Article

Predictive maintenance of journal bearings: Remaining Useful Lifetime Prediction using Isolation Forest and Temporal Convolutional Networks

Sophia Bastidas [1,‡]*, Hannes Allmaier [1,‡] and Alexander Zincke [1,‡]

[1] Virtual Vehicle Research Center GmbH, Inffeldgasse 21A, 8010 Graz, Austria
* Correspondence: sophia.bastidas@v2c2.at
‡ These authors contributed equally to this work.

Abstract: Journal bearings are machine elements that are widely used in equipment ranging from vehicles to industrial machines. Their reliable operation is in many cases critical and journal bearing failure may cause subsequent high costs. This is particularly true for bearing seizure, which is a catastrophical failure of a heavily stressed journal bearing where extremely large amounts of heat are being generated. This motivates to study bearing failure and different ways to predict it from actual operation. In this work, two approaches to predict the remaining useful lifetime of journal bearings are presented. The first attempt tries to classify the lubrication conditions - the occurrence of mixed lubrication is the required precursor to seizure. For this task, an isolation forest model was established on the basis of simulation data. The results obtained with the isolation forest showed good classification performance, although not accurate enough for real-world applications. The second attempt goes significantly beyond the first approach and utilizes measured data of seized bearings to predict the actual remaining load to failure. For this approach a Temporal Convolutional Network (Temporal Convolutional Network (TCN)) was trained using a dataset consisting of several experimental seizure tests where bearings are tested in different ways until failure. The results with the TCN showed very good Remaining Useful Lifetime (RUL) prediction accuracy, being able to predict the load to failure with a very small Root Mean Squared Error (RMSE) error of only about 10%.

Keywords: journal bearing; predictive maintenance; remaining useful lifetime; seizure; isolation forest; temporal convolutional network

1. Introduction

The prediction of the RUL of any component or system aims to estimate as precisely as possible the onset of any failure, or in other terms, determine for how long such component is expected to operate normally considering both the current and past operating conditions. For journal bearings, there exist multiple failure modes, such as deformation, fatigue, corrosion, abrasive and adhesive wear, seizure, whose failure mechanisms are intensively researched [1–5]. For the application described in this paper, the focus was set on bearing seizure which is a catastrophic failure of the bearing which generally causes major damage to the machine in which it is installed. This specific type of failure is caused by the occurrence and sustain of severe mixed lubrication in combination with thermal overload. Mixed lubrication defines one of the lubrication conditions found in (e.g. grease or oil) lubricated interfaces where the lubricant film is not sufficient to fully separate the surfaces. Therefore, the involved surfaces are in direct contact to each other during operation. When this direct contact occurs, friction increases considerably, leading to greater wear and debris generation, power dissipation and localized heat generation. Mixed lubrication is a complex, non-linear process determined by the variability of the operating conditions, as well as the specific characteristics of the surface materials and lubricant properties. During normal journal bearings operation, mixed lubrication can occur, which under certain conditions can reduce overall friction losses, and does not necessarily lead to seizure [6–9];

nonetheless when these mixed lubrication conditions are sustained and progressively increased over time, local overheating can occur leading to hot spots where the materials can weld together or material transfer can occur between the bearing shell and the shaft. An example of bearing seizure can be observed in Figure 1, where the bearing shell was already starting to weld to the shaft. Unfortunately, these processes evolve very quickly and bearing seizure is only noticed when it has already caused fatal damage to the machine. Since journal bearings are a common and critical mechanical component in multiple industrial applications such as engines, turbines and power generators, a sudden failure can result in catastrophic machine shutdowns, costly repairs and production losses [1,10,11]. Therefore, accurately predicting the RUL of journal bearings, particularly in relation to seizure, has high industrial relevance as it enables the transition from traditional reactive or preventive maintenance approaches to predictive maintenance strategies. Such a transition has the potential to reduce unplanned shutdowns, extend the lifetime of components, and improve the overall safety and reliability of operations.

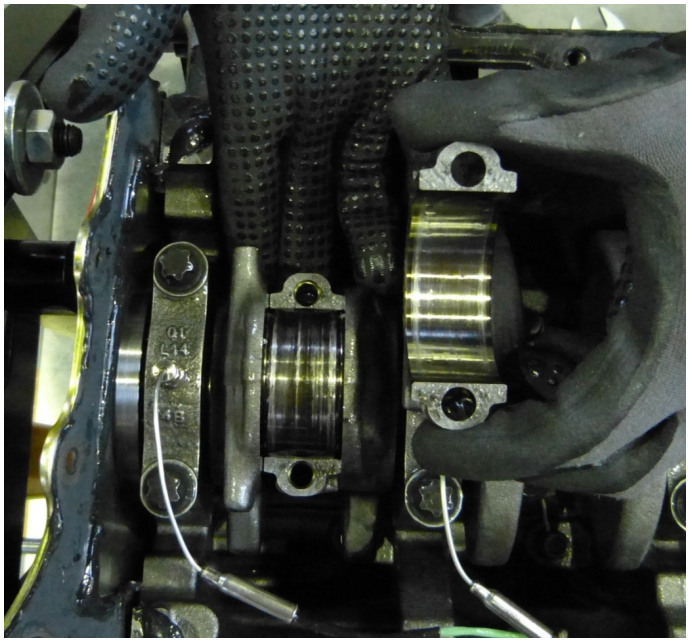


Figure 1. Journal bearing seizure: the bearing shell was starting to weld to the shaft.

The interest in machine learning and deep learning methods has been increasing recently in tribological applications, including fault diagnosis, condition monitoring and predictive maintenance [12–15]. In the context of journal bearings, some studies have been published that contribute to this field, though fewer focus on seizure failure. Moder et al. [16], for instance, evaluated the performance of different Machine Learning (ML) models, including an Artificial Neural Network (ANN), on the classification of the lubrication regime of journal bearings using experimental measurements on a bearing test-rig. It was found that both models achieved high predicting accuracy using fast Fourier transform (FFT) of the high-frequency torque measurements and that each lubrication regime presents distinctive frequencies, making FFT an effective preprocessing step to augment the data. Interestingly, a simpler logistic regression model also provided similar classification performance results. To classify the operating conditions of a self-lubricating journal bearing —steady operation to critical condition—, Prost et al. [17] proposed a semi-supervised Random Forest (RF) model trained with experimental data of lateral force sensors. The RF model showed a very good classification performance (0.939 accuracy) for unseen data. These results served as the basis for follow-up work by the same author [18], where the actual RUL of a porous journal bearing was investigated by implementing different ML algorithms and experimental measurements of torque, acceleration and wear. The experimental tests were planned to promote mixed lubrication conditions and favor accelerated wear. It was found that an ensemble model based on aggregated decision trees gave the best predicting performance, and failure was predicted up to 50 hours in advance compared to a conventional threshold-based alarm.

Using a Long Short-Term Memory (LSTM) neural network and a Fuzzy C-Means (FCM) algorithm, Ding et al. [19] proposed a model for journal bearings condition monitoring and RUL prediction. The dataset consisted of time-domain and frequency-domain features extracted from the measurements of seven sensors used during the experimental seizure tests with eight bearings. Principal Component Analysis (PCA) was used for dimensionality reduction, and the normalized first principal component (PC1) curves for each bearing were treated as degradation indicators. Ranges for the degradation indicators were pre-defined to classify the bearings' damage degree into slight, moderate, and severe seizure. In this way, at the end of the tests, it was possible to classify the bearing damage based on its final degradation indicator value. To refine this classification for the complete seizure test, an FCM algorithm was employed that classified the total bearing degradation process into four states: healthy state, slight degradation, moderate degradation, and severe degradation. Finally, for RUL prediction, authors found that a multi-stage iteration LSTM model -where the model continuously updates its predictions as degradation progresses through the different states- resulted in an improved RUL prediction accuracy and earlier failure warnings compared to the one-stage iteration approach.

