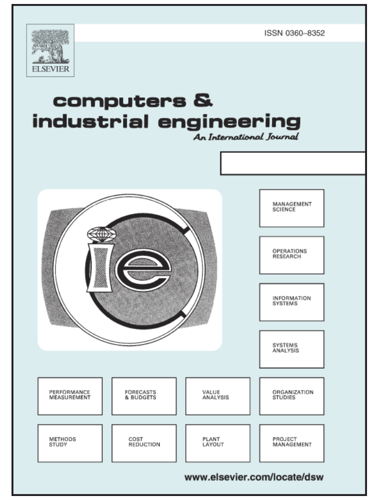
Journal Pre-proofs



Deep learning for diagnosis and classification of faults in industrial rotating machinery

Roberto M. Souza, Erick G.S. Nascimento, Ubatan A. Miranda, Wenisten J.D. Silva, Herman A. Lepikson

PII:	S0360-8352(20)30730-0
DOI:	https://doi.org/10.1016/j.cie.2020.107060
Reference:	CAIE 107060

To appear in: *Computers & Industrial Engineering*

Received Date:	15 September 2019
Revised Date:	4 November 2020
Accepted Date:	15 December 2020

Please cite this article as: Souza, R.M., Nascimento, E.G.S., Miranda, U.A., Silva, W.J.D., Lepikson, H.A., Deep learning for diagnosis and classification of faults in industrial rotating machinery, *Computers & Industrial Engineering* (2020), doi: https://doi.org/10.1016/j.cie.2020.107060

Deep learning for diagnosis and classification of faults in industrial rotating machinery

Roberto M. Souza[1,*], Erick G. S. Nascimento* (Corresponding author), Ubatan A. Miranda[2,*], Wenisten J. D. Silva[4,*], Herman A. Lepikson[4,*]

* Post Graduate Program on Computational Modeling and Industrial Technology, SENAI CIMATEC University Center, 1845, Orlando Gomes Avenue, - Piatã, 41650-010, Bahia – Brazil

* - erick.sperandio@fieb.org.br

1 - rmsouza_2@yahoo.com.br

2 - ubatan.miranda@fieb.org.br

3 - wenisten.silva@fieb.org.br

4 - herman.lepikson@fieb.org.br

***Abstract*—**Application of deep-learning techniques has been increasing, which redefines state-of-the-art technology, especially in industrial applications such as fault diagnosis and classification. Therefore, implementing a system that can automatically detect faults at an early stage and recommend stopping of a machine to avoid unsafe conditions in the process and environment has become possible. This paper proposes the use of Predictive Maintenance model with Convolutional Neural Network (PdM-CNN), to classify automatically rotating equipment faults and advise when maintenance actions should be taken. This work uses data from only one vibration sensor installed on the motor-drive end bearing, which is the most common layout present in the industry. This work was developed under controlled ambient varying rotational speeds, load levels and severities, in order to verify whether it is possible to build a model capable of classifying such faults in rotating machinery using only one set of vibration sensors. The results showed that the accuracies of the PdM-CNN model were of 99.58% and 97.3%, when applied to two different publicly available databases. This demonstrates the model's ability to accurately detect and classify faults in industrial rotating machinery. With this model companies can improve the financial performance of their rotating machine monitoring through reducing sensor acquisition costs for fault identification and classification problems, easing their way towards the digital transformation required for the fourth industrial revolution.

***Keywords*—** Mechanical Vibration, Deep Learning, Convolution Neural Network, Fault Classification, Predictive Maintenance

Acknowledgment

The authors thank the Bahia State Research Support Foundation (Fundação de Amparo à Pesquisa do Estado da Bahia – FAPESB) for partially funding the work, the SENAI CIMATEC Reference Center on Artificial Intelligence and the SENAI CIMATEC Supercomputing Center for Industrial Innovation for the computational infrastructure.

Deep learning for diagnosis and classification of faults in industrial rotating machinery

Abstract—Application of deep-learning techniques has been increasing, which redefines state-of-the-art technology, especially in industrial applications such as fault diagnosis and classification. Therefore, implementing a system that can automatically detect faults at an early stage and recommend stopping of a machine to avoid unsafe conditions in the process and environment has become possible. This paper proposes the use of Predictive Maintenance model with Convolutional Neural Network (PdM-CNN), to classify automatically rotating equipment faults and advise when maintenance actions should be taken. This work uses data from only one vibration sensor installed on the motor-drive end bearing, which is the most common layout present in the industry. This work was developed under controlled ambient varying rotational speeds, load levels and severities, in order to verify whether it is possible to build a model capable of classifying such faults in rotating machinery using only one set of vibration sensors. The results showed that the accuracies of the PdM-CNN model were of 99.58% and 97.3%, when applied to two different publicly available databases. This demonstrates the model's ability to accurately detect and classify faults in industrial rotating machinery. With this model companies can improve the financial performance of their rotating machine monitoring through reducing sensor acquisition costs for fault identification and classification problems, easing their way towards the digital transformation required for the fourth industrial revolution.

Keywords— Mechanical Vibration, Deep Learning, Convolution Neural Network, Fault Classification, Predictive Maintenance

1. Introduction

The use of non-invasive methods such as vibration analysis of rotating machines is an important tool for machine monitoring and for predicting the maintenance requirement of machines at an early stage. Vibration monitoring is one of the most commonly used methods to identify a machine state. Through its use, detecting an incipient failure, identifying the failure location, and estimating the failure time becomes possible. Monitoring the state of smart machines is critical to reduce or avoid process shutdown induced by machine faults, and to avoid environmental risks.

The detection of a vibration signal combined with an intelligent system can identify the machine condition and predict the possibility of failure. Therefore, developing a system that can help the maintenance team predict and repair rotating machines in a timely manner is possible. Unplanned machine maintenance can require materials, a long downtime process, and money. Thus, predicting and detecting machine failure in a timely manner are crucial to avoid sudden breakdown or failure of an entire system.

Continuous evolution of the wellness of machinery and monitoring of their performance have been made possible because of the application of new technologies. Because of the rapid drop in the cost of sensors and data storage and the increase in computer power, smart functionality for detecting machine failure at an early stage has become possible (Lu et al., 2017).

With developments in data-driven techniques, big data, data mining, computational intelligence, and machine-learning-based fault-diagnosis models such as support vector machine (SVM) and artificial neural network (ANN), some intelligent solutions have been applied to fault detection, prediction, and classification.

Data-driven techniques consist of the use of nonparametric methods to extract useful information from the data (Baptista et al., 2018). These techniques are very useful, especially for complex and highly efficient systems in which the development of a mathematical model may not always be feasible (Ahmad et al., 2019).

Machine-learning techniques are being applied for the classification and diagnosis of induction-motor failure, and have gained popularity in recent years. However, these techniques face a major challenge in terms of the difficulty in selecting the statistical features, which significantly limits the performance of classifiers. On the other hand, deep learning uses a method based on the representation of features, and opens up a new horizon where feature descriptors are extracted from raw signals (Chattopadhyay et al., 2018).

