PREDICTIVE MAINTENANCE OF BALL BEARINGS USING CONVOLUTIONAL NEURAL NETWORKS (CNN)

Arsema Derbie and Kibru Temesgen

School of Mechanical and Industrial Engineering, Addis Ababa Institute of Technology, Addis Ababa University, Addis Ababa, Ethiopia

Corresponding author E-mail address: arsema.derbie@aau.edu.et

DOI: https://doi.org/10.20372/zede.v41i.8629

ABSTRACT

To this day, numerous maintenance actions follow preventive and run-to-failure methods. In this work, it has been attempted to show the power of predictive maintenance (PdM) from vibration data using a machine learning technique implementing convolutional neural networks (CNN). Actual data was collected from six different bearings on a machine element fault analysis test rig. The bearings were of the ball bearing type, where one of them was healthy and the rest were faulty at the inner race, the outer race, the rolling element, or a combination of these three, and another had severe damage at either the rolling element or the rings. From the vibration signatures specific to the health status of the bearings, a powerful deep learning convolutional neural network model was built. The model was able to successfully classify the states of bearings with accuracy values ranging between 76.7 and 99.9% based on unseen data. The results indicate that this CNN model can be used as a diagnostic tool for undertaking maintenance operations.

Keywords: CNN, predictive maintenance

1. INTRODUCTION

Monitoring the state of machinery and its health condition is one of the top priorities of companies since a huge amount of money is invested in equipment. Companies want their equipment to last as long as possible, and predictive maintenance is the best and Prospective way of ensuring that. With the advancement of artificial intelligence (AI), the lowering cost of sensors, and the increasing connectivity among devices because of the explosion of the internet, now is the time to take advantage of AI for maintenance.

A maintenance program often falls into a few categories, each with its own set of challenges and benefits. PdM is one of them. It is a collection of approaches that employ condition-monitoring technologies to monitor a machine's performance and discover potential flaws before it breaks. Too early maintenance can result in money being spent unnecessarily, while too late maintenance might result in catastrophic equipment damage.

Recently, studies on bearing prognostics can be seen as a problem of pattern recognition, and several research attempts have been made to develop techniques for machine health monitoring [1]. Hrnjica et al. [2] describe an example of an explainable AI (XAI) in the form of a PdM scenario for manufacturing. PdM is performed using machine learning algorithms to predict when maintenance should be performed on equipment before it fails. The authors used CNN to predict when a bearing will fail based on vibration data. Due to improvements in machine learning algorithms and the accessibility of vast volumes of data, the application of AI in PdM has grown in popularity in recent years.

Unexpected bearing failures may force companies to pay for repairing and replacing the bearing and adjacent components, which may also sustain damage, such as housings and shafts. Bearing failures reduce a plant's operating efficiency, increase downtime, drive the cost of operations up, and, in the worst cases, may injure workers. Unplanned maintenance increases the chance of worker injury. With workers, 28% more likely to have an accident at work when performing reactive maintenance over proactive maintenance, the impact that premature bearing failure can have on worker safety is clear [3].

In manufacturing plants, optimal maintenance strategies are necessary to ensure system reliability, reduce cost, avoid downtime, and maximize the useful life of a component [4]. According to a recent article, unplanned downtime caused by poor maintenance strategies reduces a plant's overall productive capacity by up to 20 percent and costs around $50 billion each year [5].

Many researchers have studied condition monitoring (CM) using CNN and Deep Neural Networks (DNN). Zhang et al. [6] showed the domain adaptability of classifying different bearing vibration signals using CNN. It is one of the most notable deep learning models due to its shared weights and ability for local field representation [7]. CNN can extract the local features of the input data and combine them layer by layer to generate high-level features. In the field of predictive maintenance, CNN has shown dramatic capability in extracting useful and robust features from monitoring data. For one-dimension (1D) monitoring signals, Qin et al. [8] built an end-to-end 1D-CNN that reflected the raw vibration signals to fault types. The result showed that the proposed model was able to achieve about 99% accuracy through hyper parameter tuning. Furthermore, the applications of CNN in remaining useful life prediction have been widely investigated.

