Bearing failure prognostics using recurrent neural networks: a spectral data based architecture

Renato Cosin
University of São Paulo (USP)
São Carlos - SP, Brazil
renato.cosin@gmail.com

Leo Sampaio Ferraz Ribeiro
University of Campinas (UNICAMP)
Campinas - SP, Brazil
leosfr@unicamp.br

Kalinka Castelo Branco
University of São Paulo (USP)
São Carlos - SP, Brazil
kalinka@icmc.usp.br

***Abstract*—Predictive maintenance is crucial for reducing costs in the industry. With the widespread use of internet-connected sensors in industrial equipment, state-of-the-art predictive maintenance algorithms have become a prolific field for innovation. In this paper we present a new deep learning solution for predicting the remaining useful life of bearings. Bearings are widely used in industrial equipment, and their failure prognosis is highly relevant. The developed model takes spectrograms of vibration signals from bearings as input and computes their remaining useful life as output using a combination of Convolutional and LSTM neural networks. The model hyperparameters were optimized using the Hyperband algorithm. The dataset used originates from a large accelerated degradation experiment aimed at evolving bearing failure prognosis techniques, made publicly available as part of the IEEE PHM 2012 Data Challenge. The optimized model presented satisfactory results. In addition to reducing maintenance costs and downtime, the potential application in IIoT systems for online monitoring guided the architecture and data processing flow definition. Using a proposed criterion, the model successfully prescribed component replacement before failure in all test cases. While 20% of the maintenance was premature, the model accurately prescribed preventive maintenance for 80% of the test bearings. The model and data processing flow are relatively simple and compatible with IIoT systems, allowing for low-cost edge inference.**

I. Introduction

Predictive maintenance (PdM) plays a crucial role in reducing costs and downtime in industrial processes by providing accurate prognosis and diagnosis of equipment and system failures. The integration of Internet of Things (IoT) sensors in industrial equipment (IIoT) has opened up new possibilities for PdM techniques. The vast amount of data generated by IoT telemetry systems and highly automated factories has made the combination of PdM and machine learning (ML) a fertile ground for innovation.

While established failure prognostics techniques rely heavily on domain knowledge and specialized software [1], [2], recent improvements in data-driven models – specially those reliant on deep neural networks – have shown an alternative path that is easier to implement and not as dependent on extensive knowledge about the equipment being monitored. Our proposed solution is one of those methods, as we make use only of vibration data to predict Remaining Useful Life (RUL) of common rolling-element bearings.

The authors would like to thank FAPESP (grant #2022/12690)

979-8-3503-3872-0/23/$31.00 ©2023 IEEE

Online PdM systems are integrated into complex environments that involve multiple data sources and distributed processing in the edge and in the cloud. These systems must adhere to architectural constraints, computational power, and data transmission bandwidth limitations.

Regarding connectivity for IIoT sensors, a crucial data source for training and inference models, low-power wide-area networks (LPWANs) have gained significant popularity. LPWANs have low power consumption, operate in unlicensed bands, support battery-powered devices, have long range, and have low susceptibility to interference. However, their main disadvantage is their low bandwidth.

There is a wide range of applications in monitored equipment and components. However, common mechanical components such as bearings and gear boxes are particularly important due to their widespread use [1]–[3].

An essential challenge is the development of predictive methods that combine simplicity, generalization capability, and low requirements for pre-processing data. In addition, the potential for execution on edge devices might be critical for new predictive models aiming for successful field deployment. These constraints were used as guidance throughout our model design.

Recurrent Neural Networks (RNNs) [4], [5] are widely used for predictive time series data modeling. These networks handle serial data and are often applied to supervised learning problems. RNNs can process various types of data, from quasi-static sensor time series to spectral data from vibration sensors. In the latter case, RNNs can incorporate convolutional and pooling layers for reducing data dimensionality when dealing with high-dimensional and high-volume inputs.

