Predicting Remaining Useful Life During the Healthy Stage in Rolling Bearings

Sebastián Echeverri Restrepo[1,2], Sébastien Blachère[1], Simón Tamayo Giraldo[3], Daniel Pino Muñoz[4], and Cees Taal[1]

[1] *SKF Research & Technology Development (RTD), SKF B.V., Houten, The Netherlands*
sebastian.echeverri.restrepo@skf.com

[2] *Department of Physics, King's College London, London, United Kingdom*

[3] *Centre for Robotics, Mines ParisTech, PSL Research University, Paris, France*

[4] *Centre de Mise en Forme des Materiaux, Mines ParisTech, PSL Research University, Sophia-Antipolis, France*

ABSTRACT

Quantitative predictions of the time before the spall initiation phase (origination of the first spall) in pristine ball bearings running under an applied load is of great industrial relevance, especially for systems that require high running accuracy and/or high-speed performance. *Currently there are no available methodologies to predict the remaining life until the first spalling event exclusively from vibration signals.* We present an end-to-end approach, based on deep learning (one-dimensional convolutional layers combined with long short-term memory units), that is able to quantify the time before the origination of the first spall in ball bearings, having as sole input vibration measurements. The method has been validated on a set of bearings – run to failure on independent but identical test rigs – which had not been considered during training.

1. INTRODUCTION

Super-precision bearings are bearings produced to have a higher standard precision level, smaller dimensional deviations, and higher rotation accuracy. The requirements for their surface shape and quality are stricter, and they often use materials with superior properties. They are specially designed for precision applications including metal cutting machinery spindles, precision ball screws, high-speed turbochargers, machine components for the semiconductor industry, and more.

For most bearing applications, the appearance of the first spall does not necessarily indicate the failure of the bearing. Bearings can still function with satisfactory performance for a considerable duration until the spall grows to an unacceptable size. This is not the case for precision applications, where the smallest spall in a bearing can render the functioning of the containing application inadmissible.

Current methodologies for assessing bearing reliability follow three main approaches. The first approach uses statistics-based methods to determine the expected life of a bearing population (Lundberg & Palmgren, 1947). The second approach uses more advanced non-destructive methodologies, such as acoustic emissions (Lees, Quiney, Ganji, & Murray, 2011; Hase, 2020), for measuring the state of the subsurface of the bearing. The final approach uses deterministic modelling to study the evolution of the microstructure and/or the initiation and propagation of cracks in model systems subjected to rolling contact fatigue (Jalalahmadi & Sadeghi, 2010; Fu & Rivera-Díaz-del Castillo, 2018; Warhadpande, Sadeghi, & Evans, 2013; Mahdavi, Poulios, Kadin, & Niordson, 2022; Ringsberg, 2001; Cheng, Cheng, Mura, & Keer, 1994).

The three mentioned approaches have some limitations. Statistics-based methods work only for bearing populations, so they are not able to predict the failure of individual bearings. Acoustic emission requires equipment that is not yet accessible for mass implementation in regular bearing applications. Modelling methodologies are usually developed to understand failure mechanisms, help with design guidelines, and, sometimes, produce life approximations for a given set of working conditions; but not for predicting failure of individual bearings during use.

There is currently no method available in the literature capable of predicting the time to emergence of the first spall of an individual bearing exclusively from vibration signals. One of the reasons is that it *was* still not clear whether vibra-

tion signals contain information about physical phenomena occurring in the subsurface of bearings. The methodology presented in this article is the first one able to quantitatively predict the time to subsurface-initiated spalling. We note here that while there are some methods in the literature that aim to perform prognostics on individual bearings, they focus only on the degradation stage, which begins after the occurrence of the first spall (Tobon-Mejia, Medjaher, Zerhouni, & Tripot, 2012; Medjaher, Tobon-Mejia, & Zerhouni, 2012; Cui, Wang, Xu, Jiang, & Zhou, 2019; Habbouche, Benkedjouh, & Zerhouni, 2021; Lourari, Benkedjouh, El Yousfi, & Soualhi, 2024; Lu, Wang, Zhang, & Gu, 2024; Magadán, Granda, & Suárez, 2024).

