UiA University of Agder

Prognosis of ball bearings with non-characteristic degradation using recurrent neural network

AKSEL STRUKSNES JAHREN

SUPERVISORS
Van Khang Huynh
Jagath Sri Lal Senanayaka

University of Agder, 2020
Faculty of Engineering and Science
Department of Engineering Sciences

Master

Abstract

Breakdowns in rotary machines are often related to bearing failure. In recent years many prognostic and diagnostic models of bearings have been developed to prevent unexpected shutdowns. Estimating the remaining useful lifetime (RUL) of bearings have in many cases demonstrated to be problematic due to complex degradation mechanisms.

This thesis explores alternative pre-processing and neural network based models for RUL prediction of bearings. Predictive models are generated from six training bearings, where vibration data is sampled from start to end of life (EOL). The models are further tested on 11 test bearings for RUL predictions. Both training and test bearings are degraded in a unique manner as they are subjected to loads above the bearings' specification. The combination of limited training data and abnormal wear patterns suggest that accurate predictions of RULs are quite difficult to obtain.

Preface

Throughout my time as student at University of Agder I have gained knowledge in various interesting subjects. During my research project I was introduced to condition based monitoring and signal processing, which further peaked my interest. I learned about other methodologies, and eventually knew I wanted to combine condition based monitoring, signal processing and machine learning. Writing this thesis has been very challenging and rewarding as I prior had limited knowledge in programming and machine learning algorithms. I would like to thank my supervisors Khang Van Huynh, Jagath Sri Lal Senanayaka and my family for guidance and support.

Aksel Struksnes Jahren

Oslo, Norway, January 5th, 2020.

Declaration and Publishing Agreement

1.	I/We hereby declare that my/our report is my/our own work and that I/We have not used any other sources or have received any other help than mentioned in the report.	☒
2.	I/we further declare that this report: - has not been used for another exam at another department/university/university college in Norway or abroad; - does not refer to the work of others without it being stated; - does not refer to own previous work without it being stated; - have all the references given in the literature list; - is not a copy, duplicate or copy of another's work or manuscript.	☒
3.	I/we am/are aware that violation of the above is regarded as cheating and may result in cancellation of exams and exclusion from universities and colleges in Norway, see Universitets- og høgskoleloven §§4-7 og 4-8 og Forskrift om eksamen §§ 31.	☒
4.	I/we am/are aware that all submitted reports may be checked for plagiarism.	☒
5.	I/we am/are aware that the University of Agder will deal with all cases where there is suspicion of cheating according to the university's guidelines for dealing with cases of cheating.	☒
6.	I/we have incorporated the rules and guidelines in the use of sources and references on the library's web pages.	☒

Authorization for electronic publishing of the report.

Author(s) have copyrights of the report. This means, among other things, the exclusive right to make the work available to the general public (Åndsverkloven. §2).

All theses that fulfill the criteria will be registered and published in Brage Aura and on UiA's web pages with author's approval.

Reports that are not public or are confidential will not be published.

I hereby give the University of Agder a free right to

make the task available for electronic publishing: ☒JA ☐NEI

Is the report confidential? ☐JA ☒NEI

(confidential agreement must be completed and signed by the Head of the Department)

- If yes:

Can the report be published when the confidentiality period is over? ☐JA ☐NEI

Is the task except for public disclosure? ☐JA ☐NEI

(contains confidential information. see Offl. §13/Fvl. §13)

Contents

List of Abbreviations

BPFI	Ball pass frequency of inner race
BPFO	Ball pass frequency of outer race
BSF	Ball spin frequency
CBM	Condition based monitoring
CNN	Convolutional neural network
CWT	Continuous wavelet transform
DWT	Discrete wavelet transform
EOL	End of life
FC	Fully connected layer
FFT	Fast Fourier transform
FTF	Fundamental train frequency
FWHM	Full width half maximum
GRU	Gated recurrent unit
L_{10}-life	Minimum expected bearing lifetime
LSTM	Long short-term memory
MAE	Mean absolute error
MSE	Mean-squared error
ReLU	Rectified linear unit
RNN	Recurrent neural network
RUL	Remaining useful lifetime
STFT	Short-time Fourier transform

1. Introduction

One of the most frequent causes of rotary machine breakdowns are due to bearing failure. Bearings are important components in rotary machines as they supports axial and radial loads on shafts, while reducing friction. Even though bearings prevent friction and wear damage inflicted upon a machine, they are themselves vulnerable to degradation. Keeping bearings in a healthy state is in everyone interest as breakdowns do occur, which may lead to costly shutdowns. The traditional way of monitoring the remaining useful life (RUL) of a bearing is done by manually inspecting flakes, cracks, grease, and signs of exaggerate heat, which are indications of incipient faults. Forced shutdown of machinery is required to inspect bearings and is often not desirable or feasible. One solution to face this obstacle is by implementing condition based monitoring (CBM). The main objective of CBM is to observe and to anticipate RUL of bearings based on sampled data. CBM has the advantages of allowing continuous operation without unnecessary shutdowns of machine. However, for CBM to be efficient and valuable, accurate predictive models of degradation are required. Overprediction may lead to unfortunate bearing breakdown and permanent damage on other components, while underprediction leads to unnecessary change of bearing and maintenance. Bearings are regarded as of one the critical components in rotary machines that are frequently linked directly or indirectly to failure. The figure below display the importance of healthy bearings.

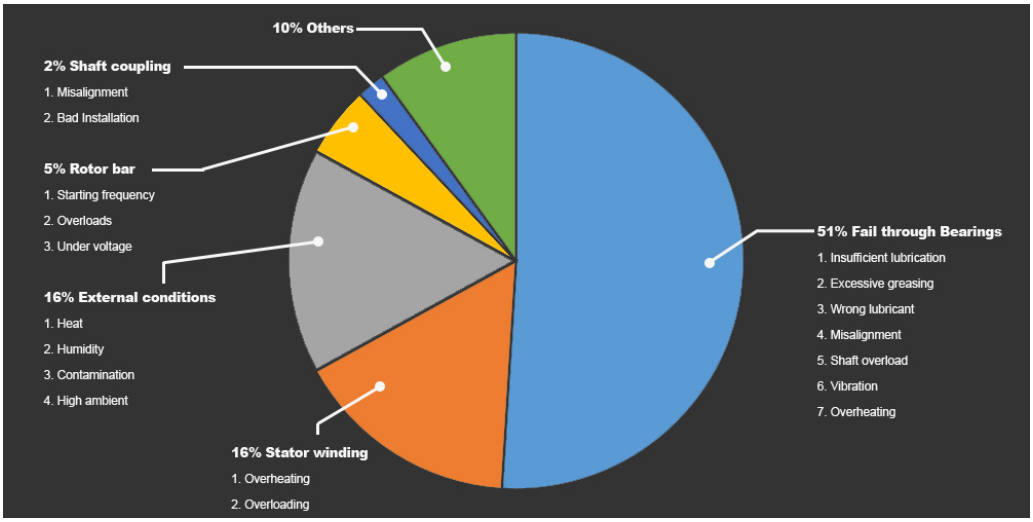


Figure 1.1: The importance of healthy bearings in machines[1]

In order to do accurate real-time prognosis and diagnosis of a bearing, data acquisition of the most characteristic parameters are crucial. Acquisition of vibration, temperature, acoustics, motors currents and lubrication debris are standard parameters used for health evaluation of ball bearings[2]. Vibration and temperature data are the most commonly used parameters for diagnosis and prognosis of bearing health. Vibration and temperature analysis are commonly used as occurring defects in bearings are often traceable in terms of increased vibrations and temperature. Typical degradation behaviour of bearings can in many cases be expressed by equations for possible characteristic vibration frequencies and their harmonics. Accelerometers and temperature sensors can normally be mounted without inconvenience on bearings, and are in many scenarios able to capture relevant data. It is also possible to capture temperature information by thermography technique which detects infrared emissions[2]. Depending on the sampled parameter, prognosis of bearings can be visualized in either time, frequency or time-frequency domain. The idea behind choosing optimal parameters and processing techniques is to effectively monitor prognosis with less uncertainty and noise. In the end CBM can be accomplished either by the use of analytical models, data-driven models or hybrid combinations of both[3]. In recent years a large number of predictive models have been made to achieve higher reliability in prognosis and diagnosis of bearings.