This study aims to evaluate an online-capable flow analysis of the remaining bearings useful life using a predictive model based on combined CNN-RNNs, that take as input time series of vibration data processed as spectrograms. The study comprises a novel data processing pipeline, model hyperparameter tuning, training and evaluation.

For this purpose, the study will assess using a combination of Convolutional Neural Networks (CNNs) and RNNs for bearing condition monitoring, focused on the algorithm potential. The model architecture evaluation was conducted. This work seeks to develop a predictive model with performance comparable to similar techniques present in the literature, especially considering accuracy in the prognosis of bearing failures.

II. Related Work

A. Vibration analysis

Vibration analysis is one of the most important and applicable used techniques in Data Science (DC) for predictive maintenance [6]. This technique involves the specialized processing and analysis of mechanical and acoustic vibration signals, usually collected through accelerometers or microphones.

Popescu et. al. [6] describe various frequency domain techniques for vibration analysis, including the characteristics extraction from the frequency spectrum, such as entropy and spectral kurtosis. Other noteworthy approaches include the spectrum envelope, change detection, and principal component analysis. Additionally, some techniques deal with vibration signals in the time domain.

One widely used technique for vibration analysis in the frequency domain is the Discrete Fourier Transform (DFT), specifically in the form of a computationally efficient algorithm called Fast Fourier Transform (FFT), as described by Jardine et al. [7].

B. Bearing Failure Prognostics

Bearings are critical components in various types of industrial equipment and are responsible for a significant fraction of failures in this equipment, as stated by Shenfield et al. [8]. Predictive maintenance of these components typically involves vibration analysis using techniques based on frequency and time domains. Data-based methods, such as ML, have also been increasingly employed, using characteristics extracted from the signals.

Given the relevance of this topic, challenges have been carried out to predict remaining useful life, and databases have been made available, as described by Nectoux et al. [9] and Hiang et al. [10].

C. RNN-CNN networks for spectrograms of Time Series Data

In time-series-dependent applications, RNNs are often used to capture the temporal data pattern. Shenfield et al. [8] propose an algorithm for RUL prediction in bearings using RNNs coupled with CNNs. Huang et al. [10] also suggest associating these two types of neural networks for the same purpose. In both cases, CNN constitutes the first layer of treatment for the vibration signal, and RNN captures relevant temporal information for RUL prediction. Our design differs from these previous studies in that we are the first to use spectrograms (instead of raw data) and 2D CNN-RNNs for bearing RUL prediction. The use of spectrograms and 2D CNN-RNN designs is a trend in sequence modeling and was successful in: typical audio applications [11], Driver Identification [12], Arm Motion Classification [13], and ECG Arrhythmia Classification [14].

III. Methodology

In this section we present our combined model for predicting RUL in rolling-element bearings. Our models is, to the best of our knowledge, the first to make use of spectral data from bearing accelerometers processed through 2D CNN-RNN architecture.

The development followed a typical workflow for ML solutions, relying on the CRISP-DM methodology [15] (Task Understanding, Data Understanding, Data Preparation, Modeling, and Evaluation).

A. Data Processing Pipeline

Our goal is to bring to this problem the benefits seen in other fields, from (a) transforming time series data to spectrograms and (b) processing those spectrograms through convolutional and recurrent neural networks. Our work starts with the extraction of spectrograms from raw accelerometer data.

To create our spectrograms we have used the power spectral density, computed with Welch et. al [16] method using 1024 points per segment with a stride of 64 (allowing then for overlap between segments). We have also experimented with other segment lengths and found – through visual inspection – this value to be the best at avoiding spurious peaks in the result. We create one spectrogram for each collection window from the original dataset and combine sequences of spectrograms to input our CNN-RNN network. A visual data representation is presented in Fig. 1.

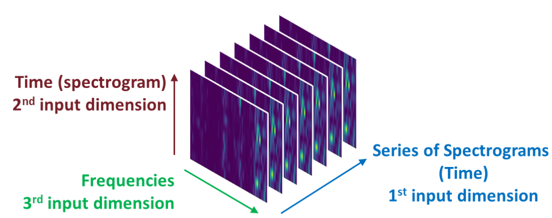

Fig. 1. Graphic representation of the structure of the entries. The fourth dimension represents the input channels of the first convolution layer. In this study one channel is used.