In the present work, we introduce an end-to-end deep learning model to predict the remaining time to the origination of a first spall in bearing tests. We will refer to the time to the origination of the first spall as the Remaining Useful Life (RUL). The model is fed exclusively with the raw vibration signal (no feature engineering is performed), and its single output is the RUL. We train and validate our model on a *set* of similarly designed test rigs and bearings rather than just one test rig to avoid any possibility for data leakage and ensure good generalisation (Hendriks, Dumond, & Knox, 2022). In this manner, we ensure that the model does not overfit to a specific test rig or bearing.

The remainder of this article is organised as follows: In section 2, we introduce the convolutional/recurrent architecture of the neural network used for the end-to-end calculation of the RUL. Subsequently, in Section 2.2, we explain the testing methodology, and how the experimental signals were grouped. In Section 3 we detail the training procedure and showcase the performance of the network when evaluated on the validation dataset.

2. MATERIALS AND METHODS

2.1. Network architecture

The architecture of the deep learning network developed in this work is presented in figure 1. It consists of three complementary parts. First, there is a set of successive convolutional, batch normalisation, and max pooling layers. These layers are followed by a set of recurrent layers, and, at the end, a dense layer. In total, the network has $57\,077$ trainable and 422 non-trainable parameters. The Keras (Chollet & Others, 2015) deep learning framework with the Tensorflow (Abadi et al., 2016) back-end is used for the generation and training of the network, as follows:

1. **Convolutional, batch normalisation, and max pooling layers**:
   As a first step, a batch normalisation operation is applied the input signal vector. We then use one-dimensional convolution filters, (additional) batch normalisation, and max pooling layers to extract features from the raw signal and to reduce the size of the input signal. This size reduction is a crucial step necessary to train the network using reasonable resources within a reasonable amount of time.
   A sequence of five stacked convolutional, batch normalisation and max pooling layers was found to be adequate for this problem (this topology is the result of a thorough exploration). We used 25, 35, 45, 50 and 55 filters with kernel sizes of 11, 9, 7, 5 and 3 for the convolutional layers, respectively. The ReLu activation function was used for the convolutions. Each of the three max pooling layers has windows of size 3. Batch normalisation operations are applied after each convolution (before applying the activation function).

2. **Recurrent layers:**
   The output of the convolutional layers is fed to a set of recurrent layers composed of Long Short-Term Memory (LSTM) units. The idea of using recurrent layers is that they are able to capture the chronological dependence of time sequences, such as the vibration signal that is captured from the tests.
   We found that using three LSTM layers gave adequate results for this problem. The first two layers contain 30 units each, and the last one contains a single unit. The first two LSTM layers return their sequences to the next layer, while the third layer does not (in order to reduce dimensionality).

3. **Dense layer:**
   After the recurrent layers, the output is finally fed to a fully connected single rectified linear unit (ReLu) that gives the prediction of the time remaining until the next bearing failure. Having a ReLu unit at the end allows the network to produce any positive real number.

2.2. Data

The data used in this work comes from bearing endurance tests performed on our premises using standard SKF R2 test rigs (Gabelli & Morales-Espejel, 2019; Harris & Kotzalas, 2006; Wan, Amerongen, & Lankamp, 1992), see Figure 2. A set of ACBB 7209 single-row angular contact ball bearings were tested until nine sub-surface initiated failures in the inner ring were obtained. The tests were performed in different test rigs (of the same type) to prevent the model from learning test rig related features (Hendriks et al., 2022). This also ensures that the model learns the bearing's behaviour under the common properties of the class of machines used.