There has been immense success in the application of CNN to image and acoustic data analysis. In this paper, rather than preprocessing vibration signals to de-noise or extract features, they investigated the usage of CNNs on raw signals; in particular, they tested the accuracy of CNNs as classifiers on bearing fault data by varying the configurations of the CNN from a one-layer up to a deep three-layer model. They also inspected the convolution filters learned by the CNN and showed that the filters detect unique features of every classification category.

In addition, they studied the effectiveness of the various CNN models when the input signals were corrupted by noise [9]. Guo et al. [10] proposed a deep hierarchical architecture of CNN in which original data was converted into 2D data to classify bearing faults and their sizes. Many of the published works of CM with CNN approaches show very high accuracy, but they were mostly tested on the same dataset.

Access to datasets is a difficulty while investigating machine learning (ML) and deep learning (DL) algorithms for fault diagnosis because it is difficult and expensive to build a realistic mechanical fault dataset-producing test-bench. The dataset created by Case Western Reserve University (CWRU) is the most well-known and easily accessible dataset for vibration-based rolling bearing failure diagnostics and has been used as the standard reference in numerous papers. In their research, Neupane and Seok [11] examined numerous articles on deep learning algorithms employing the CWRU dataset. Smith and Randall [12] studied the complete CWRU dataset and provided benchmark recommendations for diagnostic methodology.

Mohammad et al. [13] proposed a time-moving segmentation window to segment the raw vibration signal, and the segmented signals were decomposed up to two levels using the discrete wavelet transform (DWT). After that, decomposed signals were converted into grayscale images to train and test the proposed CNN model. To verify the performance of the model, the CWRU bearing dataset and the MFPT dataset were used. The proposed CNN model achieved the highest accuracy in terms of performance under different load conditions as well as in noisy situations with varying signal-to-noise ratio values. The experimental findings showed that the proposed system was effective and extremely dependable in detecting bearing faults. Junjie et al. [14] used a CNN based on improved soft maximum loss (ISM-CNN).

The constructed CNN could learn more subtle features from the bearing signals, thereby improving the accuracy of bearing signal classification. Besides, the algorithm proposed in the same paper expanded the training data set to a certain extent, so that the parameters of the ISM-CNN could be better fitted. They validated the effectiveness of the proposed algorithm on the CWRU open dataset and performed ablation experiments to prove it.

The target of this research study was to develop a PdM model for ball bearings using CNN to predict the health status of bearings before the equipment fails by using the vibration signal generated from the bearings.

2. MATERIALS AND METHODS

2.1 Test-rig and Dataset

Figure 1 shows the experimental setup that was used. The shaft was connected to a motor through a coupling. The motor could be operated to a maximum speed of 4100 rpm. The shaft was connected to the two bearing blocks. The bearing block immediately adjacent to the motor housed a normal (healthy) bearing, which was not changed during the entire data acquisition process. The second bearing block housed bearings with different types of damage during the measurement. Bearings were fixed into the block by using a retaining ring. The bearing blocks had spaces for the mounting of accelerometers.

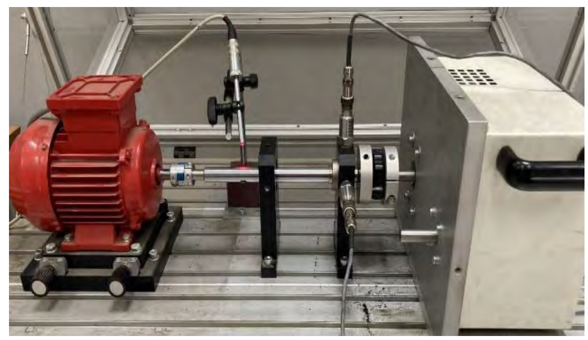

Figure1 Experimental test-rig of PT 500.12

An accelerometer was fixed to the bearing block in such a way that there would be no relative motion between them. The threaded section was located on the top and sides of the bearing block to measure both the horizontal and vertical vibration. The accelerometer (B&K Vibro’s AS-020) was used to measure the vibration characteristics. It was mounted on the bearing block that contained the damaged bearings.

The accessory setup contained six roller bearings, on which different faults could be detected and explained. The accessory setup was mounted on the base plate of the machinery diagnostic base system PT 500.12 roller bearing faults kit.