1.1 Introduction to the prognostic challenge

In this thesis neural network based models are used for RUL predictions of bearings in the IEEE PHM 2012 Prognostic challenge. Vibration data is acquired from six training bearings for modelling purposes. The training bearings were run from start to end of life (EOL), and were classified by one of three operational conditions. The objective of IEEE PHM 2012 Prognostic challenge is to predict the RULs of eleven test bearings ran at equivalent operational conditions as the six training bearings provided. The bearings in the prognostic challenge do not follow theoretical frameworks for degradation and L_{10}-life, most likely due to the fact that they are degraded with radial force higher than the maximum dynamic load capacity of the bearings in most operational conditions. Consequently, data-driven models based on training data have been made to describe the degradation of bearings, and to estimate RULs.

Deep neural networks have in recent years shown success and currently being used for many applications. Convolutional neural network (CNN) have revolutionized image classification, localization, segmentation tasks while recurrent neural network (RNN has shown to be important in time sequence modelling such as language translation. Even though the prognostic challenge does not provide a large amount of training data, deep learning based models are used in this thesis. There are multiple reasons for choosing deep learning based models, including personal reasons as this is a work field I find interesting.

Previous work in the prognostic challenge includes both machine-learning models and statistical models. Both the industrial and the academical winner models of the prognostic challenge used non-machine learning approaches to predict RULs on the test bearings.

The industrial winner proposed a RUL method based on smoothed accumulation of acceleration, where accumulated acceleration was weighted based on time and magnitude. The model functions were derived by fitting a few control parameters. Control parameters were optimized by cross-entropy optimization in a 2-fold cross validation process. This method achieved a total score of 0.248[4]. The academics winner used an algorithm based on anomaly detection, degradation feature extrapolation and survival time ratio estimation. The prognostic feature was determined by the average

of the five highest absolute acceleration values for each time observation. Sudden changes in the frequency spectrum were used to detect anomaly regions, predict upcoming anomaly regions, and eventually RUL based on anomaly duration ratio. Each training bearing made separate model weights, where the test bearings predictions were based from training bearings in same operation condition only. This method resulted in a total score of 0.31[5]. Neither of the winners had impressive scores, but this is likely due to how the prognostic challenge was set up. Senanayaka et. al proposed a neural network model based on stacked auto-encoders and use of LSTM RNN. Stacked auto-encoders were utilized as a feature extractor of continuous wavelet transformed vibration data. The encoded time features were then sequentially fed into a LSTM network for RUL predicitons predictions[6].

Participants in the prognostic challenge have tested different data-driven models with both time series and frequency as inputs. As to my knowledge none of the participants in the prognostic challenge has tried to improve the frequency resolution of vibration data. The reasoning behind improving the frequency resolution originates from how vibration time series are captured. Vibration data is captured in files of 0.1 seconds, which yields a frequency resolution of 10Hz. I believe higher frequency resolution is of importance to detect small frequency changes in time, and also for separating characteristic frequencies that occur in the same frequency band. This thesis focuses on evaluating different signal processing techniques, where processed vibration data is fed directly into a RNN network. Additionally, a convolutional neural

network is tested as a feature extractor prior to the RNN network.

2. Background Theory

2.1 Degradation of bearings

The degradation of a bearing is dependent on multiple variables such as running conditions, bearing quality, grease, and type of bearing. Because of the variety of factors that play part in bearing degradation, there is no generalized prognosis model that is sustainable for all types bearings and their circumstances. Running conditions such as speed and load are usually the most prominent variables that determine the RUL of a bearing[7]. Bearings are produced with specifications for both speed, axial and radial loads. Exposing a bearing for conditions outside its specifications, will normally diminish the RUL.

Lifetime of commercial bearings are usually rated according to the international L_{10}-life standard. L_{10}-life is the estimated lifetime 90% of similar bearings will last operating under a specific load and speed condition. However, L_{10}-life is a theoretical approximation as it assumes appropriate lubrication, and not degradation caused by improper storage, installation, maintenance, environmental or contamination, which cannot be controlled by the bearing manufacturers [7]. The equation for L_{10}-life in million revolution is shown below.

$$L_{10} = (\frac{C}{F})^p \tag{2.1}$$

L_10 is the estimated number of million revolutions 90% of similar bearings will last operating under a specific load and speed condition. F is the applied load. C is the dynamic load capacity of bearing. p is the bearing life exponent, which has a value of 3 for all ball bearings.

The equation for L_{10}-life in hours is shown below.

$$L_{10h} = L_{10} \cdot \frac{10^6}{Fr \cdot 3600} \tag{2.2}$$

L_{10h} is the estimated hours 90% of similar bearings will last operating under a specific load and speed condition. Fr is the relative rotation frequency between inner and outer race.

In many instances the load applied to a bearing will be dynamic. For L_{10}-life to be valid for dynamic loads, the loading must be converted to mean equivalent dynamic load [8]. L_{10}-life predicts bearing life due to classical fatigue failure, and is accountable for the production deviations. Since bearings are very sensitive to production tolerances and there is usually a significant difference between average life and L_{10}-life of a bearing. Minimal production differences in ball sizes and races can cause wear and vibrations. Uneven bearing surfaces caused by small production deviations will usually shorten the lifespan. Most bearings show increasing discontinuities on balls or races before breakdown[9]. Upcoming failure by insufficient lubrication is difficult to monitor through vibration signals. There are little to no signs of lubrication deficiency through vibration signals until metal to metal contact occurs. At that point the bearing is rapidly degrading [10].

A standard ball bearing consists of an inner race, outer race, balls, and a cage that keeps the balls in place. There exist equations for characteristic frequencies related to the mechanical components. The characteristic frequencies can be derived by bearing dimensions and the relative speed between inner and outer race. Defects such as cracks will produce repetitive impulse signals at the characteristic frequencies and harmonics. These characteristic frequencies may be seen as a part the expected degradation, hence they are commonly used for prognosis and diagnosis of bearings [2]. The figure below illustrates the basic mechanical components of a ball bearing.