The tests were conducted while maintaining a constant angular velocity of 6000 rpm, a temperature of $75\,°C$ and an axial load of $50\,kN$ (corresponding to a contact pressure of $3030\,MPa$). A constant value of the viscosity ratio $\kappa = 2.7$ was fixed to keep the bearing in a mixed lubrication regime and to avoid running in a boundary lubrication regime.
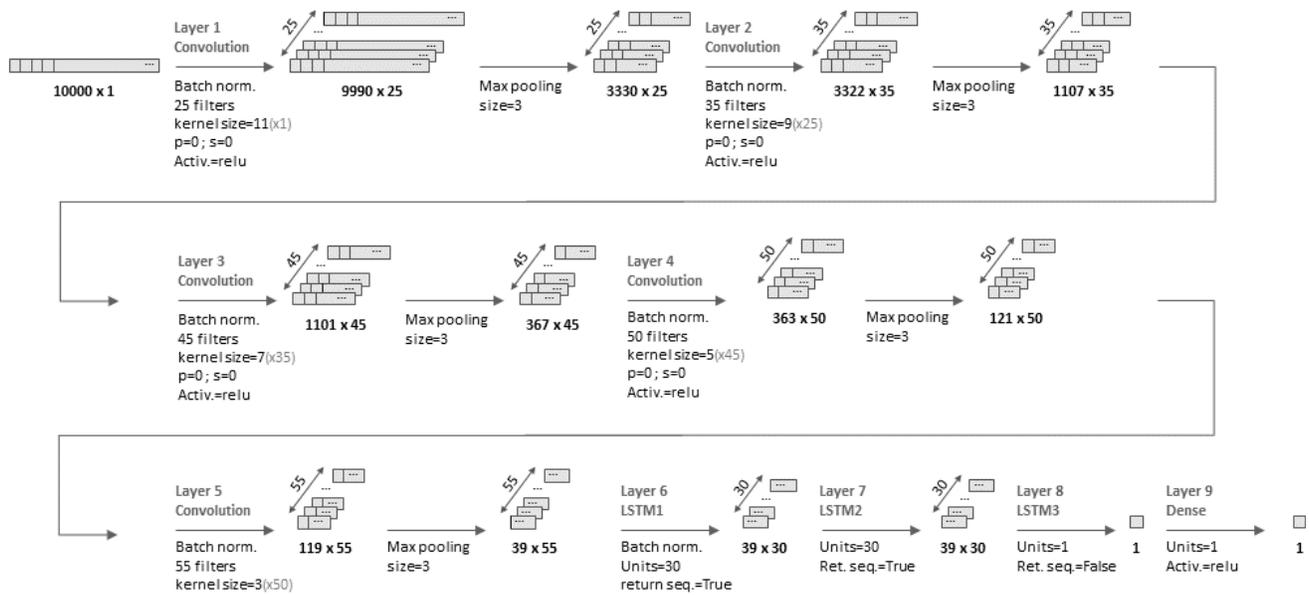
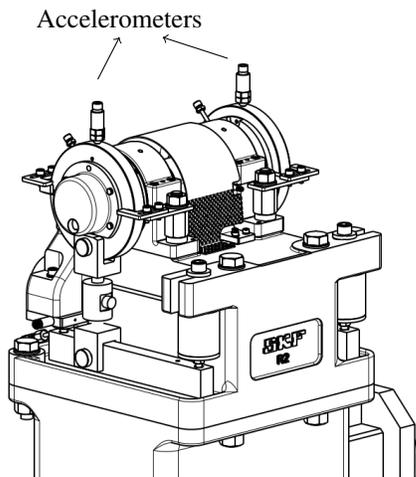Figure 1. Architecture of the convolutional/LSTM network

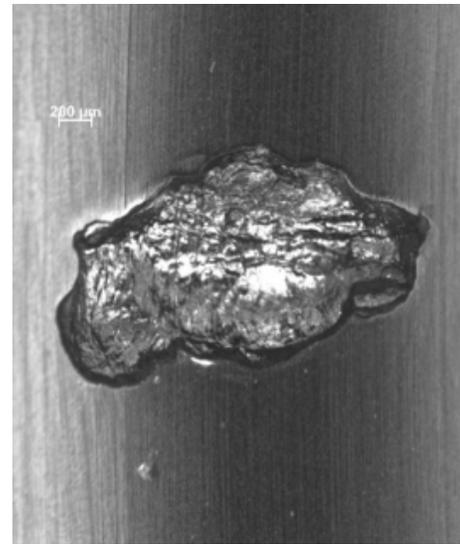Figure 2. Isometric view of an SKF-R2 test rig. The arrows indicate the position of the accelerometers.

Figure 3. Example of a typical sub-surface initiated spall used to end the endurance tests. Their origination define the moment of failure of a bearing.

Failure was defined as the moment when the first spall appears in the bearing. This event will be used as the reference to (retrospectively) measure the remaining life of each individual bearing (as detected using the vibration sensors). Note that we only selected bearings whose inner ring was the first component to fail, and only the ones that failed due to sub-surface fatigue.