Currently, deep learning that uses data-driven approaches is gaining popularity owing to its potential to achieve physical model-free solutions (Chattopadhyay et al., 2018). Deep learning is increasingly used in many different applications, such as computer vision, sound, and face recognition. The use of raw signals is a new method in the field of data-driven fault diagnosis of induction motors. Deep learning does not require extensive human labour and knowledge of manual feature designs (Zhao et al., 2019). Deep learning can work with raw data, avoiding the necessary conversion of raw data to vector and statistical features required in other model types.

In the present study, deep learning is applied to identify and classify the severity of faults in a rotating machine for different rotational speed, load level and severities, aiming to assess and identify whether the proposed CNN method, adjusted to use the accelerometers installed on the drive end bearing can provide equivalent or even greater accuracy than previous experiments using the same database that used the both drive end and fan end accelerometers sensors. To achieve this goal, the PdM-CNN model was developed and applied, in which different adjustments were made to the convolution and pooling layers aiming to fine-tune the hyperparameters. In order to compare the performance of the proposed model with previous study, the accuracy of the PdM-CNN model was verified using the real-world database [dataset] (MaFaulDa) and Case Western Reserve University (CWRU), which will be later presented and explained.

This paper is organised into six sections. Section 2 describes the literature review. Section 3 presents the deep learning and CNN concepts, and the proposed method for fault detection in induction motors along with the CNN architecture used in this study. Section 4 the proposed method is validated using two real-world database provided by [dataset] MaFaulDa and CWRU database. The results and discussions are provided in Section 5, and the conclusions and future works are presented in Section 6.

2. Literature Review

Predictive Maintenance (PdM) is an emerging technology and the extensive literature reviews on the engineer fields using intelligent systems has increasing. Currently, with the development of deep learning in various research fields, some researchers have attempted to use the Convolutional Neural Network (CNN) to classify machinery faults.

In (Zhang et al., 2017), a method called Deep-CNN with First-Layer Kernels was used to treat the input data, which were raw vibration signals, and used wide kernels in the first convolutional layer to characterise the input signal and suppress high-frequency noise.

In (Zhang et al., 2018), the authors developed a method named Neural Networks with Training Interference (TICNN) in order to use it in an environment that has signal noise, as occurs in industrial environments. Noise was added to the signal using the White Gaussian function. To solve the problem, it used six interpolated convolution and pooling layers, one fully connected and one softmax. With the objective of filtering the noise, as explained in the paper, it configured the kernel with the dimensions of 64x1 in the first layer, and in the other layers, it configured the kernel as 3x1. The evaluation of the method was performed using the [dataset] CWRU database and presented an accuracy of 96.1%.

The work in (Zhao et al., 2019) provides a wide overview on the latest deep-learning-based techniques on machine health-monitoring systems and their effect on state-of-the-art technologies. It compared the performance between different deep-learning models (Random Forest, NN, Autoencoder, Denoising Autoencoder, and CNN).

Reference (Janssens et al. 2016) presented a study that consisted of the development of a solution that, without any human intervention, can learn the characteristics of raw data representing machine failure. During the experiment, several bearing-failure types were analyzed, such as failure in external raceways, lubrication degradation, healthy-bearing failure, and rotor imbalance. To verify the generalization of the model, several tests were performed on different types of bearings and failure. In the study conclusions, two systems were compared using the same data. The precision obtained by the system developed by the author, which used feature-learning and CNN, reached precision values of 93.61% and 87.25%, which were much better than those obtained by another system that used feature learning and Random Forest classifier, respectively.

In Reference (Guo et al. 2016) proposed the use of CNN to diagnose bearing faults and determine their severity. The study achieved satisfactory performance in terms of fault-pattern recognition and fault-size evolution.

In reference (Jia et al. 2018), the Deep Normalized CNN (DN-CNN) model was presented a good accuracy even when using a small database in the model training phase. This study was conducted to demystify that the CNN method does not have good accuracy when working with a small number of data. This study expresses the practical situations that, as explained in the work, there are few database that represent equipment on failure conditions in detriment to the largest amount of database that represents the equipment functioning without. The study compared the use of the ReLU (Rectified Linear) and Sigmoid activation functions and concluded that the ReLU function responded best in enabling convergence at the pooling layer. The developed method presented an average accuracy of 97.64% using three bearing databases.

In (Khan and Yairi, 2018), a systematic review of the works presented from 2013 to 2017 was performed. Through the review, the authors try to answer whether deep learning can be used to solve System Health Management (SHM) problems. The paper states that Recurrent Neural Networks (RNN) and CNN methods are complex structures and deep learning will have difficulty replacing SHM applications using SVM and Random Forest.

3. The Predictive Maintenance Model with Convolutional Neural Network (PdM-CNN)

This section will present the details of the development, building, training, validation and test tasks of the proposed PdM-CNN model, as well as the theory and background of the main mathematics, computational principles and techniques used in this study. This section will begin by explaining the general theory and its constituent parts of deep learning and CNN. Each of the modules that make up the developed model will be described in detail below. Afterwards, it will describe the methodology of the PdM-CNN model.

3.1 General Theory Description

Deep-learning models are a subset of the machine-learning algorithm class, which is based on the cascade of multiple layers used in Artificial Neural Network (ANN) for learning in a supervised and/or unsupervised manner. It is based on learning the data representations, which enable the data to be fed into the system to naturally discover the representations required for regression or classification.

CNN is a special type of Feed Forward Multi Layer Neural Network with deep-learning capability. It explores the strategy of hierarchical computation and local receptive fields by providing a number of cascade combinations of convolution, nonlinear activation, and downsampling layers between the input and fully connected layer (Chattopadhyay et al., 2018), as shown in Figure 1.

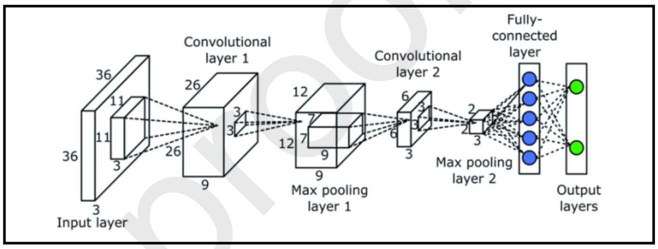


Fig. 1. Schematic representation of the operations performed in CNN. Ref: (Xu et al., 2018)

CNN is widely used in the field of computer vision, which features spatially shared weights and spatial subsampling (pooling) with two-dimensional (2D) or three-dimensional sequence. Further, it has also been used in speech, audio, text, and time-series applications with one-dimensional (1D) sequence. In contrast to images, which are spatially distributed in two dimensions, the sensor output signals used in the domain of electrical-engineering applications are more frequent and often found to be 1D time-series data.