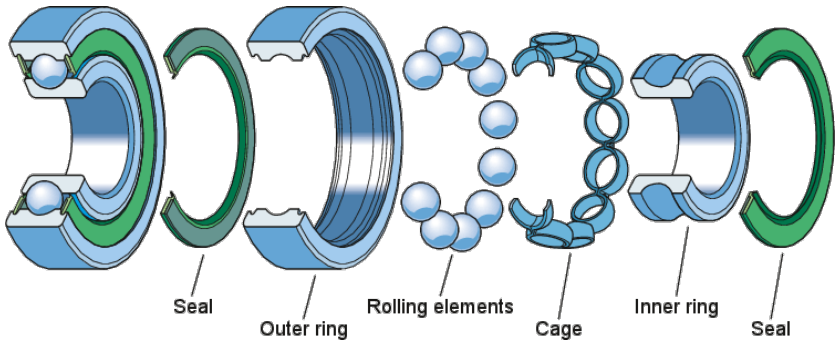


Figure 2.1: Mechanical components of a radial ball bearing [11]

The equations for the characteristic frequencies are given as:

$$\text{BPFO} = \frac{n}{2} \cdot Fr \cdot (1 - \frac{d}{D} \cdot cos\theta) \tag{2.3}$$

$$\text{BPFI} = \frac{n}{2} \cdot Fr \cdot (1 + \frac{d}{D} \cdot cos\theta) \tag{2.4}$$

$$\text{BSF} = \frac{D}{d} \cdot Fr \cdot (1 - (\frac{d}{D} \cdot cos\theta)^2) \tag{2.5}$$

$$\text{FTF} = \frac{1}{2} \cdot Fr \cdot (1 - \frac{d}{D} \cdot cos\theta) \tag{2.6}$$

BPFO is the ball pass frequency of outer race. BPFI is the ball pass of inner race. BSF is the the ball spin frequency. FTF is the fundamental train frequency or cage frequency. n is the number of ball elements in the bearing. d is the ball diameter. D is the mean bearing diameter or pitch diameter. θ is the contact angle between the ball with inner and outer race.

One thing worth noticing is the correlation between the ball diameter and the mean bearing diameter, which is relevant for all the equations[10]. The equations assume that the mechanical components have pure rolling motion. If sliding starts to occur, then the actual frequencies will start to deviate from the calculated frequency.

Healthy bearings produce relatively small vibrations, which can be difficult to distinguish from other sources. The vibration spectrum measured on a bearing may include deviations in the frequencies given by eq(2.3)-(2.6) and other external sources such as rotor unbalance, gearboxes, etc. Consequently, vibration defects generated by other components might be picked up by sensors mounted on bearings. For this reason the monitored signal may indirectly provide a health indication of the entire machine, and vibration abnormalities in the signal spectrum does not necessarily indicate a bearing defect. The possible complexity of the vibration signal may prevent equations from being effective, as the calculated frequencies are not as prominent, and can be interfered with other sources. In scenarios where the vibration signal is too complex, data-driven models may be favored over analytical models.

2.2 Data-driven models

A data-driven model will search for connections between sampled input and the desired output, based on training data. In complex systems, data-driven models have the advantages of not needing to describe the physical behaviour of the system compared to analytical based models. Data-driven models have the property of being customized to a specific set of data, whereas analytical models are preferred in simpler and more generic systems. One of the prerequisites for generating an adequate data-driven model is to have sufficient amount of relevant training data, which is often a challenge. Collecting a large amount of data can both be time consuming and costly. However in order to achieve a generalized model, the training data should extend all types of possible phenomenons. From a bearing perspective this means extending all types of degradation patterns, such as fracture in inner, outer race, etc. Inadequate training data has a probability of over-fitting, and using false correlations to predict something that is not representative of the reality.

Machine learning models learn solely based on data and are flexible in nature, whereas statistical models usually are selected based on knowledge of the problem. Statistical models are easier to interpret and often include fewer parameters than machine learning models especially deep neural networks. However, neural networks can generally have higher predictive power. The

correctness of statistical models can be evaluated using significance tests and/or confidence intervals. Evaluation of machine learning models are on the contrary done using a separate test set.

3. Datasheet

3.1 IEEE PHM 2012 Prognostic Challenge

The bearings used in the prognostic challenge are degraded by a platform called PRONOSTIA. The function of PRONOSTIA is to produce experimental bearing data without artificial defects initially. In order to degrade bearings in such manner PRONOSTIA consists of three physical parts; a rotating part, a degradation part, and a measurement setup. The rotation part is made up by an asynchronous motor, a gearbox, two shafts connected by a rigid coupling. This setup allows for different load and speed operation conditions, while the motor runs at rated conditions. The bearing of interest is mounted with a second shaft through its inner race, while the outer race remains fixed. The loading part of PRONOSTIA applies an external load forced radially on the bearing. The pressure loading is generated by a pneumatic jack and a digital electro-pneumatic regulator.

The bearing data provided by PRONOSTIA is measured by two accelerometers and one temperature sensor. The accelerometers and temperature sensors are mounted on the outer race, where the acceleration sensors are placed 90° degrees apart to capture horizontal, and vertical accelerations[12]. Accelerations are defined as vibrations, and the term vibrations will be used throughout the thesis.

The following figures display the PRONOSTIA setup.

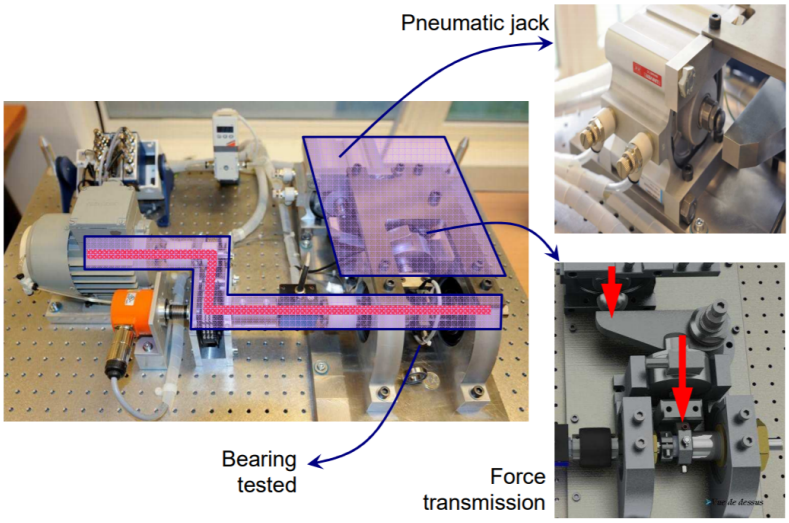


Figure 3.1: Overview of the PRONOSTIA setup [12]

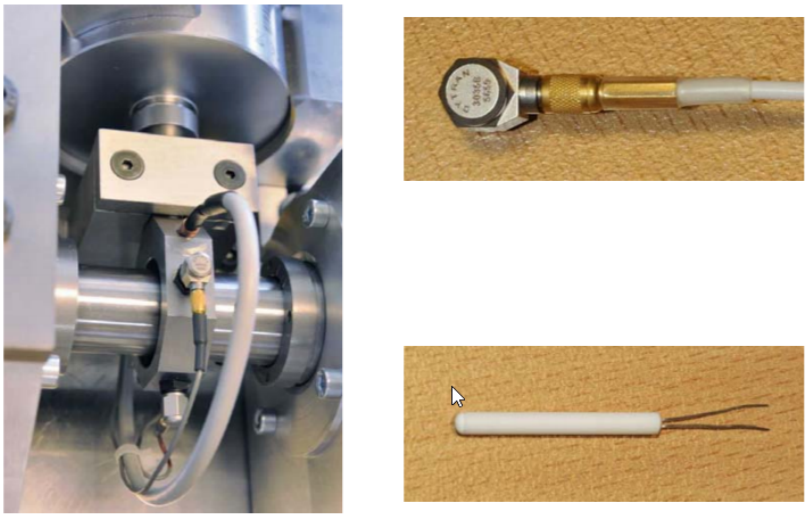


Figure 3.2: Sensors mounted on outer race [12]

The data provided by the IEEE PHM 2012 Prognostic challenge includes six training bearings, and eleven test bearings for test purposes. Training and test bearings are degraded according to one of three operating conditions.