In line with our goal of making this model useful under compute constraints, we perform dimensionality reduction on the frequency axis of the spectrograms by keeping only the maximum value every 4 points on the frequency axis, an operation equivalent to a max pooling layer. We found that this operation does not visually degrade the spectrogram.

Finally, we apply a simple yet effective augmentation on the data; the original set is comprised of data from accelerometers positioned vertically and horizontally so that each collection window has two series of points collected; we created a third series by emulating an accelerometer positioned at 45º by combining both vertical and horizontal series.

B. Dataset and test data split

The dataset used was introduced at the IEEE Prognostics and Health Management (PHM) 2012 Data Challenge by

Nectoux et al. [9]. It was created through an experiment with accelerated degradation of bearings. The data is comprised of accelerometer signals (vertical and horizontal) collected from 17 bearings up to the failure state. Each bearing was configured to have specific load and speed conditions. The signals were collected every 10s for a window of 0.1s at a frequency of 25.6kHz, and the longest-lasting bearing worked for 27710 seconds (Maximum Expected Life). Therefore, following our pre-processing pipeline each 0.1s collection window creates each spectrogram, combined in a series of 32 spectrograms to create each input sample (making for an RNN time-step of 32).

To avoid data leaks between the training and test sets, we split the bearings: 12 for training and validation and 5 exclusively for testing. The maximum life among the training bearings is 27710 seconds (a low value as this is an accelerated degradation test).

C. Loss function and RUL_{index}

The simple mean squared error (MSE) loss was chosen for this regression problem. Because MSE does not prioritize the distribution in any way, we introduce RUL_{index} as one of our core contributions: a transformation of the model's output magnitude designed to privilege greater precision for small RULs, i.e., in conditions close to bearing failure. Fig. 2 shows a diagram that illustrates the regions in a predicted RUL versus actual RUL diagram according to the model objective.

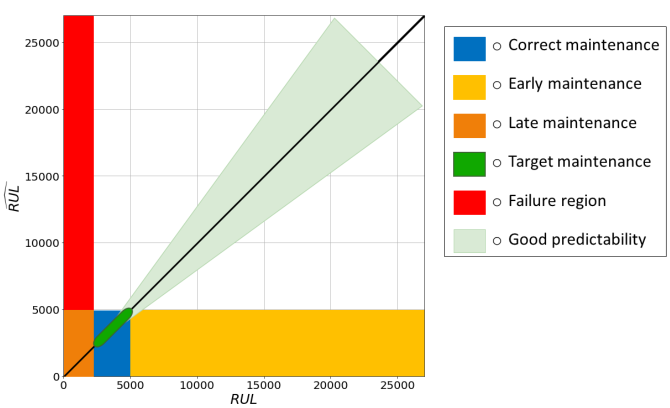

Fig. 2. Conceptual regions in the predicted RUL versus actual RUL diagram according to the model objective

We have then defined RUL_{index} as:

$$RUL_{index} = \frac{w \times MEL}{RUL + (w \times MEL)} \quad (1)$$

where $w = 0.25$ is the adjustment factor, $MEL = 27710$ is the Maximum Expected Life. In an ideal scenario MEL would be provided by a bearing manufacturer, in the absence of this information the highest RUL value found in the training set was used. In Fig. 3, we compare the distribution of the cost function as a function of RUL and the predicted $\widehat{RUL}$, where we can see how changes close to failure are prioritized when using RUL_{index}.