The vibration data was collected using accelerometers located on the housing that contains the bearings to be tested (see Fig-

ure 2) at a sampling rate of $49\,152\,\text{kHz}$. To keep the amount of data at a manageable size, only snapshots of a duration of $6\,\text{s}$ are stored every hour.

We use vibration data from ten independent bearing tests to avoid picking up test rig-specific signal changes that might not have a physical relation to the defect (Liefstingh, Taal, Echeverri Restrepo, & Azarfar, 2021). We use a classical division of such data into a training set, a validation set, and a test set. To avoid capturing features belonging to individual setups, all the tests were conducted on identical but independent test rigs (of the type SKF R2). We chose to separate the sets in the following manner: six bearings for the training set, three for the validation set, and one for the test set.

Due to the large amount of data, it is not practical (or possible on average computers) to have a static dataset with all the input signals. To solve this shortcoming, we sample the data dynamically as the network is trained. Every epoch, a batch of 1000 segments, each containing a section of the vibration signal of 10 000 points or individual measurements, is fed into the network. This corresponds to approximately 20 rotations, which is sufficient to capture the periodicity of the signal. The segment is selected in the following manner:

- A bearing is randomly selected from a uniform distribution containing the six bearings available in the training data. The idea is to avoid the network overfitting to the bearings with a longer time to failure by receiving more training samples due to unbalanced data. We want the network to be trained using the same number of segments from each bearing.

- Once the bearing is selected, a segment (i.e. 10 000 consecutive entries/points of the vibration signal) is randomly chosen together with its corresponding time to failure.

- The calculation of the validation error is performed using batches of 1000 segments extracted from the validation set following the same strategy utilised for the selection of the training samples.

3. RESULTS

3.1. Training

The parameters of the network were optimised to minimise the Mean Absolute Error (MAE) of the predictions of the time to bearing failure. The parameters of all the layers where initialised using the default Glorot (Glorot & Bengio, 2010) uniform method, except for the last (dense/output) layer, which was initialised using values of the same order of magnitude as the duration of the longest test.

The optimisation was performed using the Adam algorithm (Kingma & Ba, 2014) with a learning rate of 5.1×10^{-4}. The network was trained on an Intel Xeon Processor E5-2650 v4 CPU. The training process was left to run overnight un-
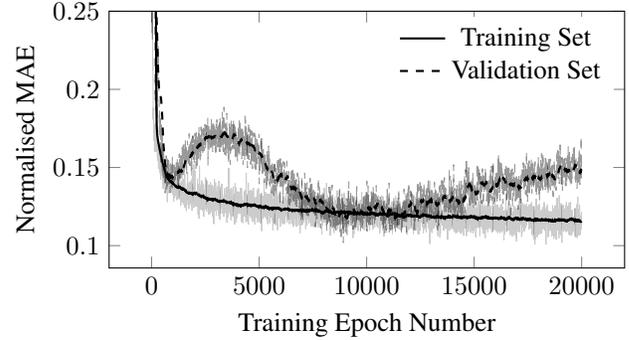

Figure 4. Convergence of the loss (MAE) for the training and validation data. The values are normalised with the total life of training bearing 6.

til completion of 20 000 epochs. Figure 4 shows the convergence of loss during training, for the training and validation data.

The set of parameters that minimises the loss (MAE) of the validation set was selected. As shown in Figure 4, such value is found after 8917 training epochs. The obtained MAE, normalised by the total life of the longest running training bearing (bearing 6), for the training set is 13%; for the validation set it is 10%. The values of the MAE are normalised due to confidentiality reasons.

For each of the training bearings we calculated the MAE of the predictions, normalised by its total life. We obtained normalised MAE values of 23%, 22%, 12%, 13%, 17%, 7% for training bearings 1, 2, 3, 4, 5 and 6, respectively.

3.2. Validation and Test

As mentioned earlier, three bearings were separated for the validation, and one for the test of the predictive power of the network. The results of the predictions of the remaining life in the validation set are presented in Figure 6, as individual points, together with the true remaining life (thin continuous lines). Similarly, the results for the test bearing are presented in Figure 7. To facilitate the visualisation of the results, we include a spline interpolation of the predictions (thick continuous line) with their corresponding standard deviation (dashed lines). Note that the plots are normalised by the total life of the bearing of the training set with the longest life (bearing 6), while the training of the algorithm was done using the raw (un-normalised) times to failure.

