XLIII CILAMCE
21-25, November, 2022
FOZ DO IGUAÇU - PR - BRAZIL

Prediction failure in electric motors bearings using vibration signals and Long Short Term-memory neural networks

Rodrigo C. Campos[1], Gabriel T. Zago1[1], Luiz A. Pinto[1]

[1]*Programa de Pós-Graduação em Engenharia de Controle e Automação (ProPECaut), Instituto Federal do Espírito Santo (IFES) - Campus Serra*
Avenida dos Sabiás, 330 - Morada de Laranjeiras, 29166-630, Serra-ES, Brasil
rodrigoccampos@gmail.com, gabriel.zago@ifes.edu.br, luiz.pt@ifes.edu.br

Abstract. Bearing failures modify the vibration regime of electric motors. The acquisition and analysis of these signals may provide important information about the operating condition of these components. In this context, the use of failure prediction techniques can ensure the rolling bearings will be always in good operating condition, ensuring the production processes continuity and avoiding accidents. This paper investigates the subject of failure prediction in rolling bearings from vibration signals using Long Short-Term Memory (LSTM) networks. The experimental was carried out on the vibration signals from the data set IMC. The method consisted of building 6 models in different training settings by using either raw data or 13 statistical descriptors in time domain. Performance evaluation was accomplished by means of accuracy, precision, sensibility, sensitivity and F1-Score. The best result (92% of accuracy, 94% of precision, 86% of sensibility, 94% of specificity and 89% of F1-score), indicates the use of LSTM aiming to predicting failures in rolling bearings can improve the reliability of production systems, by anticipating preventive actions and reducing the need for corrective maintenance.

Keywords: prediction failure, rolling bearing, electric motors, LSTM.

1 Introduction

Industrial environments comprise different types of equipment that integrate complex production systems. The most vulnerable to failure are electric motors. Due to their dynamic nature, electric motors are subject to the action of centrifugal and frictional forces, and forces resulting from the vibration of moving parts. Such system of forces acting on the equipment causes operational wear on the moving parts, which over time can lead to failures and unplanned shutdowns. Unplanned shutdowns, notably those provoked by failures, usually, interrupt the production process leading to financial losses. The use of efficient strategies to estimate and control the evolution of failures over time, allows the maintenance team to schedule shutdowns, mitigating their effects on the production. Correcting failures by means of scheduled maintenance increases equipment availability and reliability, reducing the impacts on the production.

In general, electric motors abnormal functioning conditions can be noticed in incipient stages, disturbing performance, even before their consolidation. A motor in an incipient failure stage may present several abnormal operation signs, such as, changes in mechanical vibration regime, variations in operating temperature and alterations in insulation resistance. Therefore, the development of failure diagnosis systems through signal analysis using machine learning techniques can identify failures still in initial stage. In this way, preventive interventions may be scheduled before the incipient failures evolve to more dangerous stages, resulting in equipment damages and production's break off.

Due to operational conditions, electric motors rolling bearings can be considered the most important components of this equipment. Under proper operating conditions, bearings can reduce stress on the motor structure by reducing vibrations as well as wear on mechanical parts. In this context, by using failure prediction techniques based on machine learning, preserves good operating conditions of rolling bearings, ensuring the motor functioning and, as a consequence, production process continuity. In addition, using predictive strategies in industrial environments reduces maintenance costs, as well as production downtime.

Several studies have proposed the development of bearings failures prediction techniques, through the analysis of signals, using machine learning algorithms. As bearing failures are strictly related to vibration regime, the acquisition and analysis of such signals may provide important information about the operating condition of those components. This paper investigates an approach to estimate failure occurrences in rolling bearings of electric motors. The proposed approach uses as predictor a Long Short-Term Memory (LSTM) network, trained with 13 statistics descriptors obtained from the vibration signals available in the Intelligent Maintenance Systems (IMS) data set, provided by University of Cincinnati. The remainder of the paper is organized as follows, the next section presents related works. Then, there is a description of the materials and methods, followed by the experiments, results, and discussion. Finally, we present the conclusions and future works.