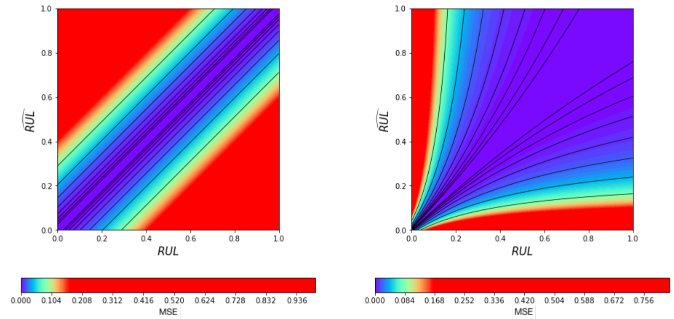

Fig. 3. Distribution of the MSE cost function as a function of RUL and $\widehat{RUL}$ using RUL directly, on the left, and using RUL_{index} on the right.

We have compared the effectiveness of RUL_{index} in Section IV; when non-transformed RUL is used min-max normalization was performed on the RUL value (using MEL).

D. Hyperparameter Tuning

Since we are the first to use a 2D CNN-RNN for this task, a network architecture had to be designed from scratch, and optimization hyperparameters were determined from validation. To avoid costly grid searches, we have used the low-cost Hyperband algorithm, described by Li et al. [17] and known for its focus on early stopping and speeding up random searches.

E. Model evaluation criteria

In addition to typical regression model evaluation metrics, the model results were processed to observe how models prescribe maintenance over time for each bearing until its failure, adhering closer to the application scenario. After computing the expected RUL (or RUL_{index}) for each of the three accelerometers, we calculate the Exponential Moving Average (EMA) for each result and take the mean of the three results as the "final" prediction. Using EMA makes sense as these systems are expected to be online, with the RUL threshold determining when maintenance should be executed.

F. Neural Network Architecture

This work aims to apply neural networks with RNN layers fed by spectrograms of vibration data. Because spectrograms can and often are interpreted as images, we use a 2D CNN module at the start to efficiently process high-dimensionality inputs and capture intermittent vibration patterns present in the spectrograms.

While designing neural networks from scratch is daunting, our architecture and the hyper-parameters we vary were chosen with edge applications and limited training data in mind. In Fig. 4, we show our proposed architecture, including optional layers which have their utility tested through Hyperband [18] tuning.

The convolution layers are intended to extract patterns from spectrograms – especially the short-term ones – while keeping the computational cost low. The series of these extracted patterns are then ingested by the RNN, which captures their

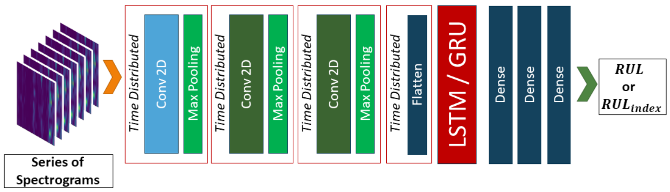

Fig. 4. Simplified representation of the neural network structure CONV 2D + RNN + DENSE. The boxes in green are optional and their presence is determined by the optimization process

evolution over time and extracts long-term trends from the vibration data. The convolution operations are then applied separately for each time series element.

For the present study, the specific instance of RNN used was the Long-short-term memory (LSTM) module [4]. This type of RNN presents several advantages when dealing with Vanishing Gradients [19] and, for our use case, does not present prohibitive computational complexity or cost. After processing through the LSTM layers the output is projected through three final fully connected layers to compute the predicted RUL value.

To keep the computational cost of training and inference as low as possible, all layers except the last one are of type 16 bits floating-point. The input data was also converted to 16 bits floating-point, significantly reducing its volume and benefiting the potential functioning of the model on IIoT systems with restricted band connectivity.

IV. Experiments and Discussion

A. Implementation Details

Our models were implemented under the Tensorflow/Keras framework. Final models (after parameter tuning) were trained with batch size 64 for at least 70 epochs, up to 90, with early stopping conditioned on the validation loss. For all models, the learning rate is exponentially decayed with a factor of 0.8 after each epoch, and 3 fully connected layers were used at the architecture top. The Hyperband tuner [18] was set to train each variant for 25 epochs, and the tuner was run for 3 iterations. The validation set was created from a random 20% split of the training set. The input dimension is $(32, 84, 25, 1)$, representing the number of spectrograms per timestep (32), the size of the spectrograms ($84x25$), and the channel dimension (1).