In addition to its strong performance, CNN tends to be computationally efficient because it requires fewer parameters than other architecture. Further, it is easier to parallelise across graphics processing unit cores. In addition, it requires a smaller pre-processing time than the multi-layer method because of the sharing of weights by employing a much smaller convolution kernel, which acts on a local region of the input data and extracts meaningful feature descriptors. When a single-layer CNN is used, the characteristics of the input signals are extracted through a signal convolution operation using a filter (or kernel). Detecting patterns captured by the kernel is possible regardless of where the pattern occurs using a convolutional operation. In CNNs, kernels are optimised to find the best parameters and to analyse the result as part of the supervised training process. The map of the characteristics consists of an array of units (or layers) whose units share the same parameters (weight and bias vector), their activation generates the result of the kernel convolution using all the input data (Ordóñez and Roggen, 2016).

The name CNN indicates that it employs a mathematical operation called convolution that is a specialized type of linear operation. Therefore, CNN are neural networks that use convolution instead of general matrix multiplication in at least one of its layers. The operation used in CNN does not exactly match the definition of convolution used in other fields, such as pure engineering or mathematics (Goodfellow et al., 2016).

The convolutional function is defined as

$$s(t) = \int_{a=-\infty}^{\infty} x(a)w(t-a)da \quad (1)$$

where w is the weight (kernel), x is the input value, t is the time, and the function *s(t)* is referenced as feature map.

As the convolution function is cumulative and can be used on more than one cartesian axis at a time, as is the case with 2D image processing. The Equation (2) describes equivalence and best represents the convolution function in machine learning applications (Goodfellow et al., 2016).

$$S_{(i,j)} = (x * w)_{(i,j)} = \sum_m \sum_n x(i-m,j-n)w(m,n) \quad (2)$$

Convolutional layers consist of multiple local filters using raw input data and generate the local features of these data. The subsequent polling layers extract resources with a fixed length over sliding-data windows following various rules such as the mean, maximum, and so on (Ordóñez and Roggen, 2016).

Depending on the application, a number of combinations of convolution, nonlinear activation function, and subsampling layers are cascaded to obtain deeper network architecture. Feature descriptors extracted from the final pooling layers are used as input to the fully connected layer of CNN to obtain the probabilistic output using the softmax function (Chattopadhyay et al., 2018).

A typical layer of a convolutional network consists of phases. In the first phase, the layer performs the convolution process in parallel to produce a set of linear activations. In the second phase, each linear activation is performed using a non-linear activation function such as ReLU. In the third phase, a pooling function is used to modify the output to the next or last layer. On the last phase, the activation function softmax is performed on the fully connected layer that classify the data.

The ReLU activation function is available in the convolution layer and is responsible for finding a non-linear function that represents the input signal (x), normalizing the weight (w) and accelerating the model convergence process (Krizhevsky et al., (2012) and Zhang et al., (2017)).

The ReLU activation function has been more widely used than sigmoidal functions. The sigmoidal functions compress the output to a short interval and the ReLU function cancels all negative values, being linear to the positive ones, as shown in Figure 2.

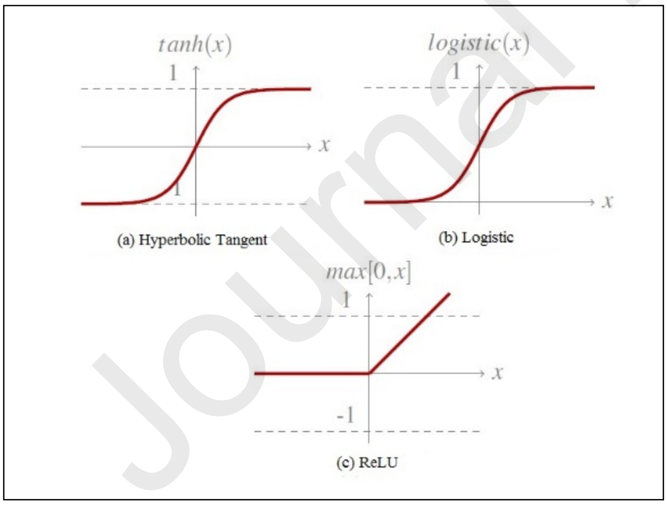


Fig. 2. ReLU Activation Function. Ref. (Vieira et al. 2017)

The ReLU activation function is represented as

$$a^{l(i,j)} = f\left(z^{l(i,j)}\right) = \max\{0, z^{l(i,j)}\} \quad (3)$$

where $a^{l(i,j)}$ is the activation function.

On the pooling layer, the most commonly function used is the maxpooling, with the aim of reducing the amount of CNN parameters, reducing the size of the data resulting from the convolution layer and accelerating the computing process. The sliding step size is configured in the function to scroll the entire length of the data volume (Jia et al., 2018).

The most common form of the pooling layer is to replace the values of a region by its function (MaxPooling, GlobalAvgPooling, GlobalMaxPooling). Figure 3 shows a reduction of the data input volume from 4 x 4 to 2 x 2, due to the use of the 2 slide step, using the MaxPooling function.

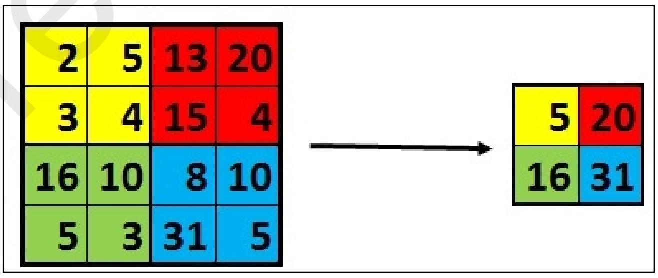


Fig. 3. MaxPooling Function

The fully connected layer is configured after the convolution and pooling layers, being the last layer of the network. Output data from the pooling layer is routed to this layer in order to sort the data and propagate the signal. These layers are similar to a conventional artificial neural network that uses softmax activation functions on the last output layer. Figure 4 shows the fully connected layer and softmax activation functions.

The softmax function is the activation function used in the fully connected layer so that the output can be interpreted as probabilities and allows assigning an example to the class with the highest probability (Vieira et al. 2017).

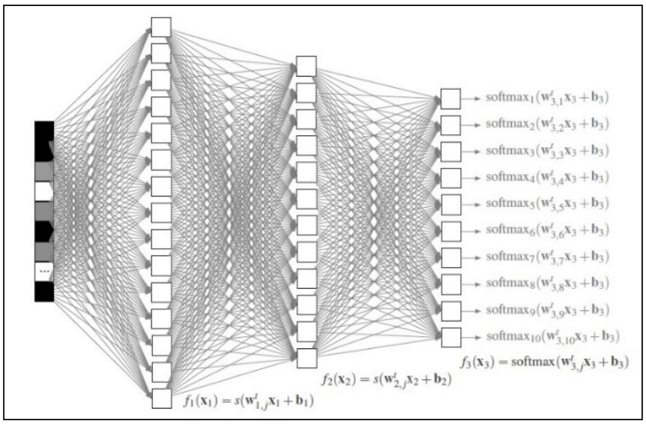


Fig. 4. Fully Connected layer and softmax function. Ref. (Vieira et al. 2017)

The softmax function calculates the probability distribution of the event over '*n*' different events, as Figure 5. The main advantage of using softmax is the output probability range of 0 to 1 and the sum of all probabilities equals one.