Bearings in the prognostic challenge have a rated dynamic load of 4000N, and in operating condition two and three, the maximum dynamic load specification is violated. The following table display the different operation conditions.

Operating condition	Frequency of bearing shaft(Hz)	Applied radial load(N)
1	30	4000N
2	27.5	4200N
3	25	5000N

Table 3.1: Different operating conditions in the prognostic challenge

Of the six training bearings, there are two bearings representing each of the operating conditions. However, for operating conditions the test bearings were distributed unevenly. The participants in the IEE PHM 2012 data challenge were supposed to predict the RUL of test bearings from truncated test sets. Training bearing data was provided from start to EOL for modelling purposes. The following tables below show training and test bearing details.

Training bearing id	Condition	Total life(s)	#files
1_1	1	28030	2803
1_2	1	8710	871
2_1	2	9110	911
2_2	2	7970	797
3_1	3	5150	515
3_2	3	16370	1637

Table 3.2: Training bearing details

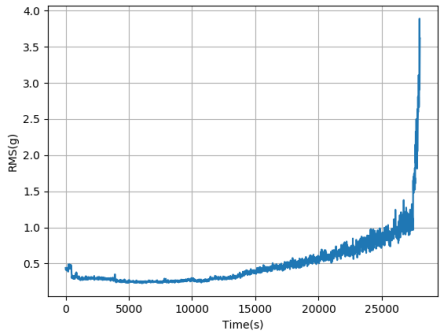
Test bearing id	Condition	Current runtime(s)	RUL to estimate(s)
1_3	1	12290	5730
1_4	1	11051	339
1_5	1	21410	1610
1_6	1	21560	1460
1_7	1	7450	7570
2_3	2	4490	7530
2_4	2	4730	1390
2_5	2	16930	3090
2_6	2	4430	1290
2_7	2	1140	580
3_3	3	2700	820

Table 3.3: Test bearing details

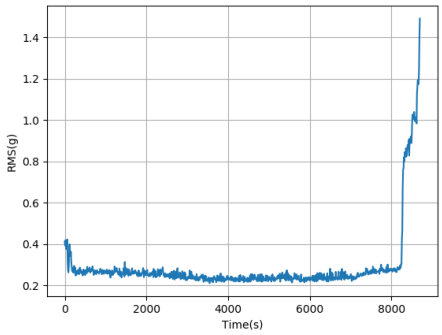
It is stated by PRONOSTIA, that the given bearings do not follow theoretical L10 estimation of bearing life, and that theoretical frequency frameworks for bearing degradation such as eq(2.3)-(2.6) do not work. Bearing EOL was defined when either of the accelerometers exceeded an acceleration signal above 20g.

The vibration data for both training and test bearings were sampled at 25.6kHz as CSV files. Each bearing is represented by 0.1 seconds long files until EOL, where the time between the sampled files is ten seconds. Each file recording consists of 2560 data points. Temperature measurements were not provided for all of the bearings, and is not utilized in this thesis.
In operation condition 1 (see table 3.1) the loading is equal to the dynamic load rating of bearings. From the equations regarding L_{10}-life eq(2.1-2.2), the L_{10}-life for bearings in operating condition 1 is 9.259 hours. Even though operation condition 1 is within the load specifications none of the train bearings in operation condition 1 meets the calculated L_{10}-life.

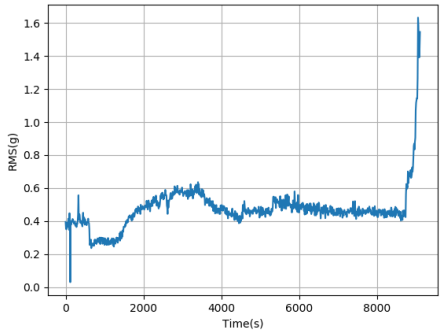
Indications of degradation can be observed by calculating the RMS vibrations for each sampled file. The figures below display how horizontal RMS vibrations and bearing life are correlated.



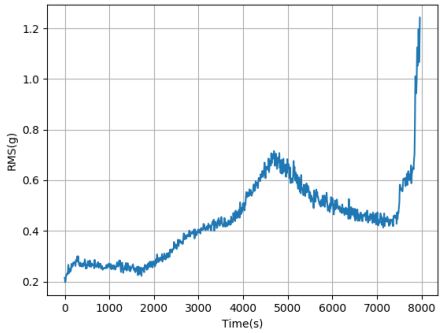
(a) Bearing1_1



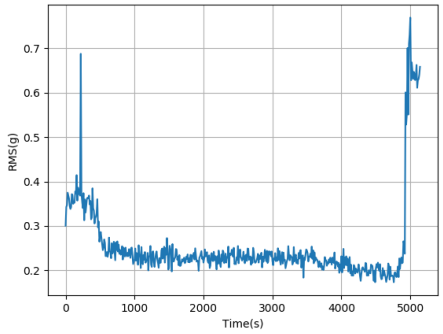
(b) Bearing1_2



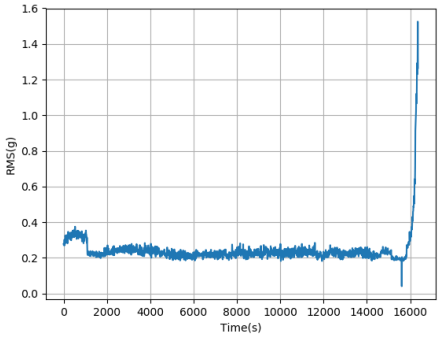
(c) Bearing2_1



(d) Bearing2_2



(e) Bearing3_1



(f) Bearing3_2

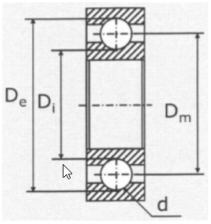
Figure 3.3: Horizontal RMS vibrations until end-of-life (EOL) (EOL)

By inspecting all the training bearings, it is clear that bearing1_1 endure

substantial longer compared to the other bearings. Bearing1_1 has an exponential degradation pattern which is similar to typical degraded bearings from the PRONOSTIA platform[13]. The normal degradation pattern of bearing1_1 is possibly explained by the more sustainable operation condition.

Most of the ball bearings in the training data set only show clear RMS vibration deviations in the later stages of failure, and do not follow a monotonic trend. Therefore, prognosis and diagnosis in earlier stages based on RMS vibrations are difficult. The RMS figures indicate higher vibrations in earlier stages likely due to a run-in period, where bearings are eventually worn smoother. As seen in figure 3.3 the RMS vibrations during the lifetime do not necessarily follow a linear or an exponential relationship, except for bearing1_1. Since the RMS figures 3.3 seemingly do not provide any general knowledge about the bearing degradation behaviour, frequency information will be used to try and characterize degradation. The sampled time domain signal is converted into frequency domain, as vibration signals are often easier to interpret in the frequency spectrum, and as different characteristic defects may correspond to certain frequencies.

The bearings used in the prognostic challenge have the following geometry:



d = 3.5 mm (Diameter of rolling elements)
Z = 13 (Number of rolling elements)
De = 29.1 mm (Diameter of the outer race)
Di = 22.1 mm (Diameter of inner race)
Dm = 25.6 mm (bearing mean Diameter)

Figure 3.4: Bearing geometry

From eq(2.3)-(2.6) the characteristic frequencies for the mechanical components are determined and listed in table 3.4.

