Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 6588–6593

A hybrid deep learning-based approach for rolling bearing fault prognostics ⋆

Domício Neto * Lorena Petrella * Jorge Henriques * Paulo Gil *, Alberto Cardoso ***

* *University of Coimbra, Department of Informatics Engineering, Centre for Informatics and Systems of the University of Coimbra (CISUC), Coimbra, Portugal (e-mails: dneto@student.dei.uc.pt, lorenapetrella@dei.uc.pt, jh@dei.uc.pt, alberto@dei.uc.pt).*

** *CTS-UNINOVA, Department of Electrical and Computer Engineering, NOVA School of Science and Technology, Caparica, Portugal (e-mail: psg@fct.unl.pt).*

Abstract: Predictive Maintenance (PdM) has the potential to revolutionize the industry by providing advanced techniques to assess the condition of an industrial system and yield key information that can help optimize maintenance planning and prevent unexpected faults and breakdowns. Nevertheless, PdM is far from being universally applied and it is still the subject of increasing research. Thus, developing new approaches has great relevance to help PdM become a practical reality for the industry. PdM can also bring benefits in terms of sustainability, by reducing human and material resources waste, which is one of the main objectives of Circular Manufacturing initiatives. In this context, rolling bearings are one of the most studied components, as most industrial systems with rotating mechanisms contain bearings, which are prone to a number of faults caused by natural and unnatural wear. In this work, an hybrid Deep Learning (DL) approach is proposed, combining a Convolutional Neural Network (CNN) with a Gated Recurrent Unit (GRU) network to predict Remaining Useful Life (RUL) using rolling bearing vibration data preprocessed with the Short-Time Fourier Transform (STFT). This model was trained and validated using the PRONOSTIA public dataset, which is a popular benchmark for rolling bearing prognostics. The obtained results are satisfactory, providing RUL estimates close to the true values in most test cases, proving the competitiveness of the approach and its potential.

Keywords: Predictive Maintenance, Circular Manufacturing, Remaining Useful Life, Rolling Bearings, Deep Learning.

1. INTRODUCTION

Industrial maintenance is a key part of any industrial business. Maintaining the machines in a good working condition and avoiding breaking faults can ensure availability and also reduce the waste of human and material resources, which is important when considering the increasingly-relevant Circular Manufacturing paradigm, which seeks to reduce the negative impact of industries on the environment (Acerbi et al. (2021)). In that sense, Predictive Maintenance (PdM) is a modern approach to the maintenance subject, where data acquired from the industrial systems is used to estimate the system's health condition and predict faults. Research in PdM has been growing steadily, as technology for acquiring and processing data from the machines evolves.

One of the promising techniques that have shown state-of-the-art performance for many PdM use cases is Deep Learning (DL) (Lin et al. (2020); Serradilla et al. (2022)), which comprises several solutions that are based on multi-layered Artificial Neural Networks (ANN). Models based on DL are capable of achieving very high accuracy for fault diagnostics and prognostics problems, generally surpassing other non-deep Machine Learning (ML) solutions.

Rolling bearings are predominant in most industrial systems that comprise rotary components. Moreover, rolling bearings are the root cause of up to 44% of faults in some machines (Cerrada et al. (2018)). Bearings are susceptible to a wide variety of wear causes, such as: lubrication problems (which can lead to overheating), overloading, improper mounting and manufacturing defects (Bernet (2019)). Worn out bearings may eventually breakdown, but first, it will gradually present malfunctioning signs, such as increased vibration and heat production. Bearings are a key subject of PdM research and finding new ways to perform fault diagnostics and prognostics with good accuracy has a great significance for the evolution of industrial maintenance.

⋆ This work has been carried out in the H2020 KYKLOS 4.0 project (Grant Agreement Number 872570), which is funded by the European Commission. This work was also partially financed by national funds through the FCT - Foundation for Science and Technology, I.P., within the scope of the projects CISUC (UID/CEC/00326/2020) and CTS (UID/EEA/00066/2019), and under grant SFRH/BSAB/150268/2019.