B. Hyperband Tuning

The Hyperband tuner was executed to perform efficient hyper-parameter search across the intervals defined in Table I; the same Table also shows the optimal parameters found. Our novel RUL_{index}, computed on the validation set, was used as the model output for the search.

With the hyper-parameters found, we have fully built and tested three further variants, ablating the use of RUL_{index} and whether a more compact version of the model works. Our Model A is our "default" model and uses RUL_{index} as output. Model B ablates the transformation, using instead RUL directly as output. Model C tests a lower compute footprint; the tuned architecture was kept except for the number of units in the LSTM layer and the first fully-connected layer, which were halved. These are the layers with the highest number of trainable parameters in the architecture.

TABLE I
SEARCH SPACE AND RESULTANT OPTIMAL VALUES FOR NUMERICAL TYPE HYPERPARAMETERS OF THE NEURAL NETWORK. THE DENOMINATIONS C1, C2 AND C3 REFER, RESPECTIVELY, TO CONVOLUTIONAL LAYERS 1, 2 AND 3. SIMILARLY, D1 AND D2 REFER TO DENSE LAYERS

Name	Type	min value	max value	Optimal
learning rate	Float	1.00E-08	1	0.000864
Layers of conv2D	Int	1	3	2
filters c1	Int	2	4	4
kernel c1	Int	(3, 3)	(7, 5)	(5, 4)
stride c1	Int	(1, 1)	(3, 3)	(1, 2)
filters c2	Int	4	8	8
kernel c2	Int	(3, 3)	(7, 5)	(6, 3)
stride c2	Int	(1, 1)	(3, 2)	(1, 1)
filters c3	Int	4	8	NA
kernel c3	Int	(3, 3)	(7, 5)	NA
stride c3	Int	(1, 1)	(3, 2)	NA
pooling kernel size	Int	(2, 2)	(3, 3)	(2, 2)
LSTM units	Int	16	128	112
size d1	Int	16	128	112
size d2	Int	8	32	16

C. Model Evaluation

1) Model training statistics: We start our analysis by comparing predicted RUL ($\widehat{RUL}$) and ground truth RUL in Fig. 5 for each model A, B, and C. As expected, predictions on the training data are very accurate. We can see, however, that when tested on unseen bearings, the model underperforms on average. However, most of the erroneous predictions are for larger values of RUL and, therefore, further from the maintenance window where most of the application value is.

To better show the phenomenon, we compute the Relative Change Difference (RCD) metric and plot a histogram of the value considering training and test predictions. The histogram in Fig. 6, computed for Model A, shows that predictions on the test set are scattered, and errors are more significant; it also shows, however, that forecasts trend toward lower than actual values for RUL, a mistake that has much lower impact, promoting early maintenance instead of post-failure maintenance. In the next section, we consider the models with these application-dependent constraints.

2) Application based model evaluation: The result of applying exponential moving average and averaging the prediction for all accelerometers is shown in Fig. 7. Good model performance on capturing the RUL trend is visible in all cases, with the caveat that because the initial remaining life is not known in a real scenario, significant variation is expected at the start of prediction.

We have additionally computed metrics for model ensembles. These ensembles are created by averaging the predicted RUL from our three model variants; we refer to them by their combined identifiers (e.g., Model AB for an ensemble of models A and B).