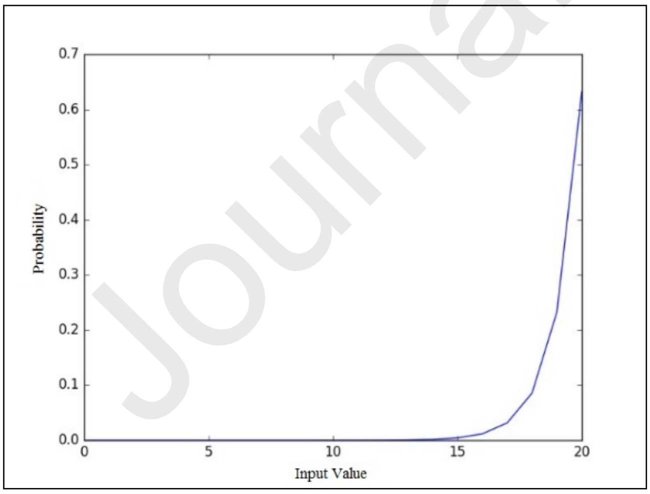


Fig. 5. Softmax function

Dropout function is implemented during the training stage. To avoid overfitting and to eliminate the co-dependence between neurons that can be developed during the learning process, it disables part of the neurons in a layer so that they do not contribute or learn any information during these updates, making the remaining active neurons to learn and reduce the error. The effect is that the network becomes less sensitive to the specific weights of those neurons. This, in turn, results in a net that is able to improve generalization and is less likely to overlap with training data. In Figure 6, it is illustrated, respectively, a neural network without the dropout function and with the dropout function, with a 0.5 adjustment, i.e., a 50% probability.

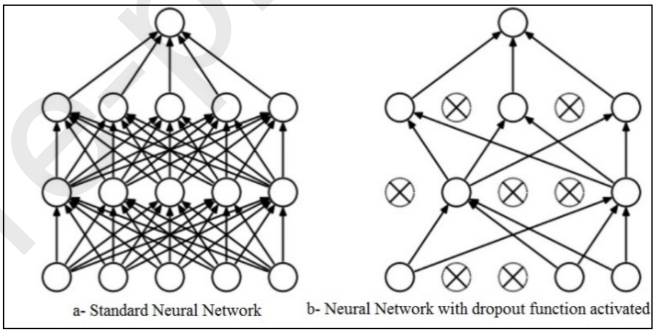


Fig. 6. The dropout function. Left and right neural networks illustrate, respectively, one without and the other with the dropout function activated. Ref. (Jerkic, 2017)

3.2 The Proposed PdM-CNN Model

This paper proposes a framework named Predictive Maintenance model with Convolutional Neural Network (PdM-CNN) for automatically classifying faults in rotating machinery with greater performance for unbalanced datasets. On the model, firstly, ReLU and weight normalization were used. Then, the softmax and loss function were applied to optimize the faults in rotating machinery. The proposed model was validated by applying and analysing its results for two publicly available and real world data datasets with different classes of faults and imbalance degrees.

To optimize the performance of the model, for each vibration sensor data, the corresponding Fast Fourier Transform (FFT) was extracted and used as the input vector for the training, validation and test tasks. Resource extraction offers a great advantage by drastically reducing the input vector size compared to the number of samples in the time series. Moreover, this step of the process allows the use of a vector with a higher discrimination capacity of the failure classes than the one obtained with the original time series.

Instead of using all the data from the vibration sensors, similar to that in previous works (Martins, 2019; Marins, 2018; Ribeiro, 2017), the proposed model was designed to use only the data from the signals captured by the three unidirectional accelerometers, which is equivalent to a triaxial accelerometer, installed at the drive end (position 4 in Figure 10b), because it is the monitoring configuration mostly used in industrial processes.

Figure 7 shows the FFT signals in some fault cases. In this example, the simulator motor (MFS) operated at a speed of 1800 rpm (30 Hz). Therefore, checking the peak of the motor-rotation frequency is possible.

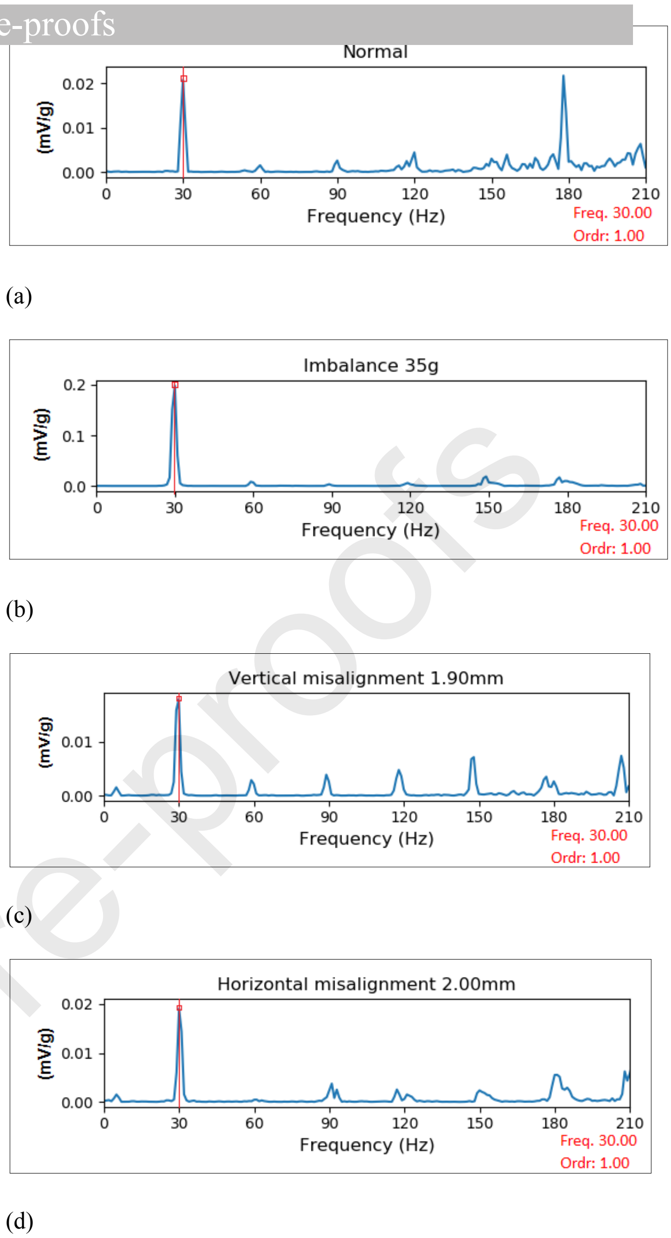


Fig. 7. FFT of signals for (a) No Faults. (b) Imbalance. (c) Vertical misalignment. (d) Horizontal misalignment

In this study, the PdM-CNN model was generated by a High-Performance Computer (HPC) using one node with two Xeon(R) Gold 6148 CPU @ 2.40 GHz processors with 40 cores and 187 GB of memory.