2 Related work

The use of machine learning to investigate failures occurrences in industrial processes and equipment has been a topic of great interest to the researchers of artificial intelligence community for several decades. Early studies in the area investigated the machine learning application to the diagnosis of failures which has already occurred. In recent years, algorithms capable to models the temporal evolution of the phenomena has been proposed, what made possible the estimation of failure even before it has been occurred.

A method to predict the bearing remaining useful life was presented in Pan et al. [1]. The proposed approach uses the Relative Root Mean Square value (RRMS) to classify the bearing's operating condition into two stages: normal operation or in failure process. According to the authors, in normal operation condition, the RRMS value keeps constant, however, when the bearing's degradation process starts, the RRMS value increases, accordingly. Since the RRMS value of the vibration signal indicates the bearing has entered the failure stage, 13 statistical descriptors are extracted and inputted in a Extreme Learning Machine neural network for the bearing's remaining useful life prediction. The reported results show the proposed method is highly accurate in short-term predictions, and has high-speed response in cases where the training sample size is limited.

In Gu [2] a combination of the wavelet transform and Long Short-Term Memory (LSTM) neural network was proposed to diagnosis the bearing's degradation level using vibration signals. The method, named DWT-LSTM consists of two stages. At first, the DWT is used to transform the vibration signals from time domain to time-scale domain. The output of the wavelet transform application step are the wavelet coefficients from the vibration signals represented as time series. Then, the LSTM network is used to model the long-term dependencies into the time series. Since long-term dependencies are characterized, failure prediction can be carried out precisely.

In Liu [3] the authors proposed a prediction approach for bearing's remaining useful life combining a Elastic Net with a Long Short-Term Memory (LSTM) network. According to the authors, in predicting the bearing remaining lifetime through LSTM network, the E-LSTM algorithm takes into account the space-time correlation. To reduce the over-fitting problem in training phase, the elastic net regularization term is inputted in the structure of the LSTM network. The authors reported the E-LSTM is able to more accurately characterizes the vibration signals changes caused by bearing degradation. In addition the authors state the E-LSTM algorithm achieves higher stability in predicting the bearing's remaining useful life compared to other approaches.

The authors in Liu et al. [4] proposed a method called LSS, which was applied to failure prediction in wind turbine bearings. The method combines together the advantages of LSTM neural network and statistical processes concepts. Initially, descriptors from the vibration signals in time domain are extracted. Since, failures present a multi-stage degradation pattern, statistical process is used to separate the samples according to the bearing degradation level. Then, the signals categorized into classes, according to the degradation stage, are provided as input to the LSS model for prediction. The authors reported the proposed method can obtain better prediction performance than recurrent neural networks and support vector regression.

A method based on Maximum Correlation Kurtosis Deconvolution (MCKD) and LSTM neural network was proposed in Ma et al. [5] to early predicts bearing failures from vibration signals. In the first stage, the Cuckoo Search (CS) algorithm is used to optimize the parameters, filter length and the period of the MCKD deconvolution. In the second step, the LSTM learning rate is adjusted, according to the behaviour of the time series deconvolution. Than, the LSTM network is trained and applied to perform failure prediction.

3 Long Short-Term Memory - LSTM

Long Short-Term Memory (LSTM) networks proposed by Hochreiter and Schmidhuber [6], addressed the issue of RNN long-term dependency, due to the vanishing gradient problem that emerges when working with longer data sequences. While on the one hand RNNs are able to make more accurate predictions based on current data, on the other hand they do not perform well with data stored in long-term memories. RNN does not provide an efficient performance as the gap length rises. LSTM is a Recurrent Neural Network that can learn order dependence, and keep information for long time. Due to its capacity to model long-term dependencies, LSTM are extensively used for time-series data processing, prediction, and classification.

A typical LSTM architecture consists of a cell, which remembers values over arbitrary time intervals, an input gate, an output gate, and a forget gate. The gates control the information flow through the cell. The forget gate controls what information should be forgotten. In the input gate occurs the addition of useful information to the cell state. The task of extracting useful information from the current cell state to be presented as output is done by the output gate.

4 Method and materials

This section describes the strategy's implementation steps to failures prediction in bearings of electric motors. The IMC data set, Lee et al. [7], is presented as well as the procedure to extracts descriptors from the vibration signals. In addition, the LSTM architecture, parameters settings and training settings used to failures predictions are described.