As an alternative to Recurrent Neural Networks (RNNs), TCNs have been shown to outperform RNNs in many applications dealing with sequential data, such as in music and language modelling [20]. This deep learning method is based on causal and dilated convolutions, meaning that the output or prediction at a given time depends only on past information —there is no leakage of future information; additionally, the TCN architecture can deal with data sequences of any length, outputting a resulting sequence of the same length. A successful application of TCNs has been presented by Zhang et al. [21] for roller bearings RUL prediction using time-series vibration signals. In addition to the basic TCN architecture, authors included a Residual Separable Convolutional Block (RSCB) and a Soft-Threshold Temporal Convolution Block (STCB) to increase the dilation rate —there is more historical degradation information that the model can capture— while filtering out noisy information and reducing computation. This architecture was evaluated on two bearings datasets and compared to other deep learning models, demonstrating a greater prediction accuracy and the advantage of not needing expert knowledge for feature extraction as the model learned from the raw vibration signals.

Given the successful results presented for TCNs so far, and more specifically for roller bearings, it was decided to apply this architecture for RUL prediction for journal bearings. In this way, this paper presents the approach to achieving the final target of predicting the remaining useful lifetime of journal bearings using seizure experimental tests. Given the complexity of the seizure phenomena and all the processes that convey its occurrence, a simple approach was proposed, starting by evaluating the potential of a random forest ML model and simulation data to classify the lubrication conditions (hydrodynamic, mixed, and severe mixed lubrication) of a journal bearing [22]. The results demonstrated excellent classification performance, establishing a strong baseline for future work. The next step, presented in this paper, consisted of predicting the actual precursor of bearing seizure, that is, severe mixed lubrication. Here, simulation data was utilized as well, as it offers the advantage of being consistent, accurate, and without any noise issues, which is different from what is seen in real-world operations. From this step, it was found that an isolation forest model was able to identify severe lubrication for a bearing under different operating conditions with a good level of accuracy. The final step consisted of predicting the RUL; for this, simulation data was not viable, as seizure involves multiple operating conditions that are not possible to reproduce using a simulation model, such as localized temperature and tolerance variations inherent to mechanical components, their assembly and operation. The authors' experience with seizure testing has demonstrated that even when testing bearings of the same physical characteristics and materials and under the same testing settings, the measured parameters showed different results for each bearing, leading to differences in the onset of seizure. Experimental seizure data were therefore selected for predicting the RUL of journal bearings through a TCN with highly accurate results. Additionally, remaining life predictions were also presented in terms of load to failure, defined as the load that can still be applied to the bearing before seizing.

2. Materials and Methods

2.1. Lubrication regime identification using an anomaly detection ML model

An isolation forest ML model was employed to identify the lubrication regime in the journal bearing. This model is an unsupervised anomaly detection algorithm, being mixed lubrication conditions the anomalies to be detected in this application. This algorithm was chosen for the task primarily because it was designed to detect outliers in unlabeled data sets, making it suitable for applications where anomalies, mixed lubrication conditions in this study, are underrepresented. Furthermore, its implementation is relatively simple being able to handle high-dimensional data without significantly increasing computation time, which makes this algorithm the first option to try and be the baseline for further developments.

Isolation forest consists of an ensemble of multiple individual decision trees; each tree isolates the samples by randomly selecting features and thresholds until all the samples are isolated. The output of each tree is the path length, resulting from the recursive splits, measured from the tree root to the last node. Finally, the average path length of each sample is calculated from the result of all the trees in the ensemble, and the anomaly score is assigned accordingly.

Given that the final target was to identify the lubrication conditions into mixed or hydrodynamic lubrication (anomaly or not an anomaly, respectively), a logistic regression model was implemented to translate the anomaly score results to classification labels. For this, the logistic regression model was trained using the training data set, from which it learned the best decision boundary for anomaly classification.

2.1.1. Data source: bearing test-rig simulation

As mentioned in Section 1, the initial approach for predicting the RUL of journal bearings involved identifying the lubrication regime within the bearing contact throughout a working cycle and under various operating conditions. This approach utilized data generated from a dedicated bearing test rig simulation model, which development and validation has been presented in detail in [23–28] for a wide range of operating conditions, including severe mixed lubrication. The model accounts for both isothermal and elasto-hydrodynamic properties in the bearing contact, as well as the elastic deformation of mechanical components, surface contact, and the rheological properties of the lubricant oil. The simulation model is based on the Reynolds equation for non-Newtonian fluids and it includes the Greenwood and Tripp contact model [29], the Patir and Cheng flow factors [30,31] to account for rough surfaces and the Jakobsson-Floberg-Olsson approach [32] for cavitation and mass conservation. Given the importance of accurately simulating the mixed lubrication regime, an oil model was implemented based on laboratory measurements to consider the oil's piezo-viscous and non-Newtonian properties. Furthermore, the surface roughness properties and a constant boundary friction coefficient were defined.

This journal bearing simulation model was implemented in the flexible multi-body solver AVL Excite Power Unit [1]. It includes a connecting rod and three elasto-hydrodynamic bearings: one serving as the testing bearing and the other two as support bearings. All are modeled as finite element structures, while the shaft is represented as a simplified beam body. The simulation model was solved in time domain, applying numerical time integration and backward differentiation. At each time step, motion equations were calculated for all the bodies as well as contact equations using the Newton-Raphson method. The elastic deformation of the test bearing and shaft was also considered and calculated for each time step.

To identify the lubrication regime of the journal bearing through ML, data was generated by applying a periodic dynamic load downward to the test bearing at a frequency of 80 Hz. Three load conditions were simulated with a maximum value of 40, 80, and 105 kN, corresponding to 50, 100, and 130 MPa specific load, respectively. Two lubricant oils were investigated: an SAE 0W20 and an

[1] AVL List GmbH, Hans-List-Platz 1, 8020 Graz Austria, www.avl.com

SAE 5W30 oil. The shaft speed varied from 1000 to 7000 rpm in a stepwise run-up in intervals of 1000 rpm, followed by a run-down to 1000 rpm. This configuration resulted in 13 simulation runs for each applied load and oil combination. Each simulation was run for 37.5 ms with a step size of 0.1 ms, corresponding to 376 data points per simulation run. The resulting data, extracted from AVL Excite Power Unit, consisted of different performance parameters of the bearing, including the friction moment generated in the bearing shell and in the shaft, the Minimum oil film thickness (MOFT), the oil viscosity, the asperity contact friction power loss, and the asperity contact area, as well as initial operating conditions such as the shaft speed and the applied load.