The Tensorflow framework with Keras was also used to develop an one dimensional CNN architecture composed of four convolution layers with ReLU activation function, two pooling layers with MaxPooling and GlobalAvgPolling activation function for each layer, a dropout function and a fully connected layer with softmax activation function, as shown in Figure 8 and 9. A dropout of 0.5 was selected as a regularisation parameter. For the convolution layers, the activation function ReLU, together with the dropout smoothing parameter, was used to avoid the problem of vanishing gradient. The convolution layers were constructed using 100 filters whose size was 10 each. For the training and validation tasks, 65% of the data were selected from the database, whereas 35% were assigned to test the PdM-CNN model.

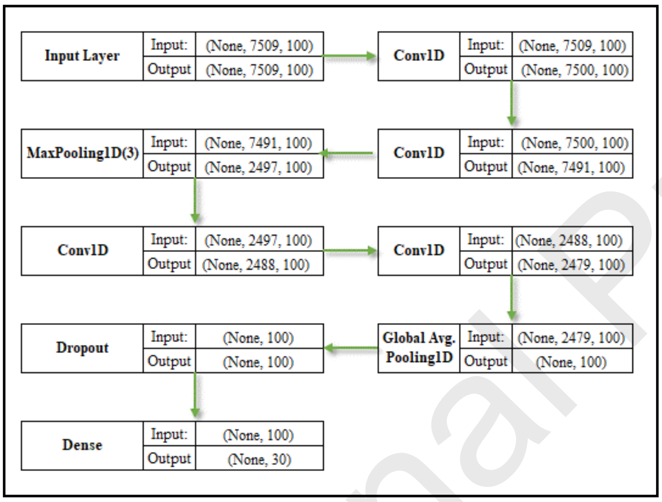


Fig. 8. The structure of the PdM-CNN model used in this work.

```
Layer (type)                 Output Shape              Param #
=================================================================
conv1d_5 (Conv1D)            (None, 7500, 100)         3100
_________________________________________________________________
conv1d_6 (Conv1D)            (None, 7491, 100)         100100
_________________________________________________________________
max_pooling1d_2 (MaxPooling1 (None, 2497, 100)         0
_________________________________________________________________
conv1d_7 (Conv1D)            (None, 2488, 100)         100100
_________________________________________________________________
conv1d_8 (Conv1D)            (None, 2479, 100)         100100
_________________________________________________________________
global_average_pooling1d_2 ( (None, 100)               0
_________________________________________________________________
dropout_2 (Dropout)          (None, 100)               0
_________________________________________________________________
dense_2 (Dense)              (None, 30)                3030
=================================================================
Total params: 306,430
Trainable params: 306,430
Non-trainable params: 0
_________________________________________________________________
```

Fig. 9. The PdM-CNN model response

Using this parameter configuration, the PdM-CNN model was generated at 1 h and 55 min.

4. Experimental Verification

To validate our model and compare its performances with previous studies we tested the PdM-CNN model on two real-world dataset. The [dataset] MaFaulDa, is a database that includes different types of failures, severity and rotation frequencies. The second database is the [dataset] CWRU, which is used to evaluate the model and test its adaptability.

4.1 MaFaulDa

The [dataset] MaFaulDa (Ribeiro, 2018) was used to develop the experiment in this study. The database includes multiple types of faults with different levels of severity and rotational speeds. It is composed of multivariate time-series signals acquired by sensors from a Machinery Fault Simulator (MFS) test rig that simulates the machine under normal operation and five fault conditions, namely, horizontal and vertical misalignment, mass imbalance, and underhang and overhang bearing faults, at different rotational speeds.

MFS, shown in Figs. 10a and 10b, is commonly used to study, simulate, and classify fault problems in rotating machines using the data-driven theory that avoids mathematical simulations. It is composed of a three-phase ¼-hp induction motor with a variable frequency drive that can provide rotational speeds in the range of 700 – 36000 rpm. In addition, a rotor-shaft system with a centre disc between the bearings and sensors is installed on the bearing housings to measure the vibration, speed, and sound signals.

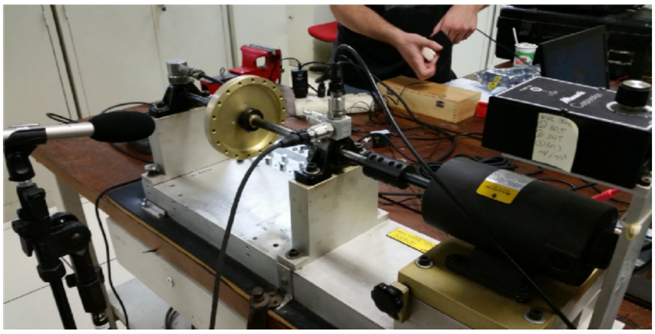


Fig. 10a. MFS used to generate MaFaulDa database Ref: (Ribeiro, 2018)

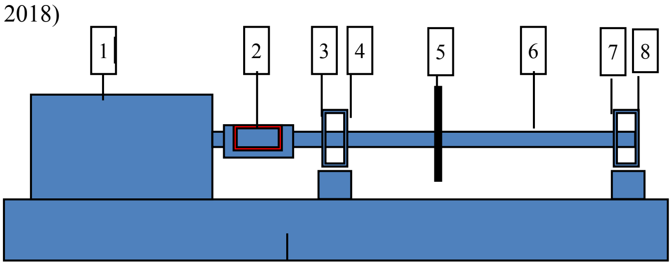


Fig. 10b. MFS schematic drawing showing each sensor position: 1. Servo-motor; 2. Coupling; 3, 7. Bearing house; 4, 8. Accelerometers; 5. Disk; 6. Shaft;

The data were acquired using two sets of accelerometers (sensitivity 102 mV/g; one set for each bearing), which were positioned at the longitudinal, horizontal, and vertical directions, a tachometer (to measure the rotor speed), and a microphone (to measure the noise). The acquired vibration signals were sampled at 51.2 kHz using the National Instruments NI 9234 data-acquisition system

A variety of faults were generated by the MFS (Ribeiro, 2018), which are described as follows.

Normal operation: This class corresponded to the generation of a database using the data from the performed tests without applying any fault. These data were generated through a set of 49 different test scenarios where the rotation speed ranged from 737 to 3686 rpm. However, during the tests, the speed was fixed, and steps of approximately 60 rpm were set.

Imbalance: This class corresponded to the generation of a database using the data from the tests carried out with the introduction of different loads installed in the rotor, such as 6, 10, 15, 20, 25, 30, and 35 g. The database included 333 different scenarios. No data was obtained when a load equal to or greater than 30 g was applied and the speed was more than 3300 rpm because the testing platform vibrated excessively.

