054), the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2022-2018-0-01417) and the ITC Creative Consilience Program (IITP-2022-2020-0-01821) supervised by the IITP (Institute for Information Communications Technology Planning Evaluation) supervised by the IITP (Institute for Information Communications Technology Planning Evaluation)" Corresponding author: Professor Hyunseung Choo and Jongpil Jeong.