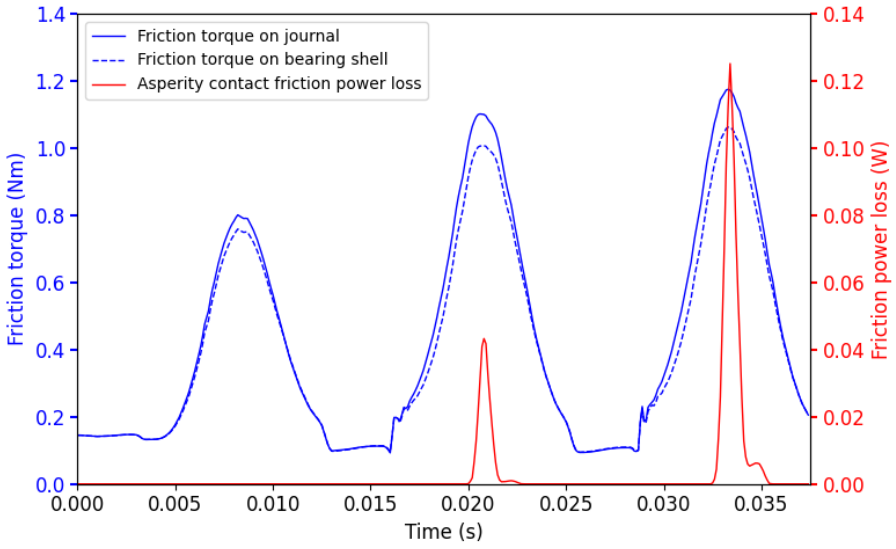
2.1.2. Data pre-process and labeling

The data pre-process and labeling consisted of multiple steps. First, all the generated data was concatenated in one data set, resulting in a matrix where the rows are the data points (at 0.1 ms intervals), and the columns are the test bearing performance parameters or features for the ML model training. Second, an additional feature was added to the data set for the Stribeck oil film parameter λ. It is defined as the minimum oil film thickness ratio to the root mean square of the combined surface roughness and was calculated from equation 1 [33]. h is the minimum oil film thickness, and the asperity surface roughness parameters $\sigma_{s,S}$ and $\sigma_{s,J}$ can be found in [24].

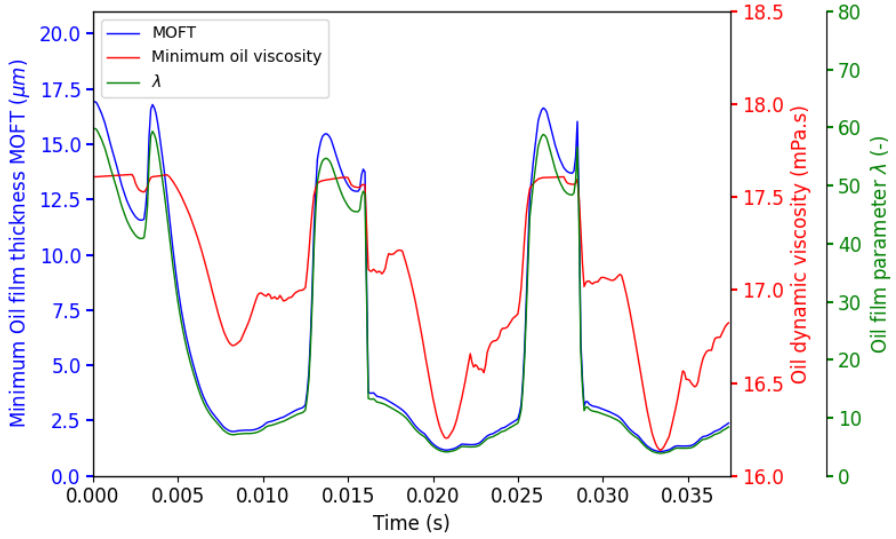
$$\lambda = \frac{h}{\sqrt{\sigma_{s,S}^2 + \sigma_{s,J}^2}} \tag{1}$$

The next step consisted of labeling the data, for this, given that the target is to identify the lubrication regime existent in the bearing contact and especially the mixed lubrication, binary labeling was established based on the normalized asperity contact area value, which is defined as the percentage ratio between the bearing area where contact between the surfaces (asperities' interaction) is present and the total area of the bearing. In this way, label 0 was assigned to asperity contact area values that are less than or equal to 0.5%, indicating hydrodynamic lubrication conditions where there is no asperity contact. Conversely, label 1 was assigned to asperity contact area values greater than 0.5%, indicating mixed lubrication conditions. These labels were determined for each data point in the data set and included as an additional labels column.

For the fourth step, feature selection was performed using Pearson correlation between all the available features and the labels. A 0.2 absolute correlation factor was set as the limit for the feature selection. From this analysis, the selected features were: the minimum oil film thickness, the minimum oil viscosity, the oil type, the applied load, the specific load on the bearing, the friction moment on the journal and on the bearing shell, the asperity contact friction power loss, and the Stribeck oil film parameter λ. Some of these features are presented in Figure 2.



(**a**)



(**b**)

Figure 2. (**a**) Simulation results used as features, part 1 (**b**) Simulation results used as features, part 2. MOFT: Minimum oil film thickness, λ: Stribeck oil film parameter.

Finally, the categorical features, oil type and test condition, were mapped to numerical values for proper utilization in the isolation forest model.

2.1.3. Isolation forest model implementation

The isolation forest algorithm was implemented with scikit-learn, which allows several hyper-parameters to be tuned. To optimize these hyper-parameters, a hold-out validation approach was used; specifically, the data set, created in the previous section, was split into three parts: the training data set (without labels), the evaluation data set for hyper-parameter tuning, and the test data set. For this, the scikit-learn function *train_test_split* was utilized with stratified sampling. This approach ensured that the proportion of anomalies (mixed lubrication data points) in the complete data set is preserved in the training, validation, and test splits. 85% of the complete data set was used for training and the remaining 15% for testing. In a second split, 20% of the data in the training data set was extracted for the evaluation data set.

The hyper-parameters considered for tuning were *max_samples* for the number of samples used to build each individual decision tree in the ensemble, *n_estimators* for the number of decision trees in the ensemble, *max_features* for the number of features selected randomly to build each individual decision tree, and *contamination* which is the proportion of anomalies expected to be identified by the model. In this application, the contamination hyper-parameter was calculated from the anomaly ratio in the complete data set. The model performance was evaluated using the validation data set using precision, recall, and F1-Measure metrics, and the best performing configuration was selected. Below are summarized the tuned hyper-parameters with the tested ranges and the final values set for the isolation forest model:

- *max_samples*: from 1 to 19 in steps of 1. Set to 14
- *n_estimators*: from 2 to 48 in steps of 2. Set to 18
- *max_features*: from 1 to 9 in steps of 1. Set to 1
- *contamination*: set to 20%

The model's final performance was evaluated with the test data set, which was not used during training or tuning, ensuring an unbiased assessment of the model's ability to distinguish between normal conditions (hydrodynamic lubrication) and anomalies (mixed lubrication). In this step, The Area Under the Receiver Operating Characteristic Curve (ROC-AUC) metric and confusion matrix, along with the previously mentioned metrics, were used to evaluate the model's performance.

2.2. Remaining Useful Lifetime (RUL) prediction

To achieve the final target of predicting the RUL of journal bearings, a specific data set is required where the performance evolution of the bearing is recorded through time until the measured conditions indicate that the bearing has failed. With the previous approach, where the bearing was subjected to a dynamic load, although it was possible to identify the exact operating point where severe mixed lubrication occurred, actual failure of the bearing was not observed. In this way, the next step consisted of using the data from experimental seizure tests performed with journal bearings. A data set was created by splitting the time series of sensor data samples into equally sized snippets. Each snippet contains measured data, including the applied load, torque, and oil temperatures, among other parameters. They were labeled with the time left before the bearing failed. In an attempt to show that the imminent failure of a bearing can be predicted using these features, a model was trained, and its predictive performance was evaluated. If it can be shown that a model's predictions are significantly better than random guessing, we expect that even better predictors than we show here are attainable.

2.2.1. Data source: bearing test-rig experimental seizure tests

To predict the RUL of journal bearings, experimental stress tests were conducted in which the bearings were subjected to increasing mechanical loads until they failed. This type of test, commonly used by bearing manufacturers, is known as seizure testing. Figure 3 is illustrative of a journal bearing test-rig and its main components. It basically consist of a test bearing, a rotating shaft and a mechanism, such as a hydraulic actuator, to apply the load. Before a bearing experiences seizure, several processes within the bearing and its surroundings must occur, leading the bearing to operate under severe lubrication conditions for an extended period. When these severe lubrication conditions persist, the lack of oil at the bearing interface causes the surfaces to weld together, resulting in a significant increase in friction torque and ultimately leading to catastrophic failure of the bearing.