Horizontal parallel misalignment: This class corresponded to the generation of a database using the data from the tests performed by moving the motor shaft horizontally by 0.5, 1.0, 1.5, and 2.0 mm. The same rotations were used during the tests at normal operation. The database in this class included 197 different scenarios.

Vertical parallel misalignment: This class corresponded to the generation of a database using the data from the tests performed by moving the motor shaft vertically by 0.51, 0.63, 1.27, 1.4, 1.78, and 1.9 mm. The database in this class included 301 different scenarios. The same rotations were used during the tests at normal operation for each vertical step.

Bearing faults: This class corresponded to the generation of a database using the test data from the tests performed using three defective bearings. Each bearing had different defects, namely, outer race, ball element, and cage defects. The experiments were performed using one defective bearing each time at two different support positions: at drive end (position 3 in Figure 10b) or at fan end (position 7 in Figure 10b). Because of the damping effect caused by the mass of the system, some bearing defects were practically imperceptible when no imbalance existed. Therefore, to induce a noticeable effect, three masses of 6, 10, and 20 g were added. The same rotations were used during the tests at normal operation for each test performed. The database included a total of 558 underhang and 513 overhang scenarios.

The aforementioned operating conditions generated a database with 1951 different scenarios. Each of these scenarios was captured by eight sensors at a frequency rate of 50 kHz with a data-acquisition time of 5 s. This dataset is available to the public from (Ribeiro, 2018).

4.2 CWRU

The data provided by [dataset] CWRU were collected from the bearing testing platform shown in Figure 11. The platform comprises a 2 HP motor, a dynamometer, a torque transducer and encoder. For the tests, faults ranging in diameter from 0.007 to 0.028 in. were seeded on the drive and fan-end bearings. There are four fault types of the bearing: normal, ball fault, inner race fault and out race fault, and each faulty bearing was reinstalled (separately) on the test rig, which was then run at constant speed for motor loads ranging from 0 to 3 horsepower (motor speeds of 1720 to 1797 rpm).

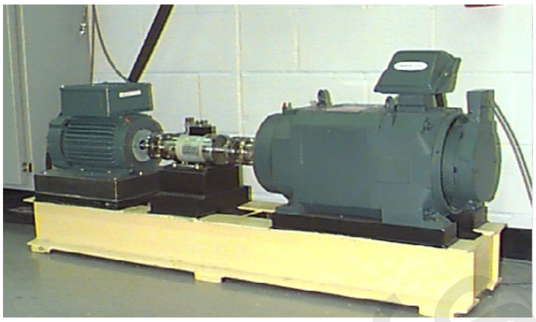


Fig. 11. Bearing testing platform (CWRU, 2020)

The data sets are grouped into four classes – 48K baseline, 48K drive end fault, 12K drive end fault, and 12K fan end fault, as to the sample rate, motor speed and fault location, making a total of 161 data sets. The 48K baseline is normal condition, without fault. During the tests, the original experiment data was obtained from one unidirectional accelerometer installed on the vertical position.

The [dataset] CWRU bearing database was selected because it is in the public domain as well as the [dataset] MaFaulDA database, and because it is widely used in scientific literature. The difficulty encountered in using the [dataset] CWRU database was the difference in the database structure, as detailed in Table 3.

In this work only scenarios containing the 48K drive-end signals were used due to their similar characteristics to the [dataset] MaFaulDa database sampling rate, allowing proper comparisons between each dataset.

5. Results and Discussions

Two failure classification methods were used in the experiments. One method used all failure classes with different severity levels as a single class, and the other method grouped all individual classes into the same base-failure class. The latter grouping made more sense once the main objective was to classify the failure according to its type (e.g. misalignment, imbalance, bearing faults, etc.) and not by its severity level. The overall accuracy of the PdM-CNN model for each classification method was 97.25% and 99.58%, respectively, when they were evaluated using the test dataset. The confusion matrix of the second approach is listed in Table 1.

In the following, comments regarding the first experiment using all classes and classifying the failure according to the severity levels are presented.

- The 'Cage Fault' rating showed accuracy of 98.3%, which reflected a practical problem when the vibration analyst must provide technical advice and the error was not always simply a problem in the bearing cage. Usually an imbalance and/or misalignment problem occurs. Then, the bearing is usually replaced, and a new imbalance/misalignment test is performed to ensure that the problem has been resolved. Bearing problems (ball, cage, or outer race fault) do not usually individually occur in practice, and bearing failure may occur together with imbalance and/or misalignment failure. Therefore, they usually appear together in the spectrum (FFT).
- With regards to the classification of the severity of the vertical misalignment failure, the accuracy was found to be 90.2%. The system was able to prove that the misalignment of more than 0.63 mm for the configuration of this machine was very high. Therefore, in this case, the more the misalignment increased, the larger was the error in the classification of the severity of failure because of the possibility of other overlapping faults.
- The system demonstrated accuracy of higher than 99% in several severity classifications such as no fault, horizontal misalignment, ball fault, and outer race fault.

Comments regarding the experiment that aimed to classify the types of failure, group the severity levels in the same group as one failure class, and present accuracy of 99.58% are also presented.

- The system demonstrated excellent accuracy (99.58%), which was higher than that in previous studies.
- The classification of no faults, imbalance, and ball fault showed 100% accuracy, as listed in Table 1.

We should note that in addition to using the CNN-type method, which is still rarely used for failure classification, only the three accelerometers installed on the motor drive end bearing were used during the experiment because this is a more common layout of machine monitoring applied in the industry. Nevertheless, obtaining accuracy of 99.58% is possible.

Table 1 – Confusion matrix - MaFaulDa

	No fault	Horizontal Misalignment	Vertical Misalignment	Imbalanced	Ball Fault	Ball Fault Imbalanced	Cage Fault	Cage Fault Imbalanced	Outer Race	Outer Race Imbalanced
No Fault	174	0	0	0	0	0	0	0	0	0
Horizontal Misalignment	0	681	1	0	0	0	1	0	0	0
Vertical Misalignment	1	1	1039	0	0	0	0	0	0	0
Imbalanced	0	0	0	1170	0	0	0	0	0	0
Ball Fault	0	0	0	0	179	0	0	0	1	0
Ball Fault Imbalanced	0	0	0	0	0	467	0	0	0	0
Cage Fault	0	0	0	0	0	0	183	0	0	0
Cage Fault Imbalanced	0	0	0	1	0	0	0	496	0	7
Outer Race	0	0	0	0	1	0	0	0	165	0
Outer Race Imbalanced	0	0	0	0	0	1	0	6	0	458

The metrics of the experiment are listed in Table 2 in which analysing the accuracy, precision, recall, and f1-score results for each fault type is possible. The average accuracy of the experiment was 99.58%. This result evidently demonstrated that the use of deep learning with CNN, which aims to distinctly and accurately classify the defects of the motor with little human intervention, is possible compared with the other model types or approaches.