Operating condition	BPFO(Hz)	BPFI(Hz)	BSF(Hz)	FTF(Hz)
1	168.3	221.7	215.3	12.9
2	154.3	203.2	197.4	11.9
3	140.3	184.8	179.4	10.8

Table 3.4: Characteristic frequencies

4. Method

4.1 Signal pre-processing

Pre-processing data is an important tool for guiding data-driven models to do accurate predictions. Processing techniques are in this thesis applied to the vibration data to increase frequency resolution, time resolution, magnitude accuracy, and to help the neural network train. Avoiding noise is especially important when modelling with small amount of data, where generalization is a problem.

The proposed processing techniques in this thesis include single file STFT (1F STFT), ten file STFT (10F STFT), ten file phase corrected STFT (10FPC STFT), single file continuous wavelet transform (1F CWT) and normalization of data (figure 4.1).

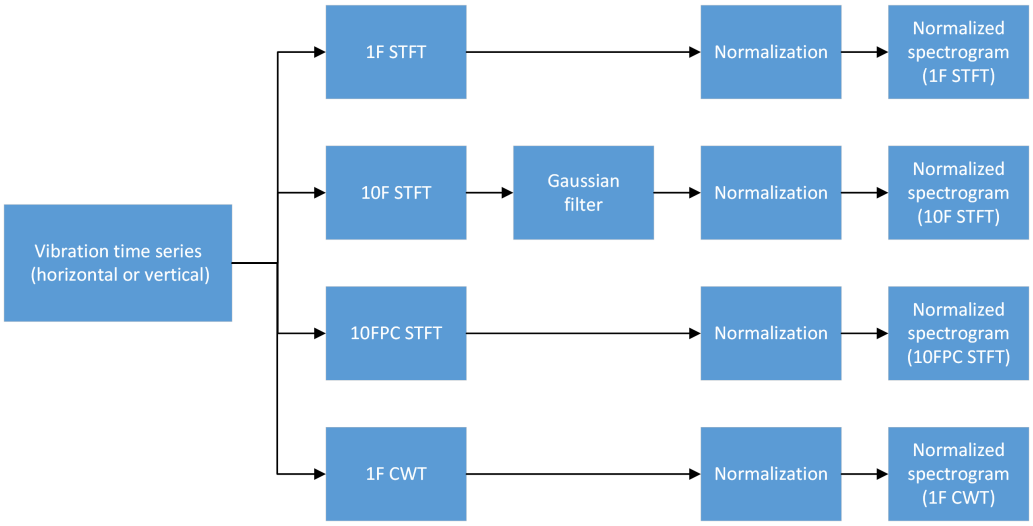


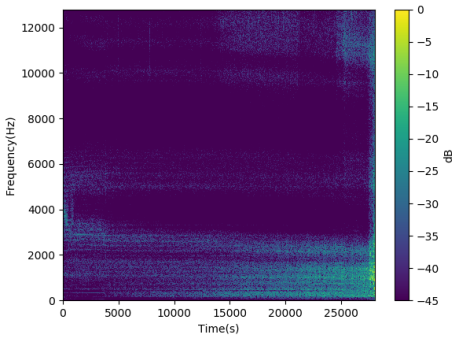
Figure 4.1: Overview of the signal pre-processing part

To compare the different processing techniques, each technique is measured in terms of mean average error (MAE) between actual RUL and predicted RUL of test bearings.

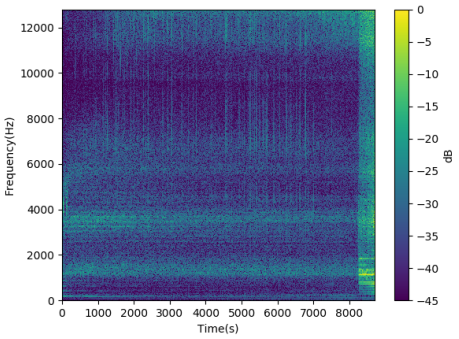
4.1.1 Single file STFT (1F STFT)

The most commonly used technique for spectral analysis is the Fast Fourier transform (FFT). One of the downsides with the FFT for prognosis is that FFT does not provide temporal information. STFT solves this issue by performing FFTs over smaller time segments. The most convenient time segment is calculating FFT for each single file of 0.1 seconds (2560 samples). By computing the FFT for each file along the y-axis in a spectrogram, it is possible to see changes in frequency related to time. As single files are sampled for a total of 0.1 seconds, the corresponding frequency resolution is 10Hz/bin along the frequency axis.

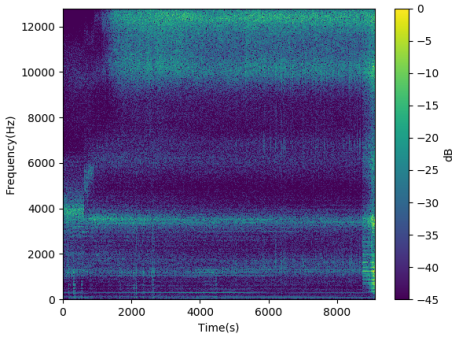
Figure 4.2 below shows STFT or time-frequency spectrograms for each training bearing. With 1F STFT one file represent a timestep in the spectrogram.



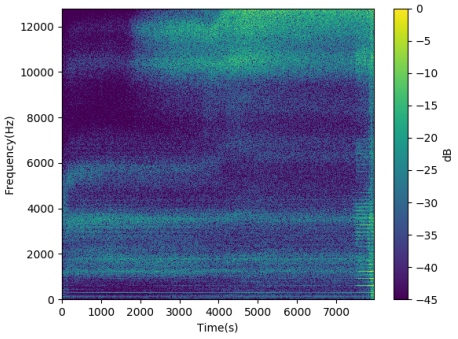
(a) Bearing1_1



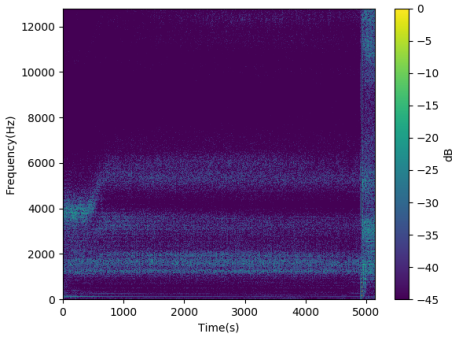
(b) Bearing1_2



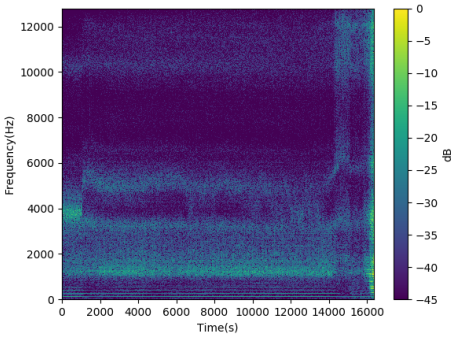
(c) Bearing2_1



(d) Bearing2_2



(e) Bearing3_1



(f) Bearing3_2

Figure 4.2: 1F STFT of training bearings until EOL

By observing the spectograms of the bearings it is reasonable to conclude that

the bearings do not share one general degradation trend. However, there are some similarities within the different operation conditions. Bearing2_1 and bearing2_2 definitely share similarities, meanwhile bearing3_1 and bearing3_2 share to some extent.