A photo-contact tachometer was used to measure the speed of the shaft. Six different bearings with different fault types were used in this test (Figure 2 and 3). After mounting them on the bearing block, vibration patterns were recorded for each type.

The test was done at a shaft speed of 1500 rpm and a torque of 0.1 Nm. The vibration data was recorded at a sampling rate of 15 kHz for 10 seconds, which resulted in 150,000 data points. The output of the accelerometer was in units of voltage. This voltage was then converted into a vibration unit known as "G,", which is an acceleration unit.

The conversion was done by using the sensitivity of the accelerometer. The sensitivity of an accelerometer is defined as the ratio of the output voltage to the acceleration being measured. After the conversion, the dataset was directly fed into a CNN architecture without the need for any preprocessing.

Figure2 Six types of Bearings (SKF 6004) with different health status

The faults were created by introducing defects in the bearings. The defects were introduced using an electro-discharge machining (EDM) by GUNT, the manufacturer of the PT 500.12 test-rig process to create artificial defects in five different parts of the bearings.

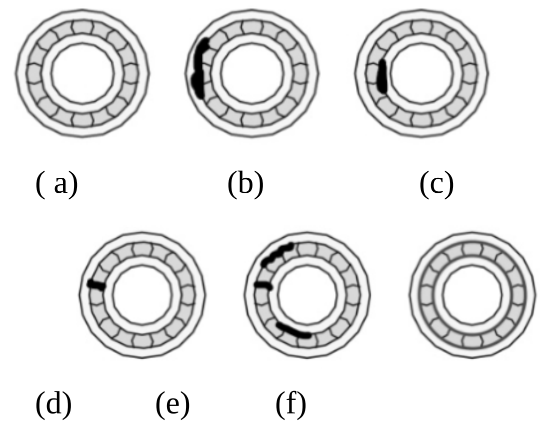

Figure 3 Types of faults and their locations; a) healthy bearing, b) damage on outer race, c) damage on inner race, d) damage on roller body, e) damage on roller body, outer and inner race, f) heavily worn bearing

2.2 Proposed CNN Architecture for Bearing Fault Classification

CNN uses the convolution operation in its architecture. This operation was used to extract features from the input data. The convolution of two functions f and g is denoted by f * g and defined as the integral of the product of the two functions after one is reflected about the y-axis and shifted. The formula for the convolution operation is given by:

$$(f*g)(t) = f(z) f(t - z) dz \tag{1}$$

where f and g are two functions, t is a variable, and z is a dummy variable for integration.

ReLu is a non-linear activation function that is used in CNN. It is used to introduce non-linearity in the output of a neuron and helps prevent over fitting. This function can be represented as:

$$f(x) = Max(0, x) \tag{2}$$

A 1-D CNN with multiple (parallel) inputs of the same data was used for this study.

Three different kernel sizes were used for the same input data. This was achieved by using three parallel convolutional layers. After the inputs passed through the three paths, they were combined and fed into a fully connected layer, which was followed by another fully connected layer and an output layer (Figure 4).

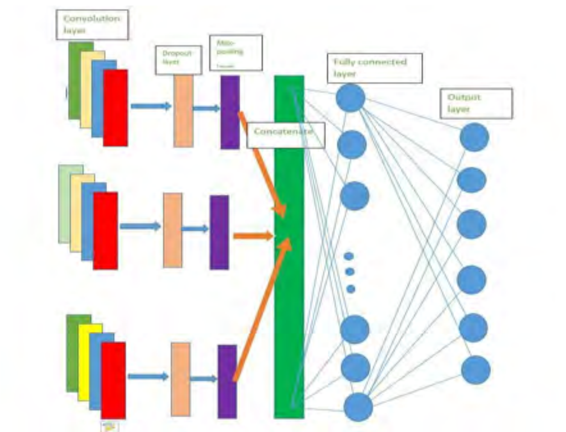

Figure 4 Proposed CNN architecture for 6 classes

2.2.1 Along upper path

A 1-D convolutional layer with 64 filters, with each filter having a kernel size of (200, 200), was used. A drop-out layer was used to avoid over-fitting, and the drop-out rate was set to 50%. This means that during each forward pass, 50% of the neurons were randomly deactivated. A max-pooling layer with a pooling size of 20 was also used. Max pooling would reduce the dimensionality of the output from the convolutional layers. This was important to make the model less sensitive to changes and more robust. It was equivalent to creating a lower resolution of the output while still retaining significant information. Networks with ReLU activation show better convergence, less vanishing, and fewer constant gradient problems, which has made them the best choice on CNN.