The accuracy obtained in the present study was higher than the best result presented so far in (Marins et al., 2018) using the same [dataset] MaFaulDa. On the previous work showed by Marins performed plenty of works to pre-process the signals has been shown to be unnecessary when compared with the present work. Instead, on this work the application of the CNN model was proven to be capable of outperforming the three steps presented in (Marins et al., 2018) and demonstrated higher accuracy. Additionally, the use of data from only one vibration sensor has been proven able of achieving optimal accuracy.

The loss was calculated using the objective function of the sparse and categorical cross-sectional entropy. Although the number of epochs to train and validate the network was set to 300, the convergence of the model loss was verified for the training and validation processes, which achieved a steady state in 135 epochs, as shown in Figure 12.

Owing to the use of ReLU, the problem of vanishing gradient did not occur in the training and validation processes. Owing to the use of the dropout, the losses quickly converged, with no considerable over or underfitting.

Table 2 – MaFaulDa metrics

Average Accuracy (%): 99.58				
Fault Class	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
No fault	99.98	99.42	100.00	99.71

Horizontal Misalignment	99.94	99.85	99.70	99.78
Vertical Misalignment	99.94	99.90	99.80	99.85
Imbalanced	99.98	99.91	100.00	99.95
Ball Fault	99.96	99.44	99.44	99.44
Ball Fault Imbalanced	99.98	99.78	100.00	99.89
Cage Fault	99.98	99.45	100.00	99.72
Cage Fault Imbalanced	99.72	98.80	98.41	98.60
Outer Race	99.96	99.39	99.39	99.39
Outer Race Imbalanced	99.72	98.49	98.49	98.49

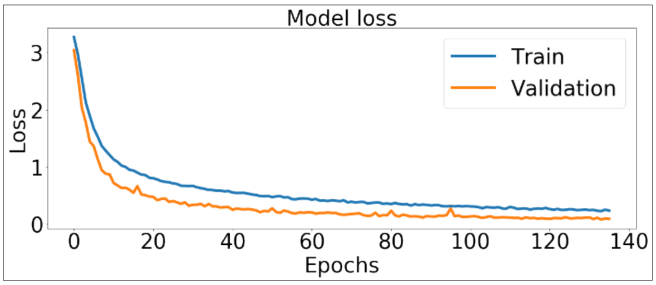


Fig. 12. The evolution of the loss during the training and validation phases.

Regarding the [dataset] CWRU database, the tests were performed using the data grouped into 48K drive end fault, which corresponds to the 48 KHz sample rate, and one unidirectional accelerometer installed at the drive end bearing position. This database has a very different structure from the [dataset] MaFaulDa database, as can be seen in Table 3.

It is noticeable that the [dataset] MaFaulDa database is a more complete dataset, which has much more fault classes and cases than the [dataset] CWRU database. Even though the structure of the [dataset] MaFaulDa database is different from the [dataset] CWRU database, as explained in the Table 3, the same model topology was used to generate and test the model using the [dataset] CWRU database. As shown in the Table 4, the overall accuracy of the model was 97.3%, demonstrating its ability to be adapted and used to new data. The confusion matrix of [dataset] CWRU test is listed in Table 5.

Although the number of epochs to train and validate the network was set to 300, the convergence of the model loss was verified for the training and validation processes, which achieved a steady state in 145 epochs, as shown in Figure 13.

Table 3 – Database comparison

	Database	
	MaFaulDa	CWRU
Faults	Normal	
	Imbalanced	
	Horizontal Misalignment	
	Vertical Misalignment	
	Outer Race	Outer Race
	Ball	Ball
	Cage	
		Inner Race
Sample Rate	50 KHz	48 KHz
Accelerometer	3 X Unidirecional	1 X Unidirecional
Direction	Axial, radial, tangential	Axial

Table 4 – CWRU metrics

Average Accuracy (%): 97.3			
Fault Class	Precision (%)	Recall (%)	F1-Score (%)
Ball	100.0	100.0	100.0
Inner Race	100.0	85.7	92.3
Outer Race	95.5	100.0	97.7

Table 5 – Confusion Matrix – CWRU

	Prediction

	Faults	Ball	Inner Race	Outer Race
Real	Ball	18	0	0
	Inner Race	0	12	2
	Outer Race	0	0	42

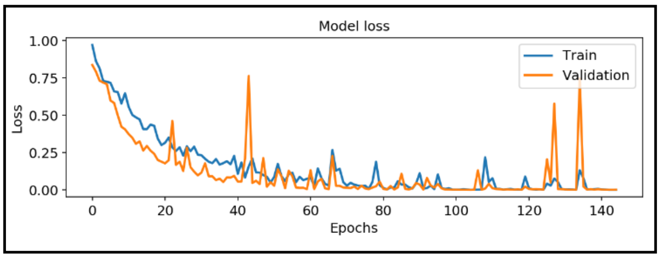


Fig. 13. The evolution of the loss during the training and validation phases for the CWRU database.

In comparison with results presented through previous published works using the same database (MaFaulDa), it was found that it reached the best average accuracy of 99.58%. Therefore, the PdM-CNN model reached an excellent level, when analyzing the metrics, and higher than previously published works, as compared in Table 6.

Table 6 – Comparison of previous studies and this work

Reference	**Model**	***Average Accuracy* (%)**
MARTINS, HEMERLY, et al., 2018	SBM + Random Forest	94.14
VIANA, LÓPES, et al., 2016	MLP	95.80
RIBEIRO, MARINS, *et al.*, 2017	SBM + RFC	96.43
MARINS, RIBEIRO, et al., 2018	SBM	98.50
ROCHA, 2018	XGBoost + Haar Wavelet	98.70
This work	PdM-CNN	99.58

The previous work with the best reported accuracy is 98.70%, presented by Rocha (2018) using the XGBoost classifier and the Haar-wavelet transform. To model the experiment, the author used the database [dataset] (MaFaulDa) and classified the misalignment (vertical and horizontal), imbalance and rolling failures (cage, ball and outer race), presenting the lowest accuracy for the vertical misalignment fault among all failures, demonstrating how difficulty is to correct identify and classify this failure.

The second previous work with the best reported accuracy of 98.50% using [dataset] MaFaulDa was presented by Marins et al. (2018) using the similarity-based modelling (SBM) and random forest models. The methodology in that work consisted of three steps: pre-processing of the data, application of SBM, and finally applying the random forest model. The pre-processing step converted the raw data (time-based values) into spatial and statistical features (FFT, kurtosis, mean, and standard deviation) of each signal. In the second step, the converted features were forwarded to the SBM model that compared the similarity between the current value and that of the existing model. In the third step, the data were forwarded to the random forest model, which was responsible for classifying the failure. The processed data resulted in 46 features, namely, rotation frequency (fr) obtained from the tachometer, 21 spectral features (fr, 2fr, and 3 fr) for each signal, and 24 other statistical features (mean, kurtosis, and entropy for each of the eight signals). That study presented the various experiments and parameter adjustments that were performed until the accuracy of 98.5% was achieved.