10.1016/j.ifacol.2023.10.311

One possible method to develop an effective PdM model is to combine different models to create a hybrid approach, taking advantage of the strengths of each model. In the work of Zhao and Wang (2021), a Convolutional Neural Network (CNN) is combined with a Support Vector Regressor (SVR) to perform RUL estimation of rolling bearings. First, a Hilbert-Huang Transform (HHT) is applied to obtain a Degradation Energy Indicator (DEI). The CNN is trained to make DEI estimations, while the SVR makes RUL estimation based on the estimated DEI. Another hybrid approach is presented by Kamat et al. (2021), where an Autoencoder (AE) is used for anomaly detection, which triggers the RUL estimation, performed by a Long Short-Term Memory (LSTM) network.

Moreover, Guo et al. (2017) construct a health indicator, combining several time, frequency, and time-frequency domain features, and apply a Recurrent Neural Network (RNN) to perform the prognostics. In the work of Chen et al. (2020), the authors propose a bi-directional Gated Recurrent Unit (GRU) network with an added convolutional layer as the first layer and with an attention mechanism to extract a health indicator from the vibration data. The RUL estimation is performed using linear regression. The approach proposed by Jiang et al. (2020) is composed by a hybrid CNN-LSTM model, capable of performing RUL prediction directly from raw data.

In the context of processing rolling bearing vibration data, the Short-Time Fourier Transform (STFT) is one of the time-frequency analysis techniques that are frequently used for extracting underlying information from the raw vibration signal. In the work of Li et al. (2019), the authors propose a CNN-based model for RUL prognostics trained with data processed by the STFT and a multi-scale feature extractor, which is also a CNN. The authors of Zhou et al. (2020) present another STFT-CNN-based approach capable of performing both fault prognostics and diagnostics.

In the review presented by Serradilla et al. (2022), the author points out the RNN, LSTM, and GRU as the most commonly applied DL models for RUL prediction. This can be explained by the fact that these models are capable to learn sequential information, thereby they are appropriate to prognostics, as RUL prediction is a time-series forecasting problem.

In this work, a hybrid DL solution combining a CNN and a GRU network is proposed for rolling bearing fault prognostics. The combination of CNNs and RNN-based networks have been shown to be more effective than using a single model by itself, as the CNN's ability to extract deep abstract features from data potentiates the overall performance of sequential learning-based models, such as the LSTM and GRU (Nasser and Al-Khazraji (2022); Wahid et al. (2022)). The dataset used for training and validation is the PRONOSTIA rolling bearing vibration dataset, a popular benchmark for bearing prognostics, provided by Nectoux et al. (2012). A comparison is made between the developed approach, a selected set of ML models created for this purpose, and other approaches from the literature.

This paper starts by presenting in section 2 the methodology used for the development of the approach, describing the components of the hybrid network, the STFT preprocessing, and the postprocessing methods. Section 3 gives a detailed description of the PRONOSTIA dataset, used in this work. Section 4 presents the prognostics results, showing some graphical examples from the obtained RUL predictions and comparisons with other approaches. Finally, section 5 features a recap about the work and its results and provides suggestions for future developments.

2. METHODOLOGY

In this section, each component of the RUL estimation approach is explained in detail. An overall representation of the approach's pipeline is shown in Figure 1.

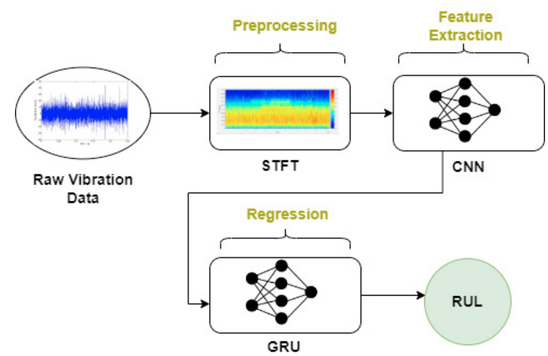

Fig. 1. Representation of RUL estimation approach's pipeline.

2.1 *Short-Time Fourier Transform (STFT)*

Vibration signals contain distinct frequency signatures for different health states. Extracting this information can improve a model's capacity to make fault diagnostics and prognostics. One popular method to perform this analysis is the STFT (Rezaeianjouybari and Shang (2020)). The STFT is based on the Discrete Fourier Transform (DFT) algorithm, as is given by equation 1, where m and ω are time and frequency indices, respectively, x is the discrete-time signal, and w is the window function.

$$\begin{aligned} STFT(m,\omega) &= X(m,\omega) \\ &= \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-i\omega n} \end{aligned} \tag{1}$$

For this work, a periodic hamming window with size equals to 128 samples is used, with a 75% overlap. The number of DFT points used is also 128.