4.1 Dataset

The experimental apparatus to generates the IMS data set, Lee et al. [7], consisted of four bearings type Rexnord ZA-2115 double row bearings. The rolling bearings were, installed on a shaft coupled to an electric motor via belts, and were force lubricated by means an oil circulation system which regulates the flow and temperature of the lubricant. Vibration signals were captured by a set of high sensitivity quartz ICP accelerometers, installed in the bearing housing (two accelerometers for each bearing [x- and y-axes] for data set 1, one accelerometer for each bearing for data sets 2 and 3). During the experiment, the shaft maintained its constant speed at 2,000 rpm. A radial force of 6,000 lbs was applied to the shaft and the bearings by a spring mechanism. The four rolling bearings ran from the beginning of its useful life, until one of them be completely damaged (test-to-failure experiment).

In order to generate the data included in the IMS set, three experiments were accomplished. The experiments consisted of run-to-failure tests and yielded the data set 1, data set 2 and data set 3. The data sets stores a specific amount of files, each one corresponding to 1 second acquisition of the vibration signal, made up of 20,480 samples collected in a sampling rate set at 20kHZ. As a decision making, in this work, it was decided to use the vibration signals of the data set 1. Data set 1 contain 2,156 files per bearing, and it was finished by inner race failure in bearing 3 and roller element failure in bearing 4. During the run-to-failure test in the bearing 1, five operational conditions (classes) were identified: (Class 1) early - initial condition; (Class 2) suspect - suspect of failure; (Class 3) normal - normal condition; (Class 4) suspect - suspect of failure e (Class 5) imminent failure. As an example, Fig. 1 illustrates the firsts 1000 vibration signals samples of the bearing 1 in each operational conditions. For illustrative purpose, suspect classes (class 2 and class 4) were merged.

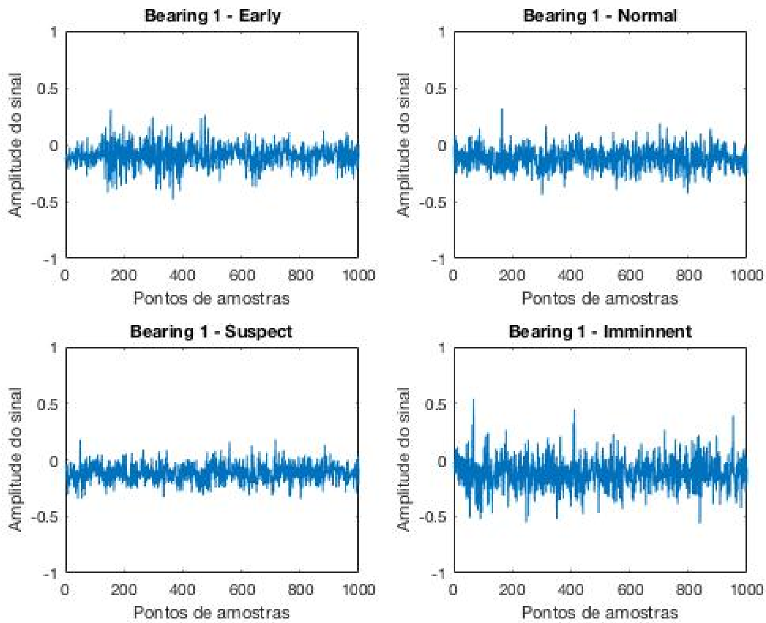

Figure 1. Vibration signals samples representing the bearing 1 operational conditions.
Source: Authors.