Figure 4.3 shows how the low frequency resolution obtained using 1F STFT does not result in an accurate representation of the shaft frequency with harmonics.

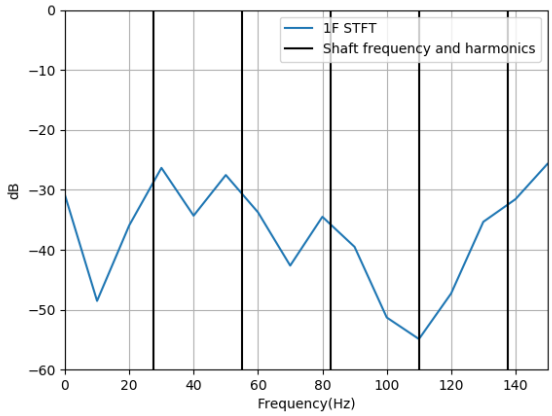


Figure 4.3: 1F FFT for bearing2_1 after 3000 seconds

4.1.2 Ten file STFT (10F STFT)

In order to detect smaller changes in the frequency spectrum higher frequency resolution is required. An example of this is if the balls start to slide and their corresponding frequency starts to deviate from the original rotation frequency. Higher frequency resolution may be obtained by merging sampled files together for longer sampling time. In this thesis ten and ten files were merged to yield a sampling length of 1 second, which gives a frequency resolution of 1Hz. The difference in frequency resolution between single file and ten file merged is shown in the FFT spectrum at one timestep in figure 4.4. When merging files temporal resolution is traded for spectral resolution. The frequency resolution is increased by 10x, while the temporal resolution is decreased by the same amount. Figure 4.4 compares 1F FFT and 10F FFT unfiltered for bearing2_1.

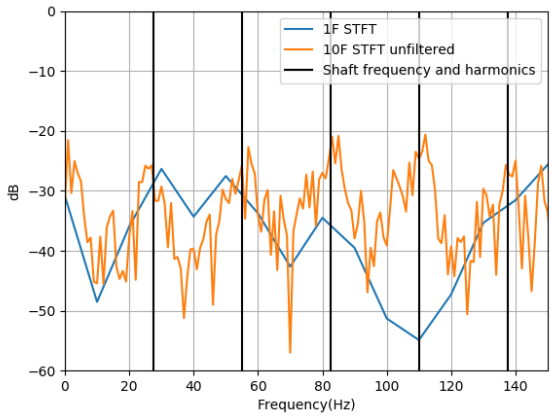


Figure 4.4: 1F FFT and 10F FFT unfiltered for bearing2_1 after 3000 seconds

As shown in the figure above the ten file FFT spectrum is noisy. This is a result of variable phase cancellation since the single files are not recorded continuously and therefore appended in random phase. For this reason a Gaussian filter is applied for noise reduction. Figure 4.5 compares 10F FFT unfiltered and 10F FFT filtered.

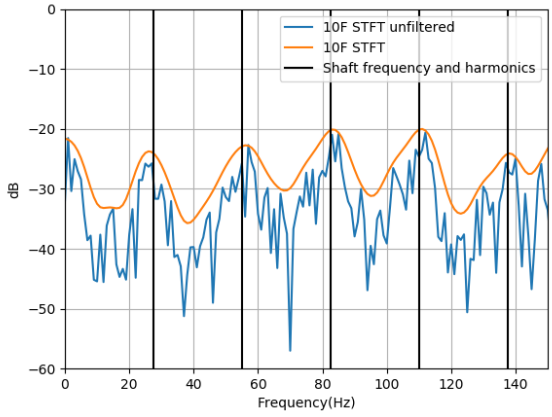


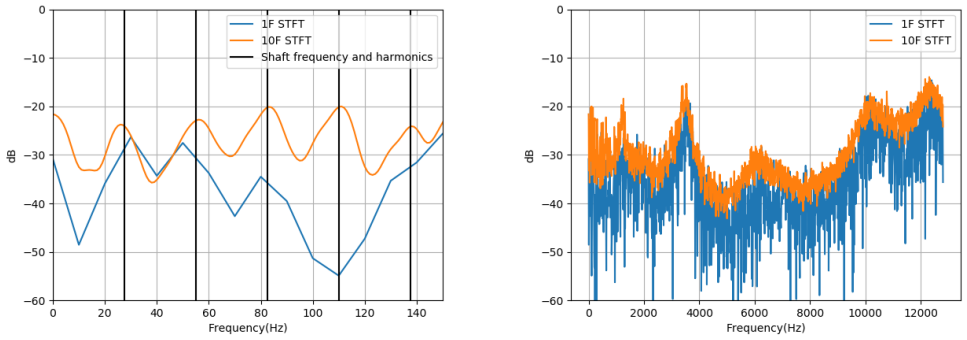
Figure 4.5: 10F FFT unfiltered and 10F FFT for bearing2_1 after 3000 seconds

In figure 4.5 the filtered signal has larger magnitudes in frequency band

0-150Hz. This is simply because the spectrum is plotted relative to the filtered max value which is slightly smaller than the unfiltered FFT maximum value.

The bearing in figure 4.5 demonstrates increased vibrations at the shaft frequency with harmonics. These frequencies are indications of imbalance or misalignment of bearings [10] [14]. The radial force is only applied on one side of the bearing, which may amplify bearing imbalances.

When comparing single file with ten file spectrums, one should be aware that these signals may be different since the time span for ten files is much longer. However, the FFT spectrums in figure 4.6 are taken from a timepoint in the middle of the bearings' life (3000 seconds), where no sudden changes in the spectrum are expected. Figure 4.6(a) indicates that the shaft vibration frequency with harmonics are better represented with 10F STFT.



(a) 0-150Hz (b) 0-12800Hz/Nyquist frequency

Figure 4.6: FFT of bearing 2_1, shaft frequency at 27.5Hz after 3000 seconds

Figure 4.7 show 10F STFT until EOL for bearing2_1.

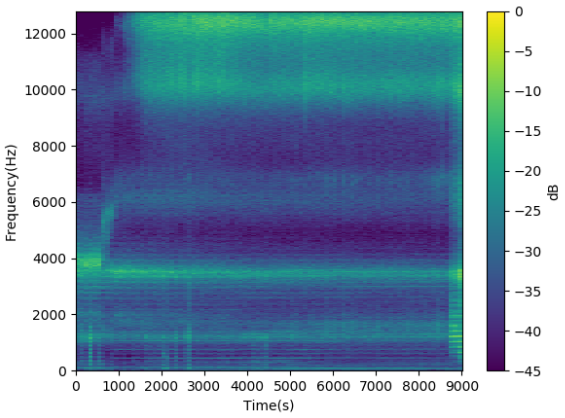


Figure 4.7: 10F STFT of training bearing2_1 until EOL. Ten files represents one timestep

The Gaussian filter introduced to reduce the noise from random phase with 10F STFT will also, to some extent, remove valid frequency information present at the accelerometers. A pre-processing technique where appended files are phase corrected are used to avoid introducing random phase noise.

4.1.3 Ten file phase corrected (10FPC STFT)

Phase correction is implemented by appending the files phase aligned for each frequency bin individually and computing the phase corrected magnitude of each bin on the phase aligned files. The phase corrected STFTs are low frequency limited to 20Hz, which is below the lowest shaft frequency of 25Hz. The phase alignment is implemented by introducing zeroes between the single files to bring them in phase for each bin. The maximum required zero insertion occurs if the file to be appended is initially nearly 360 degrees out of phase. For the 20Hz low frequency bin the phase alignment may therefore require up to 50ms of zeroes to be appended in phase. The 10FPC time series are therefore initially zero padded with 10240 samples so the frequency bins are the same for all ball bearings independent on the number of zeroes required to adjust the phase for all bins and bearings in the data set.