Figure 2 shows the STFT results for one of the rolling bearings experiments from the PRONOSTIA dataset, where the bearing gradually degrades over time until it suffers a fault. It is possible to observe how the frequency components below 4 kHz and above 10 kHz increase in magnitude as the bearing approaches the breaking point. This information is processed by the CNN and GRU hybrid model as 128x13 matrices for every 512 samples of the original vibration signal.

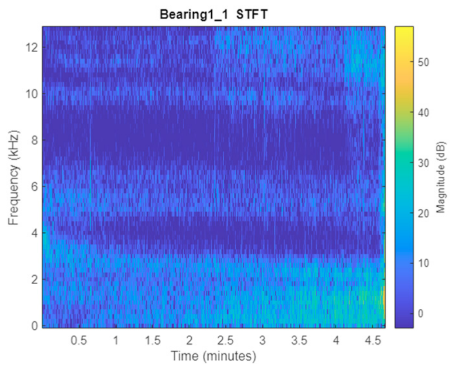

Fig. 2. STFT results visualization for one subset of the PRONOSTIA dataset.

2.2 Feature Extraction with Convolutional Neural Network

Convolutional Neural Networks (CNNs) were inspired by the brain's visual cortex. Although CNNs were initially developed for image processing tasks, where it is still the state-of-the-art technique in that subject (Guo et al. (2016); Mu and Zeng (2019)), it is now used for several different classification and regression tasks, including fault diagnostics and prognostics. Figure 3 presents a typical CNN architecture.

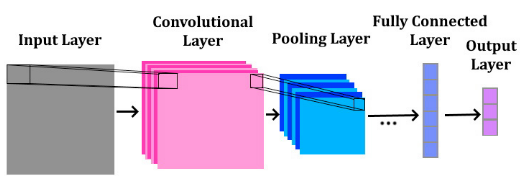

Fig. 3. Representation of the typical CNN architecture.

For a two-dimensional (2D) input, the convolutional layer performs the convolutional product between the input and the filter, which is also 2D. Assuming that the input is an image, this operation is given by the equation 2, where I is the input, K is a filter, and n_H, n_W and n_C are the image's height, width, and number of channels, respectively.

$$conv(I, K) = \sum_{i=1}^{n_H} \sum_{j=1}^{n_W} \sum_{k=1}^{n_C} K_{i,j,k} I_{x+i-1,y+j-1,k} \qquad (2)$$

A convolutional layer may have several filters, resulting in multiple outputs, one for each filter. The combination of the outputs is called the feature map. The filters' weights are updated during the training process, which ultimately influence the network's capability to extract distinct features from the input. Using the image processing analogy, this feature extraction process can be seen as the identification of certain shapes and contours.

A convolutional layer is usually followed by a pooling layer, which performs subsampling. This subsampling process helps to reduce the amount of data to be processed by the subsequent layers and to improve the network's generalization ability. The pooling operation is given by equation 3, where $a^{[l]}$ is the output calculated using input $a^{[l-1]}$, ϕ is the pooling function and f is the size of the pooling filter.

$$\begin{aligned} a^{[l]}_{x,y,z} &= pool(a^{[l-1]})_{x,y,z} \\ &= \phi^{[l]}((a^{[l-1]}_{x+i-1,y+j-1,z})_{(i,j)\in[1,2,\ldots,f^{[l]}]^2}) \end{aligned} \qquad (3)$$

The most common pooling functions are the average and max functions. The filter is applied to an area of size f and then passes to the next area, where the shift is given by the value named stride.

A CNN may include several pairs of convolutional and pooling layers and each pair are often considered to be a single layer of the CNN. Finally, the CNN contains one or more fully connected layers that map the output of the previous layers to the target output. The CNN is designed to extract deep abstract features, making it capable of performing classification/regression with raw data, without the need of previous feature extraction processes.