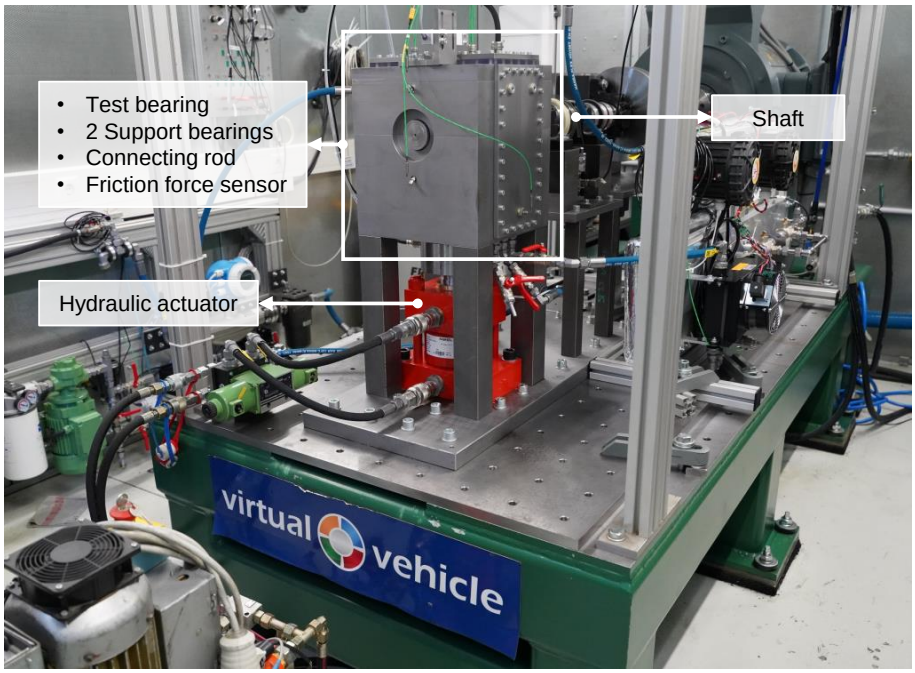


Figure 3. Test-rig for journal bearings experimental testing.

A total of twenty-three seizure tests were performed with journal bearings made from two different materials: fourteen tests used bearings of one material, while nine tests utilized bearings of the second material. Additionally, four different formulations of engine oil were utilized in the tests: one SAE 5W30, two SAE 0W20 and one SAE 0W30. For all the tests, the engine oil was supplied at a temperature of 120°C. Figure 4 summarizes the described testing protocol.

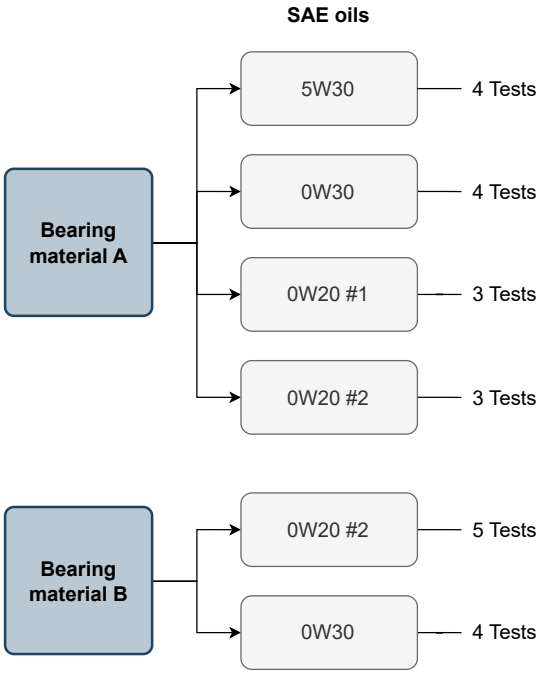


Figure 4. Testing protocol for the seizure test.

Figure 5 shows the results of one of the experimental seizure tests. In this test, as with all similar tests, seizure occurrence is clearly indicated by a sudden spike in the measured torque, reaching

nearly 25 Nm. This is followed by a rapid drop in torque to zero. Overall, from this figure, we could observe that, in addition to the direct effect of the load, the temperatures in the bearing and the surroundings are another contributing factor to bearing failure, showing a consistent increase with the load and a small spike right at the failure onset. While these results may seem straightforward, it is essential to recognize that bearing failure results from a combination of processes taking place during operation. Furthermore, factors unique to each bearing—such as slight tolerances and variations in operation—can influence the onset of failure, leading to different outcomes when conducting repeated tests with the same type of bearing.

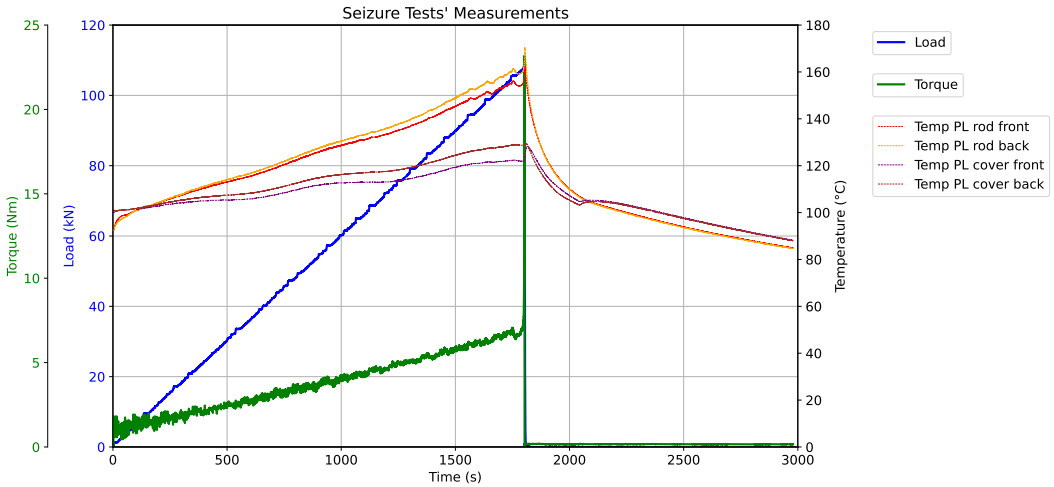


Figure 5. Measured parameters for one seizure test. PL (Prüflager) = test bearing.

Figure 5 shows the applied load, torque, and temperature measurements recorded at the test bearing. In addition to these measurements, various other operating parameters were collected during the testing process and were used for the RUL prediction. A detailed list of these parameters can be found in Table 1. Furthermore, Figure 6 shows these parameters after they have been normalized using Min-Max scaling for better visualization.

Table 1. This is a table caption. Tables should be placed in the main text near to the first time they are cited.