For each of the validation bearings we calculated the MAE of the predictions, normalised by its total life. We obtained normalised MAE values of 23%, 26% and 18% for validation bearings 1, 2 and 3, respectively. Similarly, we obtain a normalised MAE value of 25% for the test bearing.
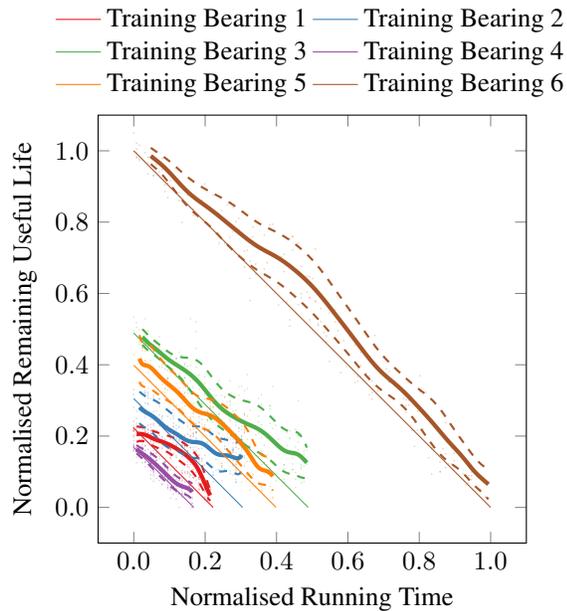
Figure 5. Remaining useful life plots for the six bearings in the training. The thick lines show the values predicted by the neural network, with their corresponding standard deviation (dashed lines). The thin lines show the true values of the remaining life. The values are relative to the total life of training bearing 6.
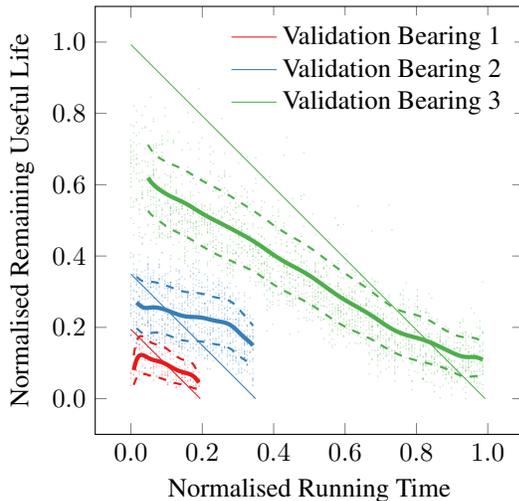

Figure 6. Remaining useful life plots for the three bearings in the validation set. The thick lines show the values predicted by the neural network, with their corresponding standard deviation (dashed lines). The thin lines show the true values of the remaining life. The values are relative to the total life of training bearing 6.
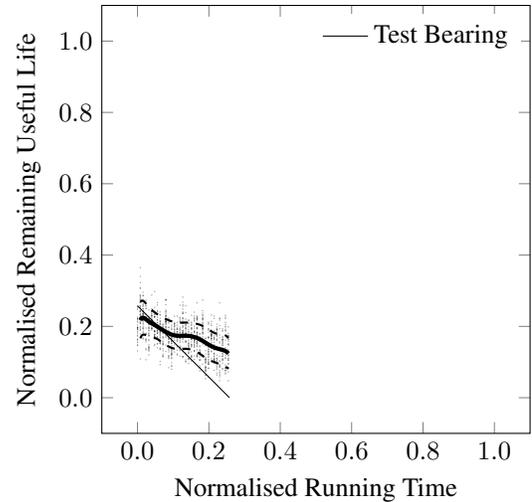

Figure 7. Remaining useful life plots for the bearing in the test set. The thick lines show the values predicted by the neural network with their corresponding standard deviation (dashed lines). The thin lines show the true values of the remaining useful life. The values are relative to the total life of training bearing 6.

4. DISCUSSION AND CONCLUSIONS

In the present work, we have developed an algorithm capable of giving a quantitative prediction of the remaining useful life (defined as the time to the origination of the first spall, or as the beginning of the spall initiation phase) for ball bearings under test conditions.