4.2 Statistical descriptors in time domain

According to Nandi and Ahmed [8], automatic monitoring by means of vibration signals analysis in time domain can be accomplished, in a efficient way, using statistical descriptors. The authors in Rauber et al. [9], Nayana and Geethanjali [10] and Tahir et al. [11], state that failures behaviors represented by vibration signals may be adequately described by statistical parameters. In this way, the following statistical descriptors were used to model vibration signals in the present work: mean value (X_m), root amplitude (X_{root}), root mean square value (X_{rms}), peak amplitude (X_{peak}) and standard deviation (X_{std}). Such attributes reflect variations in the intensity of consolidated failures, however, they are not sensitive to failure detection in early stages. To compensate for the low sensitivity of the previously mentioned parameters, other statistical measures are included, such as, skewness ($X_{skewness}$), kurtosis ($X_{kurtosis}$), crest factor (X_{crest}), clearance Factor ($X_{clearance}$), shape factor (X_{shape}), impulse Factor ($X_{impulse}$), peak to peak amplitude ($X_{peak2peak}$) and root sum of squares (X_{rss}). By using the referred parameters to represent vibration signals in time domain, may result in diagnosis systems sensitive to early failures, that still keep good performance while severity increases, Lei [12]. Mathematical formulation of each statistical parameter is shown in Table 1.

Table 1. Statistical parameters in time domain.

Parâmetro	**Descrição Matemática**
Mean value	$X_m = \frac{\sum_{n=1}^{N} x(n)}{N}$
Peak amplitude	$X_{peak} = max\|x(n)\|$
Clearance Factor	$X_{clearance} = \frac{X_{peak}}{X_{root}}$
Root amplitude	$X_{root} = \left(\frac{\sum_{n=1}^{N} \sqrt{\|x(n)\|}}{N}\right)^2$
Kurtosis	$X_{kurtosis} = \frac{\sum_{n=1}^{N} (x(n)-X_m)^4}{(N-1)X_{std}^4}$
Impulse Factor	$X_{impulse} = \frac{X_{peak}}{\frac{1}{N}\sum_{n=1}^{N} \|x(n)\|}$
Standard deviation	$X_{std} = \sqrt{\frac{\sum_{n=1}^{N} (x(n)-X_m)^2}{N-1}}$
Skewness	$X_{skewness} = \frac{\sum_{n=1}^{N} (x(n)-X_m)^3}{(N-1)X_{std}^3}$
Shape form	$X_{shape} = \frac{X_{rms}}{\frac{1}{N}\sum_{n=1}^{N} \|x(n)\|}$
Root mean square value	$X_{rms} = \sqrt{\frac{\sum_{n=1}^{N} (x(n))^2}{N}}$
Crest factor	$X_{crest} = \frac{X_{peak}}{X_{rms}}$
Peak to peak amplitude	$X_{peak2peak} = 2.X_{peak}$
Root sum of squares	$X_{rss} = \sqrt{\sum_{n=1}^{N} \|x_n\|^2}$

Source: Adapted from Lobão [13].

4.3 Preprocessing and data arrangement

In order to obtain the statistical descriptors arrangements to feed the LSTM network, the following steps were accomplished. At first, original signals (signals made up 20,480 samples) of the data set 1 were segmented in pieces with different lengths, according to the dimension of the vibration signal matrix used in the models building phase. Figure 2 (in the left) illustrates the procedure for bearing 1, considering segments with 1000 samples. Segment 1 comprises the interval between samples 1 and 1000. Segment 2 comprises the interval between sample 2 and 1001, and so on.

Secondly, resulting segments of the previous step were stacked in a vibration signals matrix. In the last step, for each line of the vibration signals matrix, the 13 statistical descriptors presented in Section 4.2 were calculated. The right side of Fig. 2 shows the descriptors matrix related to a single original signal of the bearing 1, as can be noted, the descriptors matrix is a transposed version of the descriptors matrix obtained in the last step. The referred procedure was replicated for all signals of the four bearings in data set 1. The resulting descriptors matrix that fed the LSTM network was obtained by the concatenation of the descriptors matrix of each bearing. In this way, the resulting descriptor matrix was made up of 52 lines. In addition, due the similarity between them, in the model building phase, classes related the suspect conditions (Class 2 and Class 4), in Section 4.1, and the early and normal classes (Class 1 and Class 3) were merged. Therefore, the prediction problem can be formulated as a three classes problem ans 52 input descriptors.