Parameter	Units
Load	kN
Torque	Nm
Shaft speed	rpm
Bearing surface pressure	MPa
Flow rate PL	m^3/s
Flow rate HL	m^3/s
Pressure PL	m^3/s
Pressure HL	m^3/s
Intake temperature HL	°C
Intake temperature PL	°C
Temp PL rod front	°C
Temp PL rod back	°C
Temp PL cover front	°C
Temp PL cover back	°C
Temp HL1	°C
Temp HL2	°C

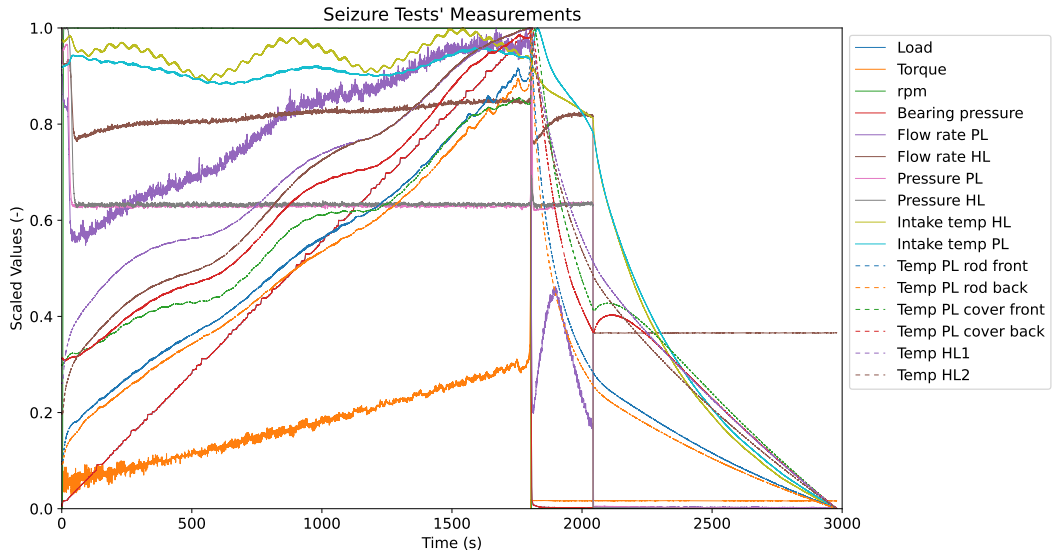


Figure 6. Example of a seizure test's features after pre-processing for the TCN model implementation.

In conclusion, the data set built for RUL prediction provides good variability in operating conditions relevant to the seizure failure process. Although it was obtained under controlled laboratory conditions, it included different bearing materials and oil formulations, along with the inherent variations due to manufacturing and installation tolerances and the operating conditions. All these factors together contributed to differences in the onset of failure. While the test protocol used, with the increasing applied load, reflects standard methods employed by bearing manufacturers to test their products, it does not reflect the broad range of operating conditions found in industrial settings, such as fluctuating loads, contamination, and lubricant starvation. As a result, while this data set offers a valuable, diverse, and controlled foundation for model development and validation, further work using extensive data sources would be needed to ensure generalizability in industrial applications.

2.2.2. Temporal Convolutional Network implementation

TCNs [20] have been shown to outperform traditional recurrent architectures in sequence modeling tasks. By employing causal and dilated convolutions, TCNs can efficiently capture long-range dependencies while maintaining low computational overhead and improved training stability. Because of recent successes[21] in predicting RUL for rolling bearings, we employed a generic implementation of TCNs to predict journal bearing's seizure. The TCN is implemented as a sequence-to-sequence model, meaning given an input sequence $X = x_0, x_1, ..., x_{t-1}, x_t$ the TCN will output a new sequence of predictions $Y = y_0, y_1, ..., y_{t-1}, y_t$. For our use-case, described in Sec. 2.2.1, we only need the final output y_t where the model output is a result of all samples in X for that snippet of data. This should mean that the causal convolutions, which are optional in the implementation we used, does not make a difference, as the function of the final output y_t has access to the entire sequence[20]. We also observed this in our experiments, hence we exclude this option in the hyper-parameter optimization experiments for the work presented here.

2.2.3. Bayesian Hyper-parameter Optimization

The generic TCN implementation we chose has many tweakable parameters, such as number of channels in each residual block, kernel size, types of activation functions, dropout percentages, and more. To fairly evaluate the capabilities of the generic TCN, we performed a hyper-parameter search using Bayesian optimization with inequality constraints[34] over a selection of these parameters. In the following we will go over some of the values and how they were parsed into hyper-parameters for the model. The full description of the parameters can be found in the Appendix 5.1. We chose relatively broad intervals for the search, to avoid biasing the hyper-parameter optimization with any preexisting assumptions. Since we trained the model directly on the original, unnormalized data, we limited the

types of available activation functions to ReLU($f(x) = max(0, x)$) and its soft variants, as we needed the model to be able to output values far exceeding 1, which is not possible with activation functions like Sigmoid or Tanh, as these are constrained to $(0, 1)$ and $(-1, 1)$ respectively. Since the data is unnormalized, we added the available model-internal normalization options to the hyper-parameter search. This is implemented as a range from 0 to 3, the result is rounded to the next integer, which are then used as indexes for the list of the options Batch Norm, Layer Norm, Weight Norm and no normalization. As the *num_channels* parameter represents both depth and layer sizes, we split these options into two hyper-parameters as follows. Firstly, we have a *depth* parameter which represents the number of residual TCN blocks, an integer in the range of 1 to 5. Secondly we have the number of channels for each block. We rely on the findings of [35], that show doubling the number of filters with each layer as a sensible default structure for a deep CNN. Using this information we decided to add a range from 0 to 8 to the optimization space which defines the number of channels in each block using the following equation:

$$\text{num\_channels}_i = 2^{\text{round}(\text{first\_block})+i} \tag{2}$$

Where i is the current layer number from the integers in 0 to round(depth). For the kernel dimensions we similarly have an integer range in 1 to total_num_features] and we round the real values of the optimizer to the nearest integer and square it to get a kernel of shape: round$(x) \times$ round(x). Dropout is a real value between 0 and 1, so we can use the result of the constrained optimization as is. For the binary decisions that toggle features on and off, we similarly search in the range from 0 to 1, round to the nearest integer and set 0 to False and 1 to True.

2.2.4. Data pre-processing and labeling

For the implementation of the TCN, the time series data needed to be pre-processed and labeled appropriately. First the actual point of seizure was identified for all tests. This identification is straightforward, as seizure is observed by a sudden and significant increase in the measured torque, followed by a rapid drop to zero. Therefore, the point of maximum torque is set as the seizure onset. In Figure 5 by instance, the seizure onset occurs at about 1800 seconds, when the torque suddenly increases from about to 7 Nm to almost 25 Nm. Following this, the time series data for each test was segmented into snippets of 10 milliseconds, stopping just before the point of failure. It is important to note that the failure or seizure point is not included in any of the snippets.

For the labeling process, the label was defined as the load to failure, which represents the additional load that is still possible to apply before the bearing seizes. At the beginning of the test, the load to failure is equal to the maximum value reached at the point of failure and decreases to zero as the test continues. Each snippet was then assigned a label using the last load to failure value present in that snippet.

Finally, the complete dataset was split randomly into training (80%) and test (20%) datasets with shuffling. For hyper-parameter tuning, a validation dataset (20%) was extracted from the training dataset in the same way.

3. Results and discussion

3.1. Lubrication regime prediction: isolation forest

Results on the performance evaluation of the isolation forest model described in Section 2.1 are presented here using the test data set. The following Figure 7 shows on the x-axis the anomaly scores predicted for the test data, and the predicted probabilities of belonging to each class (0 or 1), as determined by the logistic regression model, on the y-axis. In this plot, it can observed that negative anomaly scores, which indicate shorter path lengths in the isolation forest trees, are associated with a higher probability of being classified as anomalies. Conversely, as the anomaly score increases and exceeds the threshold of approximately -0.025, the probability of the test points being classified as normal rises. In this plot, we can observe that both probability curves transition gradually toward the

decision boundary, where the two curves intersect. This indicates that the normal and anomaly scores overlap, resulting in a range of anomaly scores where the model lacks confidence in its predictions.