The algorithm consists of a deep neural network composed of convolutional layers and LSTM units. It receives as input a raw segment of the vibration signal produced by the bearing, and outputs the remaining time before the beginning of the spall initiation phase.

The performance of the algorithm is verified by using the classic division of the data into a training, a validation, and a test set. To avoid any data leakage between the data sets, and to exclude any test rig-specific effect, each bearing is tested on an independent test rig (of the same type). We obtain mean absolute errors relative to the time to failure of 23%, 26% and 18% for each of the three bearings used in the validation set, and an error of 25% for the bearing in the test set. The algorithm is able to give an indication of the RUL even from the beginning of the test; it is able to differentiate between bearings that will be short- an long-runners. It also captures the decreasing trend of the RUL of the bearings, with respect to running time.

The presented methodology proves that vibration signals of undamaged bearings contain information about the remaining time before the beginning of the spall initiation phase (origination of the first spall). Although identifying the fea-

ture(s) or physical phenomena that evolve over time during the bearing tests, and that are captured by the proposed deep neural network, is beyond the scope of this work, we note that there is extensive literature showing that the subsurface of a bearing is constantly evolving during rolling contact fatigue (Voskamp, 1997; Echeverri Restrepo et al., 2021). The identification of the specific information that is being used by the network to make the predictions remains a topic for future investigations.

The present work opens the door to the development of novel methodologies for early failure prediction, especially for systems that are sensitive to the effect of small defects, and that require high running accuracy and/or high-speed performance.

Since it is still not clear what physical phenomena are being captured by the algorithm to make the predictions of the remaining life, we recommend to study the process of material decay that leads to the origination of spalls in the surface. This will allow us to understand the inner workings of the deep learning algorithm.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Zheng, X. (2016). TensorFlow A system for large-scale machine learning. In *12th usenix symposium on operating systems design and implementation (osdi 16)* (pp. 265–283).

Cheng, W., Cheng, H. S., Mura, T., & Keer, L. M. (1994). Micromechanics modeling of crack initiation under contact fatigue. *Journal of Tribology*, *116*(1), 2–8. doi: 10.1115/1.2927042

Chollet, F., & Others. (2015). *Keras.* https://keras.io.

Cui, L., Wang, X., Xu, Y., Jiang, H., & Zhou, J. (2019, mar). A novel Switching Unscented Kalman Filter method for remaining useful life prediction of rolling bearing. *Measurement*, *135*, 678–684. doi: 10.1016/j.measurement.2018.12.028

Echeverri Restrepo, S., Ooi, S. W., Yan, P., Andric, P., Vegter, R. H., & Lai, J. (2021, mar). Dark etching regions under rolling contact fatigue: a review. *Materials Science and Technology*, *37*(4), 347–376. doi:

10.1080/02670836.2021.1916252

Fu, H., & Rivera-Díaz-del Castillo, P. E. (2018, aug). A unified theory for microstructural alterations in bearing steels under rolling contact fatigue. *Acta Materialia*, *155*, 43–55. doi: 10.1016/j.actamat.2018.05.056

Gabelli, A., & Morales-Espejel, G. E. (2019, mar). A model for hybrid bearing life with surface and subsurface survival. *Wear*, *422-423*(October 2018), 223–234. doi: 10.1016/j.wear.2019.01.050

Glorot, X., & Bengio, Y. (2010, jan). Understanding the difficulty of training deep feedforward neural networks. In Y. W. Teh & M. Titterington (Eds.), *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (Vol. 9, pp. 249–256). Chia Laguna Resort, Sardinia, Italy: PMLR.

Habbouche, H., Benkedjouh, T., & Zerhouni, N. (2021). Intelligent prognostics of bearings based on bidirectional long short-term memory and wavelet packet decomposition. *International Journal of Advanced Manufacturing Technology*, *114*(1-2), 145–157. doi: 10.1007/s00170-021-06814-z

Harris, T. A., & Kotzalas, M. N. (2006). *Advanced Concepts of Bearing Technology: Rolling Bearing Analysis, Fifth Edition.* CRC Press.