2.2.2 Along middle path

1-D convolutional layer with 64 filters, with each filter having a kernel size of 100 by 100, was chosen. A drop-out layer of 50% and a ReLU activation function were used.

2.2.3 Along lower path

1-D convolutional layer with 64 filters, with each filter having a kernel size of (50, 50), was used. Similarly, a drop-out layer of 50% and a ReLU activation function were used.

Along each path, different kernel sizes and max-pooling sizes were used to help the model retain different information along each path, thus making it robust to changes in the data for which the model was expected to predict. The input, after passing along the three paths, was flattened and concatenated together. The concatenated output was fed into a fully connected layer with 100 neurons and the ReLU activation function. After passing through the fully connected layer, the input was fed into the output layer to produce the prediction.

2.3 Training the Model

The input data to the 1D-CNN was augmented using a window size of 4500 data points and a stride of 60 data points. So, one input to the architecture was a 4500-by-1 matrix representing 0.3 seconds of vibration duration. This was crucial in creating the large number of input data points that the model was trained on.

The model was trained using the Google Compute Engine (GPU) through *Google Collaboratory*. The GPU provided by Google Colab is the Tesla K80 with 2496 CUDA cores and 12GB GDDR5 VRAM. This compute-optimized engine took no more than 45 minutes to train the model, unlike typical CPU computers, which would have taken more than 6 hours to train. The model was trained for a total of 200 epochs. The model was normally trained until it no longer showed improvement in accuracy. For this particular work, the model converged at an epoch of around 200 (Figure 5).

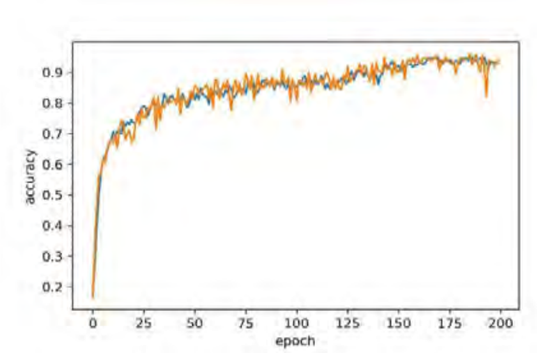

Figure5 Training accuracy Vs number of epochs over time

3. RESULTS AND DISCUSSION

3.1 Prediction

Two rounds of data were taken for this test in different sessions. During each round of data acquisition, the bearings were disassembled from the bearing block and reassembled in place. The accelerometers were unmounted. The model was built using the first round of data. In this work, the test data on which the model was evaluated was entirely from a different session. The purpose of doing so was to test the robustness of the model on completely unseen data.

The results of the prediction are shown in the next sections.

3.1.1 Ball bearing fault

The vibration data from the second session was given to the model to make a prediction, and the result was evaluated. The model was not trained on this data, but the class of the vibration data was known to be of the ball bearing fault type (Figure 6).

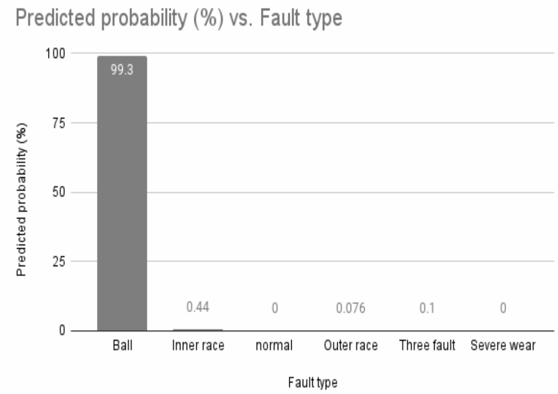

Figure 6 Ball fault Prediction on second round data

Accordingly, the model gave a probability of 99.3% that the bearing had a rolling element (ball) fault.