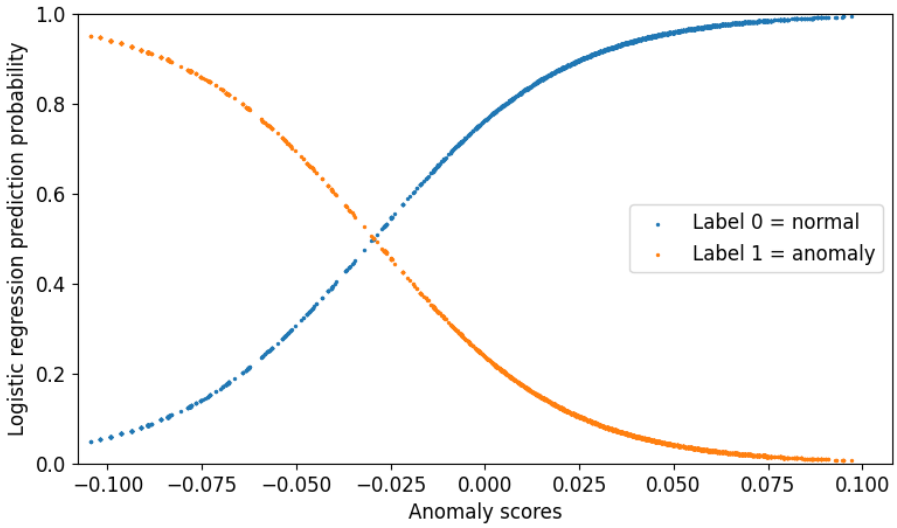


Figure 7. Anomaly score vs. prediction probability for the isolation forest model.

To evaluate the prediction performance of the isolation forest model the confusion matrix presented in Figure 8 was obtained along with the metrics: precision, recall, f1-measure and the ROC-AUC score:

- Precision: 0.99
- Recall: 0.66
- F1-score: 0.79
- ROC-AUC score: 0.83

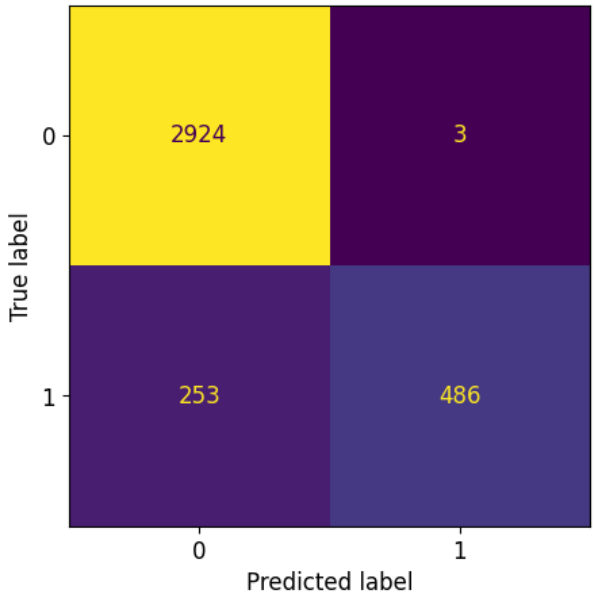


Figure 8. Confusion matrix for the isolation forest model.

From the confusion matrix and the resultant metrics, we can observe that the isolation forest model has a good performance in correctly classifying mixed lubrication when it truly occurs; this is reflected in the small number of false positives, which means that hydrodynamic lubrication conditions were minimally misclassified. From an engineering perspective, this means that false

alarms are minimized along with unnecessary maintenance interventions or shutdowns. The recall metric however, shows that the model was not able to correctly identify 34% of the mixed lubrication conditions, resulting in a significant amount of false negatives, where severe mixed lubrication was misclassified as hydrodynamic lubrication. In real-world applications, this high rate of undetected mixed lubrication conditions is unacceptable. If mixed lubrication persists or worsens, it can result in increased wear of the bearing in a short period, potentially resulting in catastrophic failure. This failure could affect not only the bearing itself but also the entire mechanical system, causing unplanned shutdowns. While the F1-score and the ROC-AUC score show that the model performs with a reasonable trade-off, an increase in recall is necessary to reduce the amount of false negatives and improve the overall model performance.

Based on the results obtained and taking into account that hyper-parameter tuning was conducted and domain knowledge was incorporated into the isolation forest model, it was concluded that the model achieved the goal of identifying mixed lubrication during bearing operation with a reasonably good level of accuracy. However, the scope of the present study is limited by the fact that the training features were obtained from a simulation model. Such data can not be directly obtained from experimental tests or in real industrial applications, which limits the generalizability of the results. Consequently, to advance this research further to ultimately predict the remaining useful lifetime of a journal bearing, it was decided to utilize an experimental data set and implement a more advanced deep learning model.

3.2. Remaining useful lifetime prediction with a Temporal Convolutional Network

After the TCN model training, it was tested with the test dataset to predict the remaining load to failure of each of the test snippets. The obtained results are presented in the flowing Figure 9. The x-axis shows the true labels or ground truth, and the y-axis shows the TCN predictions. The identity line or line of perfect predictions is represented by the black continuous line, while the standard deviation of the predictions is shown with the magenta dotted lines. The blue dots represent the true-predicted loads for each of the snippets in the test dataset. Given that each test was split into 10 ms snippets, there are about 180 points per test (the average duration of the tests was 30 minutes). To better understand this plot, it needs to be read starting from the right side, where the load needed to seize the bearing is maximum; as the applied load is increased, the one still left for failure is reduced until it reaches 0 at the moment of failure. From this plot, we can observe that the TCN model performs very well, with an RMSE of about 12 kN, which, compared to the maximum load for seizure (about 115 kN in Figure 5), is a reasonable value that allows the implementation of corrective actions on time, avoiding total failure of the bearing.

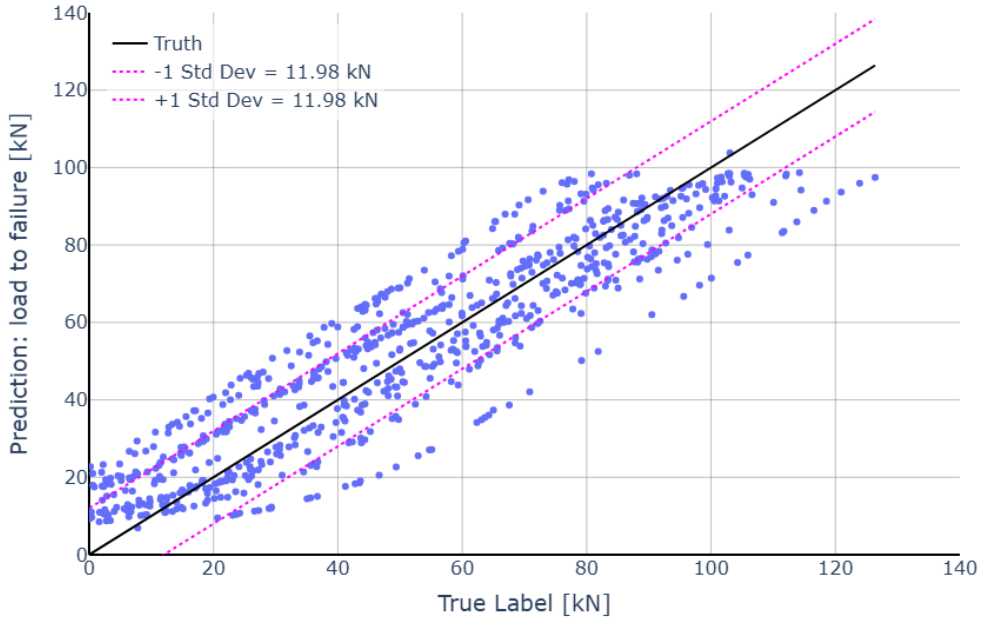


Figure 9. True and predicted load to failure values with the TCN model.