Considering the CNN's powerful feature extraction capability, it can be combined with other classification or regression models to create a hybrid approach. For this work, the features extracted from a CNN are used to train a GRU network that performs Remaining Useful Life (RUL) estimation.

The network architecture for this work was developed through manual tuning, consisting of 5 sets of convolutional, max pooling, and batch normalization layer. The batch normalization layer normalizes the output of the previous layer, scaling the data so it has a mean of 0 and a standard deviation close to 1. This normalization process ensures a faster training process. The last three layers are two fully connected layers and a regression output layer, which will map the extracted features to the target output, as previously mentioned.

In terms of parameters, the first convolutional layer has 16 filters, and each successive convolutional layer doubles the number of filters from the previous one, so more deep abstract features may be extracted from the input. Each filter has a size of 3x3. The pooling layers have a filter size of 2x2 and a stride of 2, which halves the output's height and width. Finally, the first dense layer has 64 neurons, and the second layer has 1 neuron, which is the output size.

In order to use the CNN for feature extraction, the network is first trained to make RUL estimations with the PRONOSTIA rolling bearing vibration dataset. After the network is trained, the last two layers are removed, and the rest of the CNN is used to extract the features from the same dataset. It is important to note that the new transformed dataset was originated from the whole PRONOSTIA dataset, but the CNN was trained with just the training subsets, which are predetermined (section 3).

2.3 RUL Estimation with Gated Recurrent Unit (GRU)

A Gated Recurrent Unit (GRU) network was developed to train on the data extracted by the CNN and perform RUL estimation. A GRU is a variation of the LSTM unit, while the LSTM is derived from the original Recurrent Neural Network (RNN) unit (Ashraf Zargar (2021)).

As RNN suffered from the vanishing gradient problem, which limits the capacity of learning long-term dependencies, the LSTM and the GRU were proposed to overcome that obstacle. Both LSTM and GRU rely on gates that control the information flow inside the unit, while the GRU has one less gate than the LSTM, which has 3, and as such the GRU can make the training process faster and more efficient in terms of memory, considering that it has less parameters. A GRU representation is shown in Figure 4.

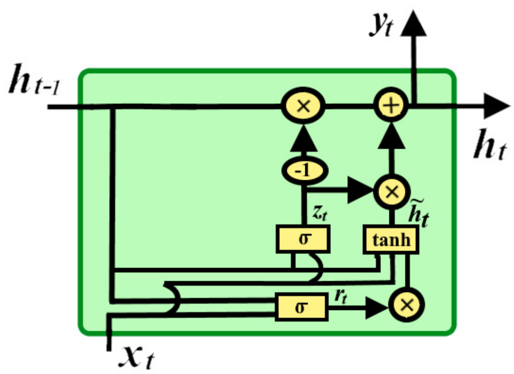

Fig. 4. Representation of a GRU unit.

The underlying GRU operations are represented by equations 4 to 6.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) + b_z \quad (4)$$
$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) + b_r \quad (5)$$
$$\tilde{h}_t = tanh(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \quad (6)$$
$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (7)$$

The main components of the unit are the update gate z and the reset gate r. The update gate determines how important the past information is to the current state. The reset gate allows the unit to eliminate past information that might be irrelevant for the future time steps. The unit returns the output y_t, which is also the hidden state h_t that is fed to the unit for the next time step.

For this work, the GRU-based network is composed of a GRU layer, with 128 units, followed by a fully connected layer with 128 neurons, followed by a dropout layer with a 50% dropout factor, and two more fully connected layers, one with 64 neurons and the last one with 1 neuron, which is the output size. The dropout layer randomly sets to 0 a percentage of a layer's output. This process helps to reduce overfitting. Once more the parameters were manually tuned.