3.1.2 Inner race fault

A prediction confidence level of 92.7% that the bearing had an inner race fault was obtained (Figure 7).

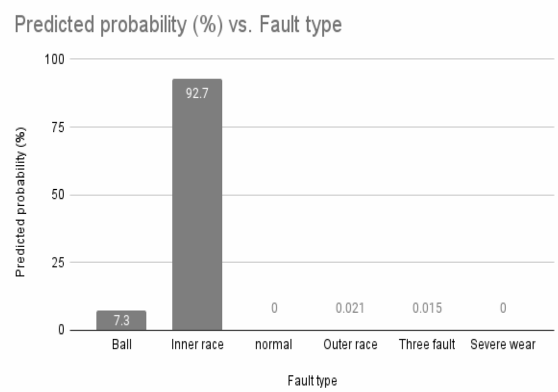

Figure 7 Inner race fault Prediction on second round data

3.1.3 Normal bearing

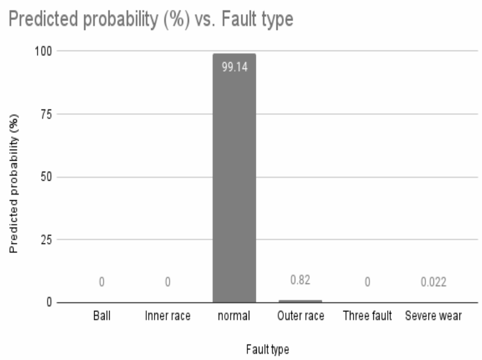

Figure 8 Normal bearing Prediction on second round data

A probability of 99.14% that the bearing was normal was obtained (Figure 8).

3.1.4 Outer race fault

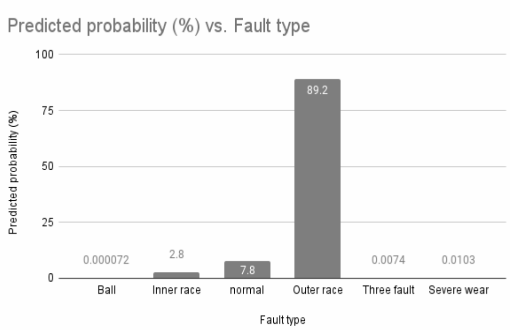

Figure 9 Outer race prediction on second round data

The model gave a probability of 89.2% that the bearing had an outer race fault (Figure 9).

3.1.5 Combination of three faults

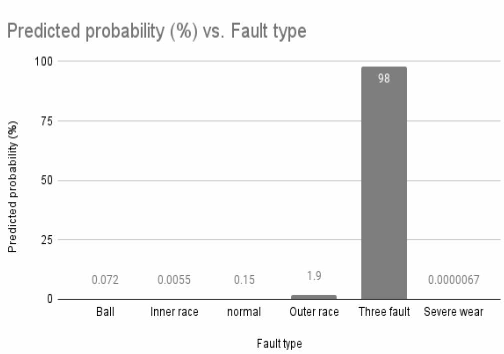

Figure 10 Combination of three fault prediction

The model gave a probability of 98% that the bearing had three fault combinations (Figure 10).

3.1.6 Severe wear faults

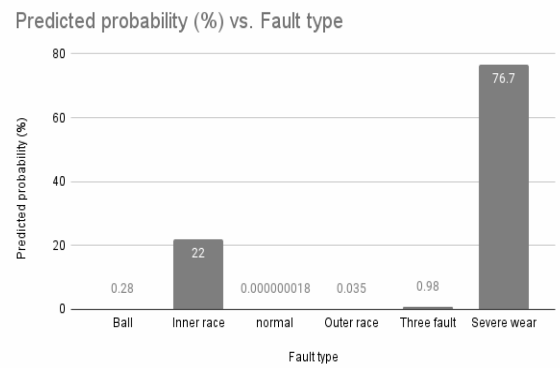

Figure 11 Severe wear fault prediction

The model gave a probability of 76.7% that the bearing had a severe wear fault (Figure 11).