6. Conclusion and future work

The proposed PdM-CNN model, built with deep learning with 1D-CNN, has been shown to be capable of outperforming the state-of-the art methods in this field, which demonstrated accuracy of 99.58% in the classification of failures using [dataset] MaFaulDa database (Ribeiro, 2018) using only one set of vibration sensor. From this study, the use of CNN method for classification of rotating-machinery-failure types was demonstrated to be possible and feasible. Classification of the failure classes using all the severity levels presented an accuracy of 97.25%, which is a high accuracy level, whilst, if considering the fault classes independently of the severity levels, an accuracy of 99.58% was reached, which is a very high level. The same topology of the model was applied in the [dataset] CWRU database, which obtained an accuracy of 97.3%, demonstrating the adaptability and applicability of the PdM-CNN model topology to other similar cases, presenting better accuracy than the work presented by (Zhang et al., 2018), whose obtained accuracy was of 96.1%. Thus, this method showed a great potential for application in the diagnosis and classification of failures of rotating machines in industrial environments with different severity levels, even with the usage of only one vibration sensor and features, i.e. only the FFT of the data from the vibration sensor. Therefore, compared with similar studies that used the same database, we have shown the high capability of the CNN method even with the use of fewer data and features, highlighting that the industry can optimize financial costs of machine monitoring by reducing sensors acquisition costs for fault identification and classification problems, easing the way towards the digital transformation required for the fourth industrial revolution.

An important research direction can be the application of model on the industrial application for the development of real-time faults classification on start condition. Another potential application is to implement a thermal sensor to detect lubrication degradation, and join the PdM-CNN model with the maintenance scheduling.

References

Ahmad, W., AliKhan, S., Islam, M. M., & Kim, J. M. (2019). A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models. Reliability Engineering and System Safety, 184, (pp. 67-76).

Baptista, M., Sankararaman, S., Medeiros, I. P.de, Junior, C. N., Prendinger, H., & Henriques, E. M.P. (2018). Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. Computer & Industrial Engineering, 115, (pp. 41-53).

Chattopadhyay, P., Saha, N., Delpha, C., & Sil, J. (2018). Deep Learning in Fault Diagnosis of Induction Motor Drives. Prognostics and System Health Management Conference (PHM-Chongqing) (pp.1068-1073).

CWRU - Case Western Reserve University Bearing Data Center. ⟨http://csegroups.case.edu/bearingdatacenter/home⟩. Last access 08/03/2020.

Goodfellow, I., Bengio, Y., Courville A. (2016). Deep Learning. MIT Press

Guo, X., Chen, L., & Shen, C. (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. Measurement, 93, (pp. 490-502).

Janssens, O., Slavkovik, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., Walle, R. V., & Hoecke, S. V. (2016). Convolutional Neural Network based fault detection for rotating machinery. Journal of Sound and Vibration, 377, (pp. 331-345).

Jerkic, M. (2017). What is dropout?. Deep learning. <https://markojerkic.com/what-is-dropout-deep-learning/>. Last access in 29/08/2019

Jia, F., Lei, Y., Lu, N.,& Xing, S. (2018). Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. Mechanical Systems and Signal Processing, 110, (pp. 349–367).

Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. Mechanical Systems and Signal Processing, 107, (pp. 241–265).

Krizhevsky, A., Sutskever, I., Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Neural Information Processing Systems, (pp. 1097-1105).

Liu, R., Yang, B., Zio, E., & Chen, X. (2018). Artificial intelligence for fault diagnosis of rotating machinery: A review. Mechanical Systems and Signal Processing, 108, (pp. 33-47)

Lu, P., Liu, H., Serratella, C., & Wang, X. (2017). Assesment of Data-Driven, Machine Learning Techiniques for Machinery Prognostics of offshores Assets. Offshore Technology Conference.

Marins, M. A. , Ribeiro, F. M. L. , Netto, S. L. , & Silva, E. A. B. da. (2018). Improved similarity-based modeling for the classification of rotating-machine failures. Journal of the Franklin Institute, 355, (pp. 1913-1930)

Martins, D. H. C. S. S., Hemerly, D. O., Marins, M., Lima, A. A., Silva, F. L., Prego, T. M., Ribeiro, F.

M. L., Netto, S. L., & Silva, E. A. B.. (2019) Application of machine learning to evaluate unbalance severity in rotating machines. Proceedings of the 10th International Conference on Rotor Dynamics – IFToMM, 61, (pp. 144-160).

Ordóñez, F. J., & Roggen, D. (2016). Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. Sensors.

Ribeiro, F. M. L. (2018). Machinery Fault Database (MaFaulDa) - multivariate time-series acquired by sensors on a SpectraQuest's Machinery Fault Simulator (MFS) Alignment-Balance-Vibration - ABVT. <http://www02.smt.ufrj.br/~offshore/mfs/page_01.html>. Last access 08/03/2020.

Ribeiro, F. M. L., Marins, M. A., Netto, S. L., & da Silva, E. A. B. (2017). Rotating machinery fault diagnosis using similarity-based models. XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais - sbrt2017.

Rocha, D. S. C. (2018) Aprendizado de máquina aplicado ao reconhecimento automático de falhas em máquinas rotativas. Universidade Federal de Minas Gerais.

Viana, D. P., López, R. Z., Lima A. A., Prego, T. M., Netto, S. L. Silva, E. A. B. (2016). The Influence of Feature Vector on the Classification of Mechanical Faults using Neural Networks. VII Latin American Symposium on Circuits and Systems (LASCAS), (pp. 115-118).

Vieira, V., Razente, H. L., Barioni, M. C. N. (2017). Tópicos em gerenciamento de dados e informações. Sociedade Brasileira de Computação (SBC).

Xu, R., Li, C., Paterson, A. H., Jiang, Y., Sun, S., & Robertson, J. S. (2018). Aerial Images and Convolutional Neural Network for Cotton Bloom Detection. Front. Plant Sci.

Zhang, W., Peng, G., Li, C., Chen,Y. & Zhang, Z. (2017). A new Deep Learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors.

Zhang, W., Li, C., Peng, G., Chen, Y., & Zhang, Z. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. Mechanical Systems and Signal Processing, 100, (pp. 439–453).

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. Mechanical Systems and Signal Processing, 115, (pp. 213-237).

Roberto M. Souza – Investigation, Writing - Original Draft, Visualization

Erick G. S. Nascimento – Conceptualization, Methodology, Formal analysis, Writing - Review & Editing, Supervision, Project administration

Ubatan A. Miranda – Conceptualization, Writing - Review & Editing,

Wenisten J. D. Silva – Software, Validation

Herman A. Lepikson – Investigation, Writing - Review & Editing, Supervision

Highlights

- Deep learning for the classification of equipment failures using 2 public datasets;
- Increased accuracy of the model using only one vibration sensor for both datasets;
- Fine tune of the model hyperparameters to increase the accuracy of the model.