2.4 RUL Postprocessing

Two postprocessing operations are performed after RUL estimation. First, a moving average smoothing is applied to reduce the effect of the erratic behavior of the RUL estimation. A sliding window of size equals to 513 is used and is represented by equation 8, where $\hat{y}_s[n]$ is the RUL estimation for a given time step and w is the window size.

$$\hat{y}_s[n] = \frac{\hat{y}[n-256] + ... + \hat{y}[n] + ... + \hat{y}[n+256]}{|w|} \quad (8)$$

The second postprocessing operation is applied following the first estimated RUL equals to 0, which marks the end of the estimation process. The true RUL 0 marks the end of the experiment for the given subset, where the bearing reached a critical point of failure. With this post-processing, the first predicted 0 points to the estimated time for the bearing's breakdown. If the estimated RUL never reaches 0 it means that the estimations were late to predict the breakdown, but if the error is small, the estimation can still be valuable, considering that a threshold-triggered alarm would normally be set with considerable advance in relation to the breakdown point.

3. DATASET DESCRIPTION

The PRONOSTIA dataset contains rolling bearing vibration data from 17 run-to-failure experiments. The bearings were not seeded with faults and were allowed to degrade naturally over time. Vibration data was collected using accelerometers installed in the rolling bearings coupled to the AC motor.

Three operation regimes were defined for the experiments, in terms of revolutions per minute (rpm) and load (N):

- Operation regime 1: 1800 rpm and 4000 N.
- Operation regime 2: 1650 rpm and 4200 N.
- Operation regime 3: 1500 rpm and 5000 N.

The 17 subsets were already divided by the authors into training and testing. This division by training, testing, and operation regime is shown in Table 1.

Table 1. PRONOSTIA dataset description.

	Operation Regime		
	Op. Regime 1	Op. Regime 2	Op. Regime 3
Training set	Bearing1_1	Bearing2_1	Bearing3_1
	Bearing1_2	Bearing2_2	Bearing3_2
Test set	Bearing1_3	Bearing2_3	Bearing3_3
	Bearing1_4	Bearing2_4	
	Bearing1_5	Bearing2_5	
	Bearing1_6	Bearing2_6	
	Bearing1_7	Bearing2_7	

Data acquisition was performed with a sampling frequency of 25.6 kHz, which favors feature extraction techniques, especially frequency-related ones, as there is more information to process. The duration of each subset depends solely on the fault onset, as such, it varies widely between less than a minute and more than 5 minutes. There is no information on what type of fault occurs, making this dataset to be a benchmark solely for prognostics algorithms. For the development of this approach, the RUL values were scaled to the interval [0, 1].

4. RESULTS

For assessing the model's performance, the Root Mean Squared Error (RMSE) is used, once it is one of the most common metrics for RUL estimation. The RMSE is given by equation 9, where y_i is the true RUL and $\hat{y}_i$ is the estimated RUL for the time step i.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (9)$$
$$RMSE = \sqrt{MSE}$$

Table 2 presents the test RMSE values for the proposed approach (CNN-GRU) and other models used for comparison, which includes the standalone CNN used for RUL estimation, a Deep Neural Network (DNN) with 5 layers, and finally a LSTM trained with raw data, which is featured in Neto et al. (2022).

Table 2. Test results for the PRONOSTIA dataset.

Model	Preprocessing method	RMSE
CNN-GRU	STFT	**0.191**
CNN		0.267
DNN		0.557
LSTM	None	0.254

The proposed method outperforms all the other compared methods by a significant margin, proving the improved prognostic capacity of the hybrid model. It is difficult to perform comparisons with other published approaches, as each study usually choose a different train test split than the one proposed by Nectoux et al. (2012). Those that use the split proposed by the author also use the RUL percent error metric, defined by equation 10. This metric is equivalent to the Mean Absolute Percentage Error (MAPE).

$$\%Er_i = 100 \times \frac{|ActRUL_i - PredRUL_i|}{ActRUL_i} \qquad (10)$$

Table 3 compares the proposed model with other literature approaches that used the MAPE to evaluate the prognostics performance.

Table 3. Comparison with other approaches from the literature.

Model	MAPE	Reference
CNN-GRU	42.34	
Bi-directional GRU	44.49	Chen et al. (2020)
RNN	32.48	Guo et al. (2017)

Analyzing the comparison, the proposed approach outperformed one of the compared methods and was surpassed by the RNN proposed by Guo et al. (2017). Nevertheless, the results are competitive and encourage further modifications to improve the model's performance in order to match the state-of-the-art.

Additionally, a comparison of the RMSE and MAPE between different operation regimes can be performed. Table 4 presents these results for the test sets of the three operation regimes for the proposed approach.