Hase, A. (2020, mar). Early Detection and Identification of Fatigue Damage in Thrust Ball Bearings by an Acoustic Emission Technique. *Lubricants*, *8*(3), 37. doi: 10.3390/lubricants8030037

Hendriks, J., Dumond, P., & Knox, D. A. (2022). Towards better benchmarking using the CWRU bearing fault dataset. *Mechanical Systems and Signal Processing*, *169*(November 2021), 108732. doi: 10.1016/j.ymssp.2021.108732

Jalalahmadi, B., & Sadeghi, F. (2010, apr). A Voronoi FE Fatigue Damage Model for Life Scatter in Rolling Contacts. *Journal of Tribology*, *132*(2), 1–14. doi: 10.1115/1.4001012

Kingma, D. P., & Ba, J. (2014, dec). Adam: A Method for Stochastic Optimization. , 1–15.

Lees, A. W., Quiney, Z., Ganji, A., & Murray, B. (2011). The use of acoustic emission for bearing condition monitoring. *Journal of Physics: Conference Series*, *305*(1). doi: 10.1088/1742-6596/305/1/012074

Liefstingh, M., Taal, C., Echeverri Restrepo, S., & Azarfar, A. (2021, nov). On the interpretation of deep learning models in bearing vibration diagnostics. *Annual Conference of the PHM Society*, *13*(1), 1–9. doi: 10.36001/phmconf.2021.v13i1.3047

Lourari, A. W., Benkedjouh, T., El Yousfi, B., & Soualhi, A. (2024, jan). ANFIS-based Framework for the Prediction of Bearing's Remaining Useful Life. *International Journal of Prognostics and Health Management*, *15*(1). doi: 10.36001/ijphm.2024.v15i1.3791

Lu, W., Wang, Y., Zhang, M., & Gu, J. (2024, jan).

Physics guided neural network: Remaining useful life prediction of rolling bearings using long short-term memory network through dynamic weighting of degradation process. *Engineering Applications of Artificial Intelligence*, *127*(PB), 107350. doi: 10.1016/j.engappai.2023.107350

Lundberg, G., & Palmgren, A. (1947). Dynamic capacity of rolling bearings. *Acta Polytechnica - Mechanical Engineering Series*, *1*, 4–51.

Magadán, L., Granda, J., & Suárez, F. (2024, apr). Robust prediction of remaining useful lifetime of bearings using deep learning. *Engineering Applications of Artificial Intelligence*, *130*(November 2023), 107690. doi: 10.1016/j.engappai.2023.107690

Mahdavi, H., Poulios, K., Kadin, Y., & Niordson, C. F. (2022, mar). On the effect of microplasticity on crack initiation from subsurface defects in rolling contact fatigue. *International Journal of Fatigue*, 106870. doi: 10.1016/j.ijfatigue.2022.106870

Medjaher, K., Tobon-Mejia, D. A., & Zerhouni, N. (2012, jun). Remaining Useful Life Estimation of Critical Components With Application to Bearings. *IEEE Transactions on Reliability*, *61*(2), 292–302. doi: 10.1109/TR.2012.2194175

Ringsberg, J. (2001, aug). Life prediction of rolling contact fatigue crack initiation. *International Journal of Fatigue*, *23*(7), 575–586. doi: 10.1016/S0142-1123(01)00024-X

Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., & Tripot, G. (2012, jun). A Data-Driven Failure Prognostics Method Based on Mixture of Gaussians Hidden Markov Models. *IEEE Transactions on Reliability*, *61*(2), 491–503. doi: 10.1109/TR.2012.2194177

Voskamp, A. P. (1997). *Microstructural changes during rolling rolling contact fatigue* (Doctoral Thesis). Delft University of Technology.

Wan, G. T. Y., Amerongen, E. V., & Lankamp, H. (1992, jan). Effect of extreme-pressure additives on rolling bearing fatigue life. *Journal of Physics D: Applied Physics*, *25*(1A), A147–A153. doi: 10.1088/0022-3727/25/1A/023

Warhadpande, A., Sadeghi, F., & Evans, R. D. (2013, may). Microstructural Alterations in Bearing Steels under Rolling Contact Fatigue Part 1—Historical Overview. *Tribology Transactions*, *56*(3), 349–358. doi: 10.1080/10402004.2012.754073