3.2 Discussion

The prediction on the second round of data was quite satisfactory for all faults except in the case of severe wear. The model predicted the fault types with an accuracy of above 90% for most of the faults. The fault of the outer ring is usually difficult to predict because the rings are stationary and don't exhibit peculiar vibration patterns like the rest of the components. However, on the outer race, an accuracy of 89% was obtained, which is a good result.

For companies, it is usually not that important to know what type of fault is going to occur. The most important question is the health status of the bearing (healthy or faulty). The accuracy of the normal bearing was above 99%. This is significant for increasing the reliability of the model. The model is not susceptible to giving false positive results. It would have been a bad result if the model had predicted a normal bearing status as faulty. This would mean that operations are not halted for maintenance.

Fairly good accuracy has been obtained without using tiresome preprocessing techniques that used to be performed on ML techniques. Especially for the normal bearing health status, the model predicts with an accuracy of 99.9%, indicating its dependability in identifying whether a bearing is healthy or faulty. The CNN was able to construct feature representations that assisted in the categorization of "normal" and "faulty" data, outperforming previous approaches, which relied on feature selection, and displaying a higher accuracy rate [1-5, 16, 19].

Many papers validate their work by setting aside 30% of their data, of which 70% is allocated for training their model's performance [11, 12, 15, 17-19]. This study, however, used an entirely different session dataset to validate the robustness of the model.

Two fault classes (outer race fault and severe wear fault) had the lowest prediction accuracy: 89.2% and 76.7%, respectively. As pointed out previously, the outer race doesn't rotate as much as the rest of the components; therefore, it doesn't exhibit the peculiar vibration pattern associated with it. On the other hand, the severe wear fault doesn't have a localized fault, and the distribution of the fault over the whole surface doesn't make the bearing vibrate every cycle as with other fault types. Hence, this prevented the bearing from giving a purely distinct pattern.

The model's accuracy of over 99% in predicting the bearing's health status is a noteworthy accomplishment with broad ramifications for businesses that depend on machinery with bearings. Its excellent accuracy rate demonstrates the model's dependability in determining whether a bearing is in good condition or not. Outperforming earlier methods that depended on feature selection, the CNN was able to create feature representations that helped classify "normal" and "faulty" data. Also, this method does away with the necessity for time-consuming preprocessing methods that were previously utilized with ML techniques. The model's dependability in determining whether a bearing is healthy or faulty is demonstrated by its accuracy of 99.9% in predicting normal bearing health status.

The significance of this achievement cannot be overstated, as it has implications for increasing the reliability of machinery and reducing downtime due to unexpected failures. Companies can now have greater confidence in their machinery and plan maintenance schedules more effectively. This approach also eliminates the need for costly and time-consuming manual inspections that are often required to identify faulty bearings. The model's ability to predict normal bearing health status with an accuracy of 99.9% signifies its dependability in identifying whether a bearing is healthy or faulty and provides companies with greater confidence in their machinery.

4. CONCLUSIONS

The CNN framework proposed in this work can detect faults by using indicators that precede failures and advise maintenance. This work demonstrates the potential of deep learning methods for fault classification and bearing health condition monitoring.

This study has shown that although CNN networks were designed for image classification, they are sometimes even more powerful in classifying vibration patterns and tackling the problem of over fitting than artificial neural networks (ANNs).

The paper has also shown that deep learning (1D-CNN) is robust to noise and data variation. by showing acceptable levels of accuracy without the need for preprocessing (cleaning) of the vibration data and by learning complex patterns in the signal.

As the quantity of data available expands over time, deeper CNNs may be constructed, resulting in higher levels of feature representation. The model's accuracy can be improved if this is achievable.

CONFLICT OF INTEREST

The authors do not have any personal or financial interests.

ACKNOWLEDGEMENTS

We would like to thank God for everything, especially for the gift of life. We are also grateful to our families for their love and support. Finally, we would like to acknowledge the Addis Ababa Institute of Technology for providing us with the resources and facilities that we needed to complete this research, especially the mechanics lab.

REFERENCES

[1] Cerrada, M., Sánchez, R.V., Li Chuan, Pacheco, F., Cabrera D., Oliveira J.V.D. and Vásquez, R.E., "*A Review on Data-Driven Fault Severity Assessment in Rolling Bearings*", Mechanical Systems and Signal Processing, Vol. 99, 2018, p. 169-196.