The best results were obtained for the first op. regime. The model achieved poorer performance for the two remaining op. regimes. It is important to note that the last regime only has one test set, while the first two have 5 sets each. One possible explanation is the difference in the load for each regime (4000 N, 4200 N, and 5000 N for op. regimes 1 to 3, respectively). As the load increases, the rolling bearing will degrade more rapidly, making abrupt failures more likely, which can be confirmed by visually inspecting the data and observing that the first op. regime's vibration data show a more well-behaved gradual degradation, as opposed to the remaining regimes, that contain sudden spikes and aggravation of vibration, leading to the bearing's breakdown.

Table 4. Comparison of results between the different operation regimes, for the CNN-GRU model.

Op. Regime	RMSE	MAPE
1	0.143	0.262
2	0.316	0.288
3	0.314	0.321

Figures 5 to 6 show the results for one training subset and two test subsets.

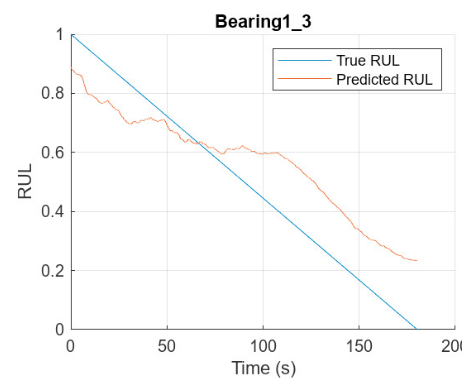

Fig. 5. Bearing1_3 test subset RUL predictions.

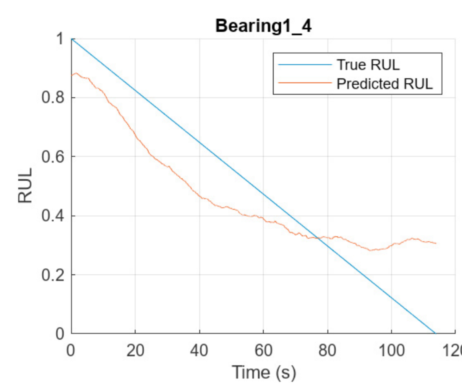

Fig. 6. Bearing1_4 test subset RUL predictions.

By visually inspecting the RUL predictions, it can be concluded that the model was able to estimate the overall degradation trend of the bearings, with a considerably small varying error degree. In practical terms, a degradation threshold can be set to activate an alarm that warns that the machine has reached a critical degradation level. A 0.2 RUL threshold would prevent breakdown in 45% (5 out of 11) of the test cases and a 0.4 threshold would prevent 72% (8 out of 11).

The results are satisfactory, considering the difficulty involved with this dataset, where there are several different degradation patterns, including abrupt faults, meaning that relating these varying degradation scenarios to a linear RUL is a challenging task.

5. CONCLUSION

A hybrid CNN-GRU model is proposed for rolling bearing fault prognostics, using the PRONOSTIA bearing prognostics benchmark to train and validate the approach. The CNN model was used for feature extraction of the STFT-processed dataset, which then were used to train the GRU model for RUL estimation. The proposed model was compared with several different ML approaches, which pointed to a satisfactory and competitive prognostic performance.

For future work, a more in-depth hyperparameter optimization process can be applied to the network, such as the Bayesian Optimization algorithm (Wu et al. (2019)), as the network's performance is sensitive to the parameter change, which can provide a significant improvement compared to the model obtained with manual tuning in this work. Another possibility is to develop an intermediate health indicator that can improve the prognostic performance. This technique is based on the extraction of a univariate condition indicator that is then mapped to the RUL and is present in most of the state-of-the-art approaches. Finally, it would be of great interest to make use of Transfer Learning to apply the developed model to real-world use cases, without the need to collect large amounts of data, as the model would already be equipped with the knowledge obtained from past training, and would "transfer" it to the new use case. Ultimately, this can be advantageous to develop new PdM solutions for different industrial machines that employ rolling bearings.

REFERENCES

Acerbi, F., Sassanelli, C., Terzi, S., and Taisch, M. (2021). A systematic literature review on data and information required for circular manufacturing strategies adoption. *Sustainability (Switzerland)*, 13(4), 1–27.