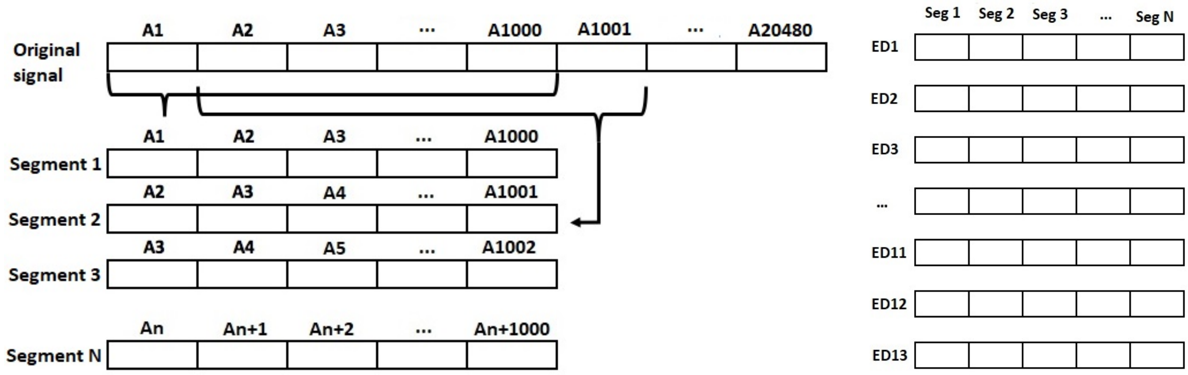

Figure 2. Procedure to obtain the descriptors matrix. In the left the original signal segmentation. In the right the descriptors matrix. Source: Authors.

4.4 LSTM architecture and training settings

LSTM network architecture consisted of a input layer with 52 dimension. To modelling the temporal relations in the vibration signals, it was used a LSTM structure made up of 50 hidden layers. A fully connected layer with three neurons, followed by a softmax activation function was used in the output of the LSTMs layers. The output classification layer classifies samples in one between three possible classes (normal, suspect and imminent). The optimization method used was the Adam, because, according to Kingma and Ba [14], it is the most appropriate method for non-stationary problems.

To build the models different training settings were used (Table 2 in Section 5). Each model consisted of a specific combination of values of the following parameters: descriptor type (Raw Data and Statistical Descriptors), mini-batch size (10 and 64), segment length of the vibration signal (24480, 100 and 1000), and the method for model validation (holdout and k-fold). In validation applying holdout, training and validation sets were partitioned in 70% and 30%, respectively. When validation was carried out by means of k-fold, the k value was set at 3. All models were trained in 1,500 epochs. In order to evaluate performance of the prediction failures models, the following metrics were applied: accuracy, precision, sensitivity, specificity, and f1-score.

5 Results and discussions

Table 2 presents the prediction results of the 6 models. In columns 2, 3 and 4, Descrip, LenSeg and SizBatch refer to Descriptor Type, Segment Length and Mini-Batch size, respectively, and EstDesc refers to Statistical Descriptors. As can be seen, when compared to the models built using statistical descriptors (Model 4, Model 5, and Model 6), the models built using the raw data (Model 1 and Model 2), regardless of segment length, did not perform well in predicting the failures. Such result indicates that, while LSTM layers have good capability for representing temporal relationships in the vibration signals, they are limited as features extractors. *NaN* values of accuracy and F1-Score in Table 2 indicate the corresponding models showed low predictive ability for the suspect and imminent classes.

If compared performances of models 3 and 4, whose only difference in training settings is the segment length, can be noted the use of larger segments results in models with better predictive ability. One can speculate that, for modelling the vibration signals of IMS set, the use of small segments may not make possible the LSTM layers get good representation of the temporal relationships embedded in the data.

The best results were obtained with models 4, 5 and 6. These models were built with the same segment length and all used statistical descriptors. However, for the construction of model 4 the mini-batch was set to 10, and for models 5 and 6, this parameter assumed value 64. It can be seen that the segment length equal to 1000, significantly improves the performance of these models, compared to model 3, also built using statistical descriptors.

According to Table 2, the model 6 is the best one in predicting failures in the IMS dataset. Compared to model 5 (both were built with statistical descriptors, segment length equal to 1000 and mini-batch 64 - the only difference is validation strategy), model 6 performed slightly better, considering sensitivity and F1-Score. Figure 3 illustrates the plot of accuracy and loss of model 6, trained with 1500 epochs. It can be seen the model converges to the final accuracy and loss values in about 1000 epochs.