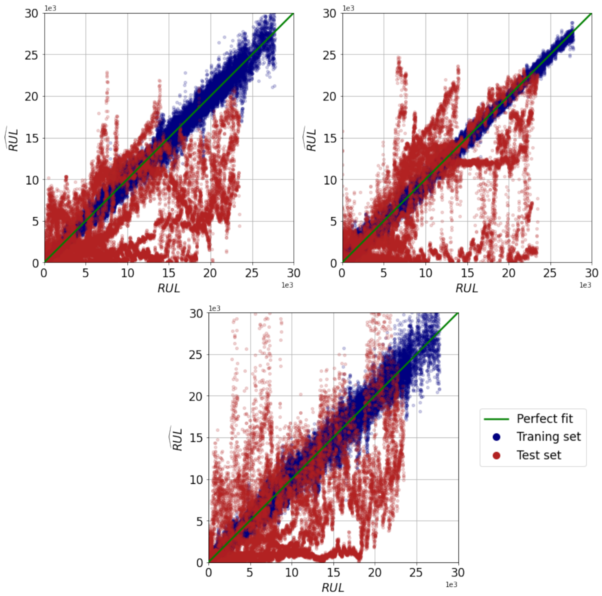

Fig. 5. Predicted versus ground truth RUL for models A, B and C, shown respectively from left to right, top to bottom.

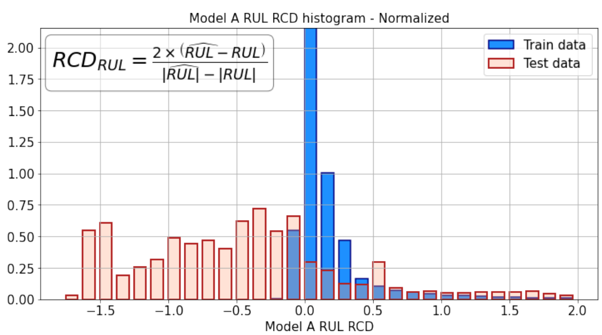

Fig. 6. RCD metric histogram for training and test data. The histogram of the training data had part of its bars truncated for better visualization of the test base results.

For all tested models, a time called $t_{replace}$ was computed as a threshold at which the model would prescribe bearing maintenance. The value was calculated by considering a target RUL of 10% of the maximum life expected.

By comparing this predicted $t_{replace}$ against the ground truth $RUL = 10\% \times MEL$ we can infer when a model prescribes premature or late maintenance of the bearing. This metric is illustrated for all bearings using the best model (Model BC) in Fig. 8. The results for all models in terms of this error as a percentage of MEL are presented in Table II. It can be observed that models A, B, and C deliver good performance given the challenges already discussed. Combined models generate positive outcomes and better meet the proposed objectives despite higher computational costs.

3) Comparison to related work: To allow for comparison against the work of Nectoux et al. [9] we have also computed the percent errors of RUL predictions as was done for the IEEE PHM 2012 challenge. These results are presented in

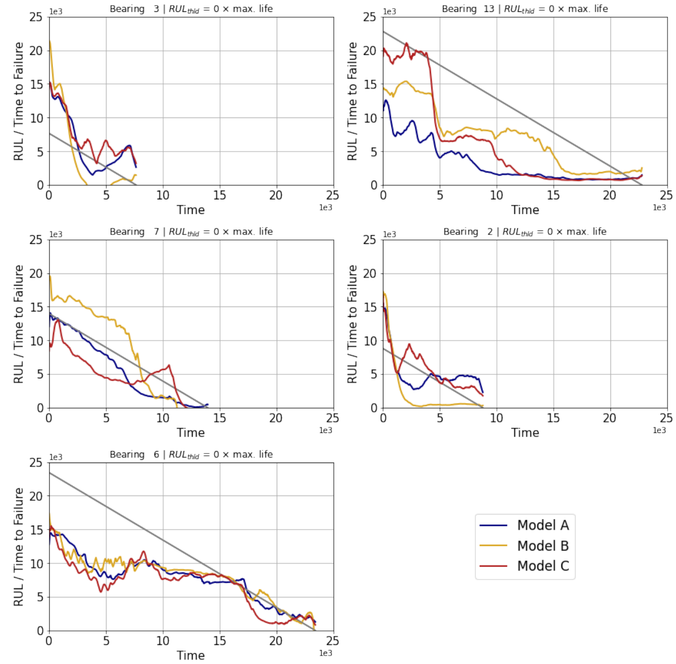

Fig. 7. Predicted vs Ground Truth RUL values shown for each of the 5 test bearings. The values were smoothed through exponential moving average and the average prediction from the three accelerometers is used.