The phase alignment is implemented with the resolution given by the raw sampling rate. For this reason the phase alignment is less accurate for the higher frequency bins and starts to become inadequate at frequencies above about Fs/8 (3200Hz). This limitation could be avoided by up-sampling the signal. Up-sampling was not implemented since the algorithm is already very computational intense.

Figure 4.8 below show 10F STFT unfiltered/filtered and 10FPC STFT for comparison.

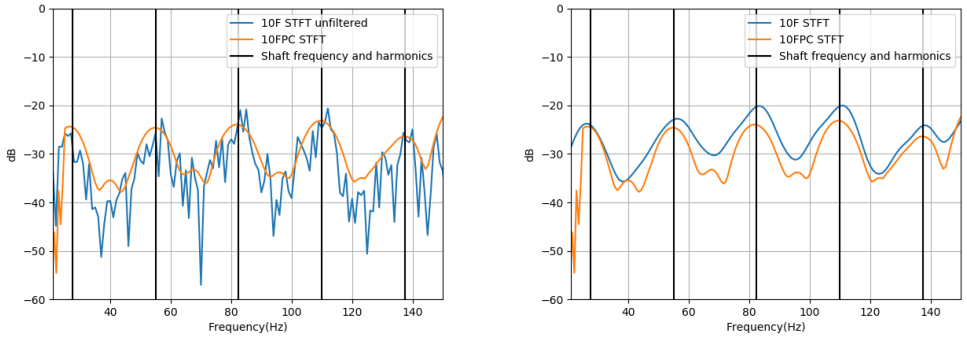


Figure 4.8: 10FPC FFT and 10F FFT/unfiltered of bearing 2_1 after 3000 seconds

Figure 4.9 illustrate 10FPC STFT of bearing 2_1.

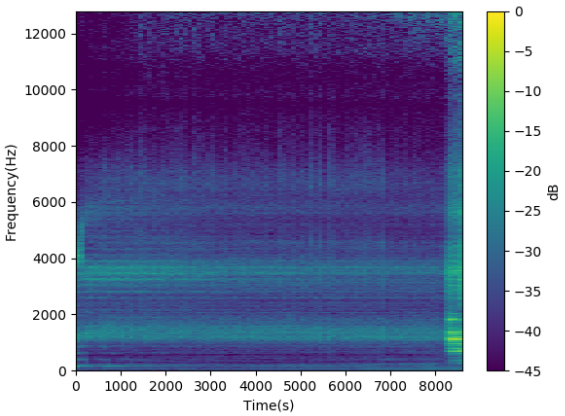


Figure 4.9: 10FPC STFT of training bearing2_1 until EOL. Ten files represents one timestep

4.1.4 Single file continuous wavelet transform (1F CWT)

One of the limitations with STFT for time-frequency analysis is the compromise between time and frequency resolution. However, wavelets filters can be customized to fulfill this issue. Wavelet filters are essentially band-pass filters that are located in time, which is unlike FFT where cross-correlation between the signal and the sinusoidal frequency is performed all at once. The wavelet transform extract information from the signal by convolving scaled wavelet filters with the signal. For time-frequency analysis the continuous wavelet transform (CWT) is usually preferred over the discrete wavelet transform (DWT).

Wavelet filters are normally designed with a logarithmic increasing center-frequency, where high frequency resolution in lower frequency bands and high time resolution in higher frequency bands are obtained. However, by evaluating the STFT figures from previous sections it can be observed that high frequency resolution may be important in higher frequency bands as well. For that reason wavelets were linearly scaled 1Hz apart, similar resolution to the 10F STFT. Wavelets can be configured in terms of mother wavelet, scaling, and full-width half-maximum value (FWHM). The width of the wavelet frequency spectrum is described by the FWHM value and the

type of mother wavelet. If a narrower frequency band is desired the FWHM is set at higher value. FWHM can be changed dynamically according to the center frequency of the wavelet such that the wavelet filters have the same frequency spectrum width[15]. In this thesis Morlet wavelets with dynamic scaling of the FWHM value are used to create narrow band-pass filters. The idea behind this is to prevent the band-pass filters from overlapping too much in the frequency spectrum.

The wavelets filters used are designed from a center frequency of 20Hz to 12800Hz (Nyquist frequency). In order to create a narrow wavelet filter with a center frequency of 20Hz, the wavelet filter length has to be longer then 0.1 seconds for a good representation. Meaning that the wavelet filters are actually longer then the data file length. Convolving a file/timestep of 2560 samples with 12780 wavelets filters yields a unsustainable amount of wavelet coefficients. Especially, considering that each bearing contains many files/timesteps. In order to reduce the number of coefficients, the max value for each wavelet filter within in each file is selected. The max value is the location where the wavelet filter and signal correlate most. Each file will in the end be represented by 12780 wavelet coefficients. This way 1F CWT provide the same time resolution as 1F STFT, and nearly the same frequency resolution as 10F STFT. However, the amount of data is close to tenfold. Figure 4.10 show 1F STFT and 1F CWT for comparison.

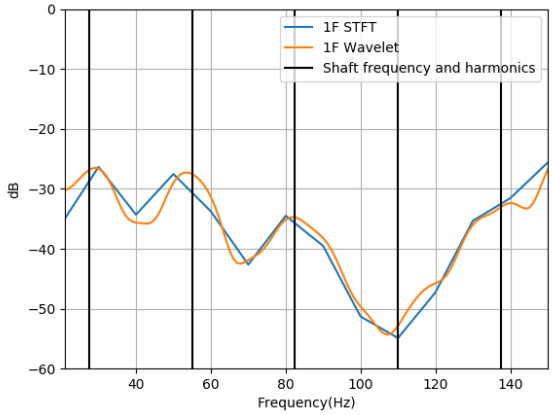


Figure 4.10: 1F FFT and 1F CWT of bearing1_2, with shaft frequency of 27.5Hz

Figure 4.11 illustrates the 1F CWT spectrogram of bearing1_2.

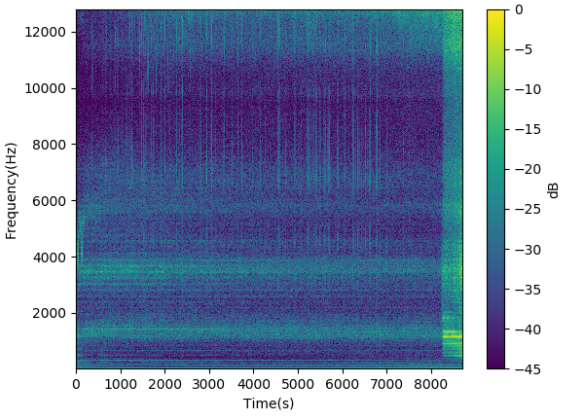


Figure 4.11: 1F CWT of bearing1_2 until EOL. One file represents one timestep.

4.1.5 Normalization

The bearing spectrograms were normalized by scaling and mean subtraction before entering the neural networks. The mean value was acquired from the spectrograms of the training bearings. Scaling was performed such that the vibration values remain in approximately the same ranges of magnitude regardless of signal pre-processing technique. Normalization of data helps neural networks train easier and to converge. Proper utilization of activation functions require the data input to be within certain ranges, otherwise training issues may arise in early stages of training. Generally, normalization is considerably more critical when the input data consist of multiple parameters. Even though vibration is the only parameter, normalization was used.