[2] Bahrudin, H. and Softic, S., "*Explainable AI in Manufacturing: A Predictive Maintenance Case Study*", Springer International Publishing, 2020, doi:10.1007/978-3-030-57997-5_8.

[3] Daniyan, I., Muvunzi, I., Mpofu, K., Oyesola, M., Ramatsetse, B. and Adeodu, A.," "*Artificial Intelligence System for Enhancing Product's Performance During its Life Cycle in a Railcar Industry*", Procedia CIRP, 2021.

[4] Kaiser, K., and Gebraeel, N., "*Predictive Maintenance Management Using Sensor-Based Degradation Models*", IEEE Trans Syst Man, Cyber Part Systems Humans 2009;39:840–9. doi:10.1109/TSMCA.2009.2016429, 2009.

[5] Coleman, C., Damofaran, S. and Deuel E., "*Predictive Maintenance and the Smart Factory*", Deloitte University Press, 2017.

[6] Zhang, W., Peng, G., Li, C., Chen, Y. and Zhang, Z., "*A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals*", Sensors, 2017

[7] Yan, L., Yoshua, B, and George, H., "*Deep Learning*", Nature, vol. 521, no. 7553, 2015, p. 436.

[8] Haiqin, Q., Kejun, X. and Likun, R., "*Rolling Bearings Fault Diagnosis via 1d Convolution Networks*", IEEE 4th International Conference on Signal and Image Processing (ICSIP), pp. 617–621, 2019.

[9] Dean, L., Vincent, S., Rick C. and Charles Y., "*Convolutional Neural Net and Bearing Fault Analysis*", Semantic Scholars, vol. 1, 2016.

[10] Guo, X., Shen, C. and Chen, L., "*Deep Fault Recognizer: An Integrated Model to Denoise and Extract Features for Fault Diagnosis in Rotating Machinery*", Applied Sciences, vol. 7, no. 1, 2017.

[11] Dhira,j N. and Jongwon, S., "*Bearing Fault Detection and Diagnosis Using Case Western Reserve University Dataset with Deep Learning Approaches*", EEE Access. 8. 9315-93178, 2020.

[12] Wade, A.S. and Robert, B.R., "*Rolling Element Bearing Diagnostics Using*

the Case Western Reserve University Data. a Benchmark Study", Mechanical systems and signal processing. Vol. 72,2016, pp. 303-315.

[13] Mohammad, M. and Mohammad, S.I., "*Rolling Element Bearing Fault Detection and Classification Technique Using Vibration Signals*", Engineering Proceedings,vol. 27, no. 1, 2022. https://doi.org/10.3390/ecsa-9-13339, 2022.

[14] Junjie, D., Gege L. and Caidan, Z., "*A New Bearing-Fault Classification Method Based on Deep Learning*", 15th International Conference on Computer Science & Education (ICCSE), Delft, Netherlands, 2020, pp. 191-196. doi:10.1109/ICCSE49874.2020.9201672.

[15] Ahmad S., Xu-Cheng, Y. and Waheed, Y.R., "*Robust Feature Extraction on Vibration Data Under Deep-Learning Framework: An Application for Fault Identification in Rotary Machines*", International Journal of Computer Applications, vol. 167, no. 4, 2017, pp. 37-45.

[16] Yin, Z. and Hou, J., "*Recent Advances on Svm Based Fault Diagnosis and Process Monitoring in Complicated Industrial Processes*", Neurocomputing, vol. 174, 2016,pp. 643–650.

[17] Guo, X., Chen, L. and Shen, C., "*Hierarchical Adaptive Deep Convolution Neural Network and its Application to Bearing Fault Diagnosis*", Measurement, vol. 93, 2016, pp. 490–502.

[18] Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X. and Chen, X., "*A Sparse Auto-Encoder-Based Deep Neural Network Approach for Induction Motor Faults Classification*", Measurement, vol. 89,2016, pp. 171–178.

[19] Sawaqed, S. and Alrayes, M., "*Bearing Fault Diagnostic Using Machine Learning Algorithms*", Progress in Artificial Intelligence, vol. 9, 2020, pp. 341–350.