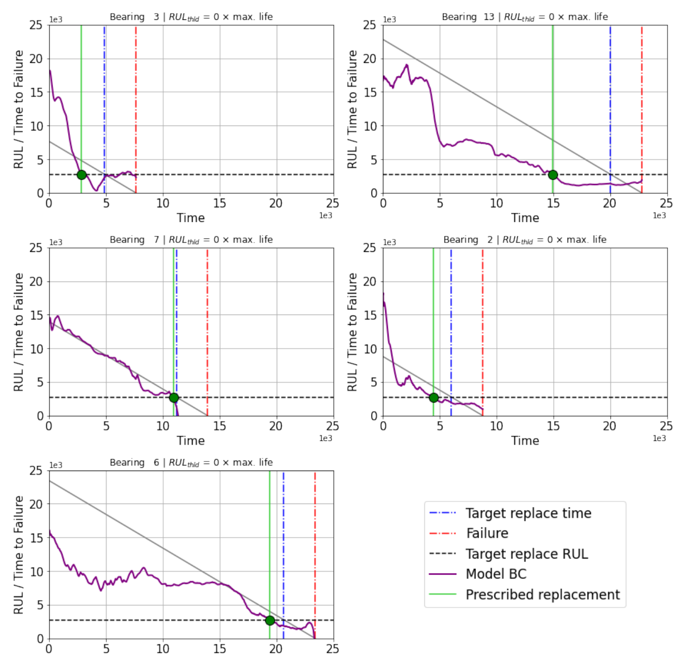

Fig. 8. Prognostic of bearing maintenance as predicted by Model BC using the $RUL = 10\% \times MEL$ criteria.

Table III, where it is observed that the best model was the combination of B and C, with the best non-combined model being A. The excellent performance of model C is also regarded, despite having approximately half of the parameters of models A and B.

TABLE II
ERROR OF PREDICTED $t_{replace}$ AS RELATIVE PERCENTAGE OF MEL FOR ALL MODELS AND TEST BEARINGS. POSITIVE VALUES ARE INDICATIVE OF PREMATURE MAINTENANCE AND NEGATIVE VALUES OF LATE MAINTENANCE.

	Rol. 3	Rol. 13	Rol. 7	Rol. 2	Rol. 6
Model A	6%	42%	13%	12%	2%
Model B	9%	16%	8%	16%	1%
Model C	18%	33%	0%	-8%	9%
Model AB	8%	18%	10%	15%	2%
Model AC	3%	36%	1%	-9%	7%
Model ABC	3%	36%	1%	-9%	7%
Model BC	7%	18%	1%	6%	4%

TABLE III
COMPARISON OF MODEL SCORES CALCULATED ACCORDING TO THE WORK OF NECTOUX ET AL. [9]. THE BEST RESULT IS HIGHLIGHTED IN BOLD.

Model	**Score**
Nectoux et. al. [9]	0.31
Our Model A	0.24
Our Model B	0.23
Our Model C	0.23
Our Model AB	0.23
Our Model AC	0.25
Our Model ABC	0.25
Model BC	**0.32**

V. CONCLUSION

This work proposed a new architecture for the prognosis of the useful life of bearings. The model selection and the data processing flow involved were designed with the constraints of IIoT systems in mind. The main pillars defining the proposed architecture are the use of spectral data and a combined RNN-CNN model.

We emphasize that our proposed model is not dependent on knowing the load or rotation speed of the bearing and that it has a meager inference cost. We also provide an even smaller but well-performing model for embedded systems.

After adopting the criterion of replacing the bearing with RUL of 10% of the maximum expected life, in none of the 5 cases evaluated, did the bearings fail before the criterion was reached. A premature replacement was prescribed in only one case, bearing 13. It is worth noting that in a real potential application in industrial systems, unscheduled replacements, and failures tend to be very expensive, and a premature replacement, although sub-optimal, is highly preferable.