By observing the TCN prediction results in Figure 9 in more detail, it is notable how the TCN is able to predict the RUL of the bearings keeping a rather constant accuracy throughout the lifetime of the bearings. Intuitively, it was expected that the TCN would perform poorly at the beginning of the seizure tests, where the load is very small compared to the one needed for failure, and that the model accuracy would improve as the applied load increased and the bearing was closer to failure. However, this assumption was not observed in the TCN prediction results, which could be attributed to the properties of temporal convolutional networks, which leverage past time steps during training, allowing the model to incorporate historical information when evaluating the current time step.

When discussing the generalizability and potential applications of the method described here to predict the RUL of journal bearings, the specific characteristics and limitations of the method should be considered. The data set used was the result of a specific test protocol performed under controlled laboratory conditions where the load was linearly increased, and although various lubricant oils and bearing materials were tested, generalization to other operation conditions, and especially to real industrial settings, is not straightforward and would require further development and testing. In contrast, the TCN approach has a high transferability potential. It is well-suited for multi-dimensional time-series data, making it applicable to different rotating machinery that utilizes sensor signals, such as vibration and acoustic emissions. The application described here demonstrated that the TCN was able to learn temporal dependencies leading to seizure failure. This means TCNs could, in principle, generalize well to other bearing failure modes if representative data is available for training. From an industrial perspective, TCNs have a broad potential in supporting the implementation of effective predictive maintenance strategies.

4. Conclusions

- The isolation forest model demonstrated an overall good performance in classifying the lubrication regimes of a journal bearing using simulation data. As expected, the model accurately identified hydrodynamic lubrication as it is the regime with the most representation. For mixed lubrication, although the model showed good overall precision, the recall metric and the presence of a significant number of false negatives highlighted the model's limitations in identifying all instances of mixed lubrication. In real-world applications, this performance is unacceptable for reliable condition monitoring as it fails to account for all occurrences of mixed lubrication that

could indicate bearing failure, consequently increasing the risk of unplanned shutdowns and their associated costs.

- To achieve the final target of predicting the RUL of journal bearings, a TCN model was proposed and implemented using a data set of experimental stress tests consisting of time-series data where bearings were subjected to an increasing load until they failed due to seizure. The TCN model was then used to predict the load to failure, i.e. the load that can still be applied before the bearing seizes. The model showed a very good prediction performance with a small RMSE value; furthermore, the prediction accuracy was kept constant throughout the lifetime of the bearings, demonstrating the TCNs capacity to accurately raise early failure warnings in predictive maintenance strategies.
- While utilizing simulation data offers benefits such as data consistency and reduced noise, certain applications, like seizure failure, are heavily influenced by the specific characteristics of mechanical components and the operation variability. This makes simulation not viable or, at least, not sufficiently representative for its use in RUL prediction. The results presented here demonstrate the feasibility of integrating deep learning tools with experimental data, which aligns more closely with real-world applications.
- The promising results obtained with the TCN model set the basis for future research aimed at expanding the bearings RUL prediction to more realistic industrial applications. These applications often involve variable operating conditions and load applications, which introduce additional variability and noise to the data, making it more challenging to predict the onset of failure. Continued exploration of advanced ML and deep learning techniques, complemented by comprehensive datasets, is essential for further improving RUL prediction and predictive maintenance strategies.

Funding: Research leading to these results has received funding from the EU KDT Joint Undertaking under grant agreement n° 101096387 (project PowerizeD) and from the partners national programs/funding authorities. Additionally, the research received funding from SoliDAIR, a Horizon Europe project under grant agreement n° 101120276.

Research leading to these results are co-funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union Key Digital Technologies Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them. The project is supported by the Chips Joint Undertaking and its members including top-up funding by the program "KDT 2021" of the Austrian Federal Ministry for Innovation, Mobility and Infrastructure (BMIMI).The publication was written at Virtual Vehicle Research GmbH in Graz and partially funded within the COMET K2 Competence Centers for Excellent Technologies from the Austrian Federal Ministry for Innovation, Mobility and Infrastructure (BMIMI), Austrian Federal Ministry for Economy, Energy and Tourism (BMWET), the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG). The Austrian Research Promotion Agency (FFG) has been authorised for the programme management.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study will be openly available in the SoliDAIR Zenodo community: https://doi.org/10.5281/zenodo.17530066

Conflicts of Interest: The authors declare no conflicts of interest.

Acronyms

ANN Artificial Neural Network

FCM Fuzzy C-Means

FFT Fourier transform

LSTM Long Short-Term Memory

ML Machine Learning

MOFT Minimum oil film thickness

PCA Principal Component Analysis

RF Random Forest

RMSE Root Mean Squared Error

RNN Recurrent Neural Network

ROC-AUC Area Under the Receiver Operating Characteristic Curve

RSCB Residual Separable Convolutional Block

RUL Remaining Useful Lifetime

SAE Society of Automotive Engineers

STCB Soft-Threshold Temporal Convolution Block

TCN Temporal Convolutional Network

References

1. Vencl, A.; Rac, A. Diesel engine crankshaft journal bearings failures: Case study. *Engineering Failure Analysis* **2014**, *44*, 217–228. https://doi.org/https://doi.org/10.1016/j.engfailanal.2014.05.014.
2. Muzakkir, S.M.; Lijesh, P.; Harish, H. Failure mode and effect analysis of journal bearing. *International Journal of Applied Engineering Research* **2015**, *10*, 37752–37759.
3. Hase, A.; Mishina, H.; Wada, M. Fundamental study on early detection of seizure in journal bearing by using acoustic emission technique. *Wear* **2016**, *346-347*, 132–139. https://doi.org/10.1016/j.wear.2015.11.012.
4. Xu, F.; Ding, N.; Li, N.; Liu, L.; Hou, N.; et al.. A review of bearing failure Modes, mechanisms and causes. *Engineering Failure Analysis* **2023**, *152*, 107518. https://doi.org/10.1016/j.engfailanal.2023.107518.
5. Dou, Q.; Luo, H.; Song, Y.; Zhang, Z.; Zhang, J. Failure analysis of diesel engine connecting rod Big-End bearing wear considering coupled clearance lubrication joints. *Engineering Failure Analysis* **2025**, *169*, 109136. https://doi.org/10.1016/j.engfailanal.2024.109136.
6. Krithivasan, R.; Khonsari, M.M. Thermally Induced Seizure in Journal Bearings During Startup and Transient Flow Disturbance. *Journal of Tribology* **2003**, *125*, 833–841. https://doi.org/10.1115/1.1538619.
7. Ligier, J.L.; Noel, B. Friction Reduction and Reliability for Engines Bearings. *Lubricants* **2015**, *3*, 569–596. https://doi.org/10.3390/lubricants3030569.
8. Guo, H.; Bao, J.; Zhang, S.; Shi, M. Experimental and Numerical Study on Mixed Lubrication Performance of Journal Bearing Considering Misalignment and Thermal Effect. *Lubricants* **2022**, *10*. https://doi.org/10.3390/lubricants10100262.
9. Ogihara, H.; Iwata, T.; Mihara, Y.; Kano, M. The effects of DLC-coated journal on improving seizure limit and reducing friction under engine oil lubrication. *International Journal of Engine Research* **2022**, *23*, 1267–1274. https://doi.org/10.1177/14680874211012934.
10. Mehdizadeh, M.; Khodabakhshi, F. An investigation into failure analysis of interfering part of a steam turbine journal bearing. *Case Studies in Engineering Failure Analysis* **2014**, *2*, 61–68. https://doi.org/10.1016/j.csefa.2014.04.001.
11. Peterka, P.; Krešák, J.; Vojtko, M.; Mantič, M. Failure analysis of the journal bearing pulley of the cargo cable way. *Engineering Failure Analysis* **2020**, *111*, 104329. https://doi.org/10.1016/j.engfailanal.2019.104329.
12. Wang, J.; Liang, Y.; Zheng, Y.; Gao, R.X.; Zhang, F. An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples. *Renewable Energy* **2020**, *145*, 642–650. https://doi.org/10.1016/j.renene.2019.06.103.
13. He, M.; Zhou, Y.; Li, Y.; Wu, G.; Tang, G. Long short-term memory network with multi-resolution singular value decomposition for prediction of bearing performance degradation. *Measurement* **2020**, *156*, 107582. https://doi.org/10.1016/j.measurement.2020.107582.
14. Bienefeld, C.; Kirchner, E.; Vogt, A.; Kacmar, M. On the Importance of Temporal Information for Remaining Useful Life Prediction of Rolling Bearings Using a Random Forest Regressor. *Lubricants* **2022**, *10*. https://doi.org/10.3390/lubricants10040067.
15. Saad, A.; Arif, S.; Liwicki, M.; Almqvist, A. Bearing Fault Detection Scheme Using Machine Learning for Condition Monitoring Applications. 04 2023. https://doi.org/10.53375/icmame.2023.137.