Table 2. Models's performance considering the different training settings

Model	Descrip	LenSeg	SizBatch	Validation	Accuracy	Precision	Sensibility	Especificity	F1-Score
Model 1	RawData	20480	10	Holdout	0.5753	*NaN*	0.4065	0.7582	*NaN*
Model 2	RawData	100	10	Holdout	0.6164	*NaN*	0.3333	0.6667	*NaN*
Model 3	EstDesc	100	10	Holdout	0.6164	*NaN*	0.3333	0.6667	*NaN*
Model 4	EstDesc	1000	10	Holdout	0.8971	0.9307	0.6459	0.9263	0.6616
Model 5	EstDesc	1000	64	Holdout	0.9187	0.9333	0.8202	0.9356	0.8562
Model 6	EstDesc	1000	64	K-Fold	0.9196	0.9451	0.8616	0.9409	0.8952

Source: Authors.

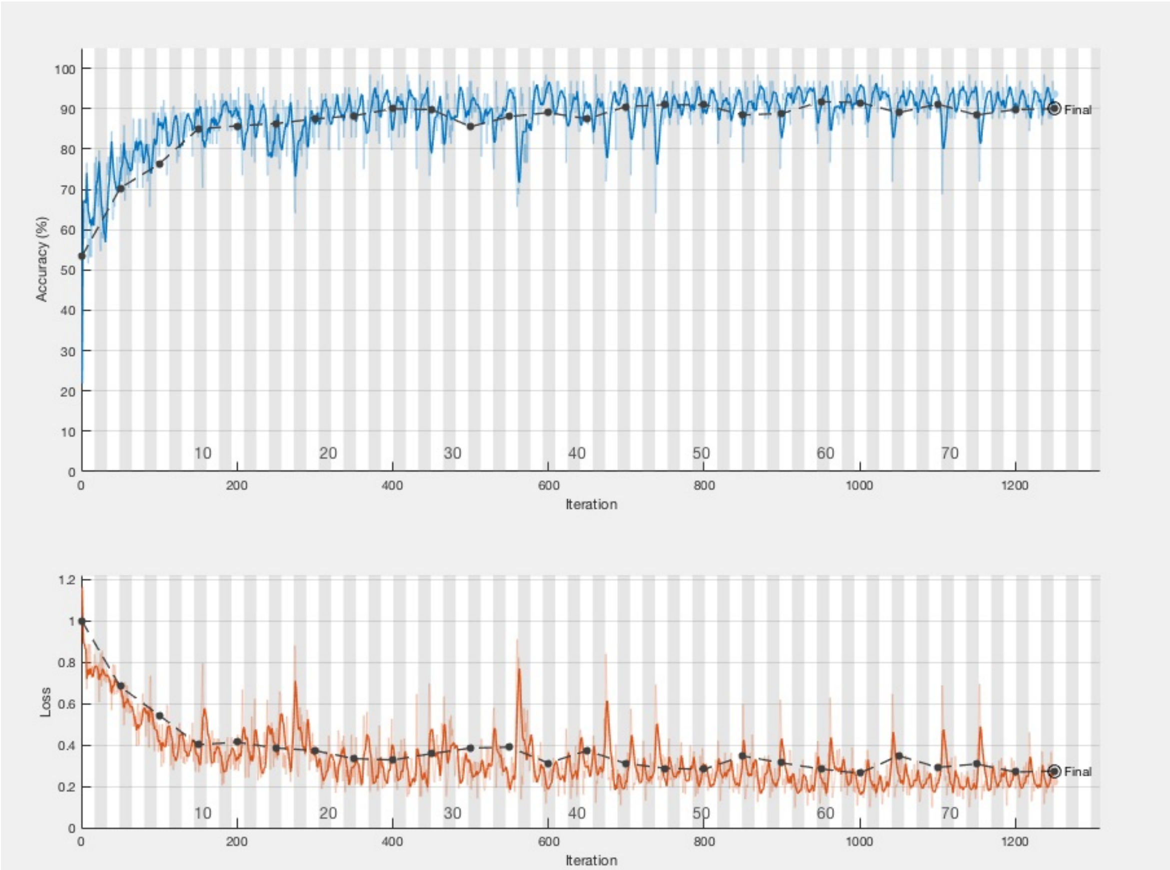

Figure 3. Performance of the best model (model 6), considering accuracy and loss after 1,500 epochs.
Sorce: Authors.