The main contribution of this work is the proposal of a new model and data processing flow for RUL prognosis in rotating machines and components; the first study to make use of spectrograms and 2D CNNs to address this task. A model combining CNNs and RNNs was developed, trained, and evaluated that predicts the RUL of bearings from time series of spectrograms. The proposed model can be considered relatively simple and compatible with IIoT systems.

The model results were satisfactory, with complexity and computational cost within acceptable limits and aligned with the proposed objectives. Therefore, this study is expected to serve as a basis or reference for future work in the area of failure prognosis. Although challenging in modeling and data acquisition, this area within predictive maintenance has many applications and allows significant gains for users.

REFERENCES

[1] D. Hochmann and E. Bechhoefer, "Envelope bearing analysis: Theory and practice," in *2005 IEEE Aerospace Conference*, Mar. 2005, pp. 3658–3666.

[2] Y. Liu, Q. Qian, F. Liu, S. Lu, Q. He, and J. Zhao, "Wayside Bearing Fault Diagnosis Based on Envelope Analysis Paved with Time-Domain Interpolation Resampling and Weighted-Correlation-Coefficient-Guided Stochastic Resonance," *Shock and Vibration*, vol. 2017, p. e3189135, Jan. 2017.

[3] H. Huang and N. Baddour, "Bearing vibration data collected under time-varying rotational speed conditions," *Data in Brief*, vol. 21, pp. 1745–1749, Dec. 2018.

[4] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "Lstm: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222–2232, 2016.

[5] J. Zhu, H. Chen, and W. Ye, "A Hybrid CNN–LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar," *IEEE Access*, vol. 8, pp. 24 713–24 720, 2020.

[6] T. D. Popescu, D. Aiordachioaie, and A. Culea-Florescu, "Basic tools for vibration analysis with applications to predictive maintenance of rotating machines: an overview," *The International Journal of Advanced Manufacturing Technology*, pp. 1–17, 2021.

[7] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[8] A. Shenfield and M. Howarth, "A novel deep learning model for the detection and identification of rolling element-bearing faults," *Sensors (Basel, Switzerland)*, vol. 20, 09 2020.

[9] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, "Pronostia: An experimental platform for bearings accelerated degradation tests." in *IEEE International Conference on Prognostics and Health Management, PHM'12.* IEEE Catalog Number: CPF12PHM-CDR, 2012, pp. 1–8.

[10] H. Huang and N. Baddour, "Bearing vibration data collected under time-varying rotational speed conditions," *Data in Brief*, vol. 21, pp. 1745–1749, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352340918314124

[11] L. Wyse, "Audio Spectrogram Representations for Processing with Convolutional Neural Networks," Jun. 2017.

[12] S. Hernández Sánchez, R. F. Pozo, and L. A. H. Gómez, "Driver Identification and Verification From Smartphone Accelerometers Using Deep Neural Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 1, pp. 97–109, Jan. 2022.

[13] Z. Zeng, M. G. Amin, and T. Shan, "Arm Motion Classification Using Time-Series Analysis of the Spectrogram Frequency Envelopes," *Remote Sensing*, vol. 12, no. 3, p. 454, Jan. 2020.

[14] S. Y. ŞEN and N. ÖZKURT, "ECG Arrhythmia Classification By Using Convolutional Neural Network And Spectrogram," in *2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, Oct. 2019, pp. 1–6.

[15] C. Schröer, F. Kruse, and J. M. Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Computer Science*, vol. 181, pp. 526–534, 2021.

[16] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967.

[17] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6765–6816, 2017.

[18] ——, "Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization," *Journal of Machine Learning Research*, vol. 18, 2018.

[19] S. Basodi, C. Ji, H. Zhang, and Y. Pan, "Gradient amplification: An efficient way to train deep neural networks," *Big Data Mining and Analytics*, vol. 3, no. 3, pp. 196–207, Sep. 2020.