16. Moder, J.; Bergmann, P.; Grün, F. Lubrication regime classification of hydrodynamic journal bearings by machine learning using torque data. *Lubricants* **2018**, *6*.
17. Prost, J.; Cihak-Bayr, U.; Neacşu, I.A.; Grundtner, R.; Pirker, F.; Vorlaufer, G. Semi-supervised classification of the state of operation in self-lubricating journal bearings using a random forest classifier. *Lubricants* **2021**, *9*.
18. Prost, J.; Boidi, G.; Puhwein, A.; Varga, M.; Vorlaufer, G. Classification of operational states in porous journal bearings using a semi-supervised multi-sensor Machine Learning approach. *Tribology International* **2023**, *184*, 108464.
19. Ding, N.; Li, H.; Yin, Z.; Zhong, N.; Zhang, L. Journal bearing seizure degradation assessment and remaining useful life prediction based on long short-term memory neural network. *Measurement* **2020**, *166*, 108215. https://doi.org/10.1016/j.measurement.2020.108215.
20. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *CoRR* **2018**, *abs/1803.01271*, [1803.01271].
21. Zhang, Y.; Zhao, X. Remaining useful life prediction of bearings based on temporal convolutional networks with residual separable blocks. *Journal of the Brazilian Society of Mechanical Sciences and Engineering* **2022**, *44*, 527.
22. Bastidas, S.; Allmaier, H. Friction and Wear in Journal Bearings: Accurate Testing and Simulation with an Outlook on Predictive Maintenance with Machine Learning. In *Artificial Intelligence Annual Volume 2024*; Papakostas, G.A.; Aceves-Fernandez, M.A.; Aydin, M.E., Eds.; IntechOpen: Rijeka, 2024; chapter 3. https://doi.org/10.5772/intechopen.1003822.
23. Allmaier, H.; Priestner, C.; Sander, D.; Reich, F. Friction in Automotive Engines. In *Tribology in Engineering*; Pihtili, H., Ed.; IntechOpen: Rijeka, 2013; chapter 9. https://doi.org/10.5772/51568.
24. Sander, D.; Allmaier, H.; Priebsch, H.; Reich, F.; Witt, M.; Füllenbach, T.; Skiadas, A.; Brouwer, L.; Schwarze, H. Impact of high pressure and shear thinning on journal bearing friction. *Tribology International* **2015**, *81*, 29–37.
25. Sander, D.; Allmaier, H.; Priebsch, H.; Reich, F.; Witt, M.; Skiadas, A.; Knaus, O. Edge loading and running-in wear in dynamically loaded journal bearings. *Tribology International* **2015**, *92*, 395–403.
26. Sander, D.; Allmaier, H.; Priebsch, H.; Witt, M.; Skiadas, A. Simulation of journal bearing friction in severe mixed lubrication - Validation and effect of surface smoothing due to running-in. *Tribology International* **2016**, *96*, 173–183.
27. Sander, D.; Allmaier, H.; Priebsch, H.H. Friction and Wear in Automotive Journal Bearings Operating in Today?s Severe Conditions. In *Advances in Tribology*; Darji, P.H., Ed.; IntechOpen: Rijeka, 2016; chapter 7. https://doi.org/10.5772/64247.
28. Allmaier, H.; Sander, D.; Priebsch, H.; Witt, M.; Füllenbach, T.; Skiadas, A. Non-Newtonian and running-in wear effects in journal bearings operating under mixed lubrication. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology* **2016**, *230*, 135–142.
29. Greenwood, J.A.; Tripp, J.H. The Contact of Two Nominally Flat Rough Surfaces. *Proceedings of the Institution of Mechanical Engineers* **1970**, *185*, 625–633.
30. Patir, N.; Cheng, H.S. An average flow model for determining effects of three-dimensional roughness on partial hydrodynamic lubrication. *ASME Journal of Lubrication Technology* **1978**, *100*, 12–17.
31. Patir, N.; Cheng, H.S. Application of average flow model to lubrication between rough sliding surfaces. *ASME Journal of Lubrication Technology* **1979**, *101*, 220–229.
32. Jakobsson, B.; Floberg, L. The finite journal bearing, considering vaporization. Gumperts Förlag, 1957.
33. Rahnejat, H. *Tribology and dynamics of engine and powertrain: fundamentals, applications and future trends*; Woodhead Publishing: Oxford [etc.], 2010.
34. Gardner, J.R.; Kusner, M.J.; Xu, Z.E.; Weinberger, K.Q.; Cunningham, J.P. Bayesian optimization with inequality constraints. In Proceedings of the ICML, 2014, Vol. 2014, pp. 937–945.
35. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *CoRR* **2015**, *abs/1512.03385*, [1512.03385].

Appendix

5.1. Bayesian optimization parameters

In the following Table 2 the ranges and possible values for each parameter optimized by Bayesian Optimization are listed; and Table 3 summarizes the parameters' values set for the TCN for RUL prediction.

Parameter	Possible Values
Hidden Activation Functions	ReLU, Leaky ReLU, ELU, GELU, SELU, SiLU
Normalization	Batch Norm, Layer Norm, Weight Norm, None
#Channels in first Block $2^{round(x)}$	$x \in [0, 8]$
#Blocks $round(x)$	$x \in [1, 5]$
Dropout	$[0, 1]$
Kernel dimensions $round(x) \times round(x)$	$x \in [1, \text{num\_features}]$
Use Gate	True, False
Use Skip Connections	True, False

Table 2. Bayesian Optimization hyper-parameter values.

These were used as hyper-parameters for the TCN implementation that can be found here: Github:paul-krug/pytorch-tcn.

Parameter	Value
Hidden Activation Functions	GELU
Normalization	Layer Norm
#Channels in first Block $2^{round(x)}$	3.77
#Blocks $round(x)$	4.42
Dropout	0.79
Kernel dimensions $round(x) \times round(x)$	4.86
Use Gate	False
Use Skip Connections	True

Table 3. Hyperparameters for the TCN model.