6 Conclusion and future works

This paper investigated the bearing failures prediction using LSTM neural networks. The IMS dataset was applied to build the models, by using original data and statistical descriptors extracted from the vibration signals. By analyzing the models's performances, one may conclude the set of 13 statistical descriptors used to represent the vibration signals can provides a good modelling of the phenomena embedded in the original signals. Furthermore, the poor performance of the models with the raw data shows the limited ability of LSTM networks to extract complex features from nonstationary signals.

Another important aspect is that proper tuning of the parameters, segment length and mini-batch size is critical for obtaining models with good predictive ability. In general, the best values for these parameters are obtained by trial and error. Future works may consider the integration of CNN and LSTM networks. In this case, the CNN module would be used to extracts features of the vibration signal and the LSTM would model the temporal relations in the signals. The authors believe such strategy may, significantly, improves the prediction performance.

Acknowledgements. This work was supported by the CAPES/FAPES (process: 2021-CFT5C, nº FAPES 133/2021) in PDPG (Programa de Desenvolvimento da Pós-Graduação - Parcerias Estratégicas nos Estados). Authors thanks the support from Instituto Federal do Espírito Santo (IFES).

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] Z. Pan, Z. Meng, Z. Chen, W. Gao, and Y. Shi. A two-stage method based on extreme learning machine for predicting the remaining useful life of rolling-element bearings. *Mechanical Systems and Signal Processing*, vol. 144, pp. 106899, 2020.
[2] Y. L. X. L. H. R. M. K.; Zhang. Gu. Dwt-lstm-based fault diagnosis of rolling bearings with multi-sensors. *Eletronics*, vol. 10, 2021.
[3] X. D. W. H. L. C. L. L. B. L. W. Z. H. C. L. Z. H.; Meng. Liu. A regularized lstm method for predicting remaining useful life of rolling bearings. *International Journal of Automation and Computing*, vol. 18, 2021.
[4] J. Liu, C. Pan, F. Lei, D. Hu, and H. Zuo. Fault prediction of bearings based on lstm and statistical process analysis. *Reliability Engineering System Safety*, vol. 214, 2021.
[5] L. Ma, H. Jiang, T. Ma, X. Zhang, Y. Shen, and L. Xia. Fault prediction of rolling element bearings using the optimized mckd–lstm model. *Machine*, vol. 10, 2022.
[6] S. Hochreiter and J. Schmidhuber. Long short-term memory. *Neural Computation*, vol. 9, n. 8, pp. 1735–1780, 1997.
[7] J. Lee, H. Qiu, G. Yu, and J. Lin. Bearing data set. ims, university of cincinnati. nasa ames prognostics data repository, nasa ames, moffett field, ca. Disponível em: `http://https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/`. Acesso em: 20 de novembro de 2020, 2015.
[8] A. K. Nandi and H. Ahmed. *Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines*. John Wiley & Sons, 2020.
[9] T. W. Rauber, F. de Assis Boldt, and F. M. Varejão. Heterogeneous feature models and feature selection applied to bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, vol. 62, n. 1, pp. 637–646, 2015.
[10] B. R. Nayana and P. Geethanjali. Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal. *IEEE Sensors Journal*, vol. 17, n. 17, pp. 5618–5625, 2017.
[11] M. M. Tahir, A. Q. Khan, N. Iqbal, A. Hussain, and S. Badshah. Enhancing fault classification accuracy of ball bearing using central tendency based time domain features. *IEEE Access*, vol. 5, pp. 72–83, 2017.
[12] Y. Lei. *Intelligent fault diagnosis and remaining useful life prediction of rotating machinery*. Butterworth-Heinemann, 2016.
[13] D. A. Lobão. Técnicas de aprendizado de máquinas aplicadas ao diagnóstico de falhas em equipamentos industriais. Mestrado profissional em engenharia de controle e automação, Universidade Federal do Espírito Santo, Serra, 2020.
[14] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.