Ashraf Zargar, S. (2021). Introduction to Sequence Learning Models: RNN, LSTM, GRU. (April).

Bernet, J. (2019). Major causes of roller bearing failure and prevention tools. https://www.piprocessinstrumentation.com/bearings-seals/article/15564253/major-causes-of-roller-bearing-failure-and-prevention-tools. Visited in March 2023.

Cerrada, M., Sánchez, R.V., Li, C., Pacheco, F., Cabrera, D., Valente de Oliveira, J., and Vásquez, R.E. (2018). A review on data-driven fault severity assessment in rolling bearings. *Mechanical Systems and Signal Processing*, 99, 169–196.

Chen, Y., Peng, G., Zhu, Z., and Li, S. (2020). A novel deep learning method based on attention mechanism for bearing remaining useful life prediction. *Applied Soft Computing Journal*, 86, 105919.

Guo, L., Li, N., Jia, F., Lei, Y., and Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, 240, 98–109.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., and Lew, M.S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187, 27–48.

Jiang, J.R., Lee, J.E., and Zeng, Y.M. (2020). Time series multiple channel convolutional neural network with attention-based long short- term memory for predicting bearing remaining useful life. *Sensors (Switzerland)*, 20(1).

Kamat, P.V., Sugandhi, R., and Kumar, S. (2021). Deep learning-based anomaly-onset aware remaining useful life estimation of bearings. *PeerJ Computer Science*, 7, e795.

Li, X., Zhang, W., and Ding, Q. (2019). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliability Engineering and System Safety*, 182(October 2018), 208–218.

Lin, S., Zhang, G., Chee, H., Leow, A., Zhao, S., Kruse, D.E., Ferrara, K.W., Madarasz, F.L., and Lang, J.E. (2020). Review—deep learning methods for sensor based predictive maintenance and future perspectives for electrochemical sensors. *Journal of The Electrochemical Society*, 167, 037552.

Mu, R. and Zeng, X. (2019). A review of deep learning research. *KSII Transactions on Internet and Information Systems*, 13.

Nasser, A. and Al-Khazraji, H. (2022). A hybrid of convolutional neural network and long short-term memory network approach to predictive maintenance. *International Journal of Electrical and Computer Engineering*, 12, 721–730.

Nectoux, P., Gouriveau, R., Medjaher, K., Ramasso, E., Chebel-Morello, B., Zerhouni, N., and Varnier, C. (2012). PRONOSTIA : An experimental platform for bearings accelerated degradation tests. *IEEE International Conference on Prognostics and Health Management, PHM'12*, (June), 1–8.

Neto, D., Henriques, J., Gil, P., Teixeira, C., and Cardoso, A. (2022). A Deep Learning Approach for Data-Driven Predictive Maintenance of Rolling Bearings. In L. Brito Palma, R. Neves-Silva, and L. Gomes (eds.), *CONTROLO 2022*, 587–598. Springer International Publishing, Cham.

Rezaeianjouybari, B. and Shang, Y. (2020). Deep learning for prognostics and health management: State of the art, challenges, and opportunities. *Measurement: Journal of the International Measurement Confederation*, 163, 107929.

Serradilla, O., Zugasti, E., Rodriguez, J., and Zurutuza, U. (2022). Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. *Applied Intelligence 2022 52:10*, 52, 10934–10964.

Wahid, A., Breslin, J.G., and Intizar, M.A. (2022). Prediction of machine failure in industry 4.0: A hybrid cnn-lstm framework. *Applied Sciences (Switzerland)*, 12.

Wu, J., Chen, X.Y., Zhang, H., Xiong, L.D., Lei, H., and Deng, S.H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, 17(1), 26–40.

Zhao, Y. and Wang, Y. (2021). A Deep Learning-based Remaining Useful Life Prediction Approach for Engineering Systems. *Chinese Control Conference, CCC*, 2021-July(3), 6249–6254.

Zhou, S., Xiao, M., Bartos, P., Filip, M., and Geng, G. (2020). Remaining Useful Life Prediction and Fault Diagnosis of Rolling Bearings Based on Short-Time Fourier Transform and Convolutional Neural Network. *Shock and Vibration*, 2020.