4.2 Neural network models

For predicting the RUL from the spectrograms neural network models have been used. Of the two main approaches, one is based on a recurrent neural

network (RNN) and the other is based on a combination of a convolutional neural network (CNN) and a RNN.

4.2.1 Recurrent neural network model

The idea behind utilizing a recurrent neural network for prognosis of bearings originates from the data structure. A RNN is able to transfer temporal information in sequence data, which is a property not present in other types of neural networks such as convolution neural network and feed forward neural network. Temporal convolutions have been introduced in sequence modeling, but their maximum temporal memory length is limited by the CNN structure [16]. Transfer of temporal information may be useful, e.g. being able to track and store information about the run-in period as features for later use.

The simplest form of a recurrent neural network is given by the following equation:

$$h_t = tanh(W^x x_t + W^h h_{t-1} + b) \tag{4.1}$$

x_t and h_t is the input vector and hidden unit vector at timestep t, h_{t-1} is the previous hidden unit vector and W^h, W^x and b are trainable weights as matrices and a vector. Challenges related to RNNs are to preserve long range dependencies and to prevent vanishing and exploding gradients during the training phase. From eq 4.1, it can be noticed that the hidden units are forced to be updated for every timestep.

Advanced recurrent cell structures have been developed to improve long range dependencies in recurrent neural networks. Two commonly used methods are gated recurrent unit (GRU) and the Long short-term memory (LSTM) [17] [18]. In this thesis LSTM was chosen without consideration of GRU. LSTM cells are more capable of learning long-term dependencies and better at dealing with vanishing and exploding gradients as LSTM cells utilize gates for regulating the flow of information [19]. The LSTM cell is described by the following equations:

Forget gate	$\Gamma^f = \sigma(W^f x_t + U^f h_{t-1} + b^f)$	(4.2)
Input gate	$\Gamma^i = \sigma(W^i x_t + U^i h_{t-1} + b^i)$	(4.3)
Output gate	$\Gamma^o = \sigma(W^o x_t + U^o h_{t-1} + b^o)$	(4.4)
Potential cell memory	$\tilde{c}_t = tanh(W^c x_t + U^c h_{t-1} + b^c)$	(4.5)
Cell memory	$c_t = \Gamma^f \circ c_{t-1} + \Gamma^i \circ \tilde{c}_t$	(4.6)
Cell state	$h_t = \Gamma^o \circ \ tanh(c_t)$	(4.7)

Without going to deep into how the different gates operate, it is observed that the potential cell memory is governed by a similar equation as for the simple RNN cell. The update of the cell memory is however regulated by the input gate and forget gate. The input gate adjust how much the new information should be weighted and the forget gate adjust how much old information should be kept. The cell state is also controlled by the output gate, giving the LSTM cell the ability to select features particularly important for the prediction at the current timestep. It is also worth noticing that the number of hidden units (features) of the cell memory and cell state determine the capacity of the LSTM.

An overview of the RNN model is given in figure 4.12.

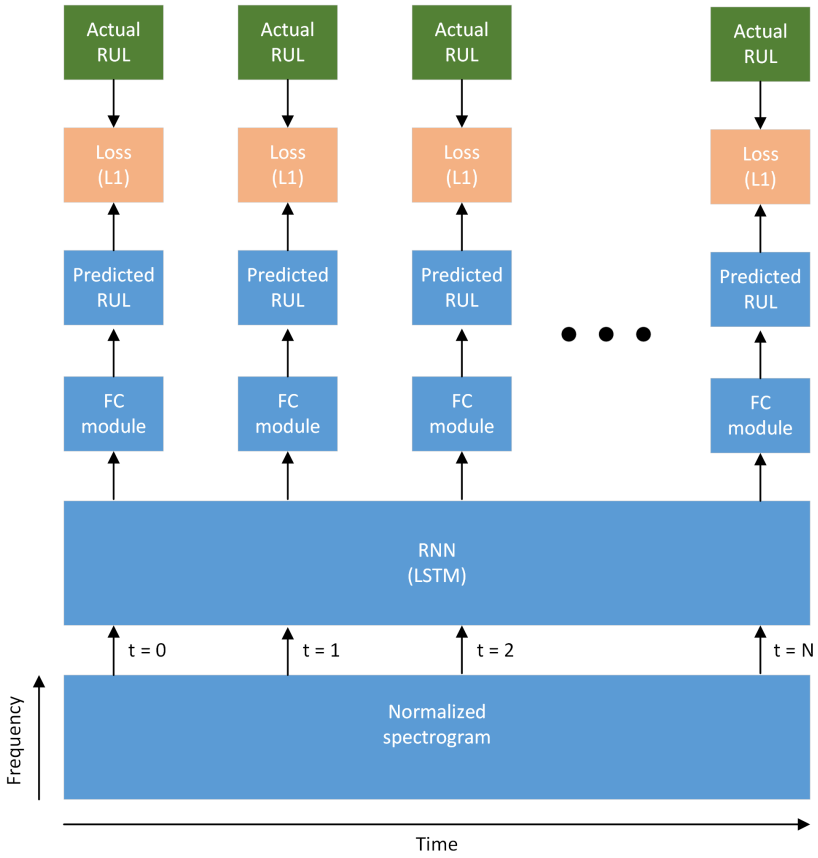


Figure 4.12: Overview of the RNN model showing how the spectrogram, the RNN model and the FC module are connected. Dimensions are not shown in the figure since they are dependent on spectrograms input dimensions and the number of hidden units which will later be tuned.

Image data is seldom passed to RNNs as features directly since the RNN structure has no spatial invariance. However, a spectrogram can be thought of as a special image where the pixel position has a particular meaning. The pixel position on the vertical axis is linked to the frequency content and for this reason discarding a feature extraction layer is reasonable.

For each timestep in the spectrograms the frequency vector is passed into the LSTM. The LSTM cell updates the cell memory and the cell state based on the frequency vector and the previous LSTM cell configuration. The features

in the cell state are further passed to a fully connected (FC) module which produces the predicted RUL for the current timestep. The structure of the FC module is shown in table 4.1.

	Input-features	Output-features
Input:		
FC layer	# LSTM HUs	(# LSTM HUs)/2
ReLU	-	-
FC layer	(# LSTM HUs)/2	1

Table 4.1: FC module

The FC module consists of two fully connected layers with a ReLU activation function in between. The number of features is a function of the number of hidden units (HUs) in the LSTM cells. The weights and biases are initialized using He initialization [20]

It is common to link the output features from the LSTM cell state using fully connected layers and activation functions for non-linear relations. In a fully connected layer all input features are connected to all output features with trainable weights. Stacking multiple fully connected layers directly is meaningless, as two linear operations could be expressed as one linear operation. Adding an activation function makes the neural network capable of introducing non-linear relations between the features. Typical activation functions are the sigmoid function, the tanh function and the rectified linear unit (ReLU) function. The sigmoid and the tanh functions are prone to the vanishing gradient problem in deep neural networks, for this reason the ReLU function is generally able to converge the training faster [21]. Although the RNN model is shallow, the ReLU function has been chosen. The ReLU function is also less computational heavy and have simple derivatives of 0 and 1 for values below and above zero respectively. The ReLU function is illustrated in figure 4.13.

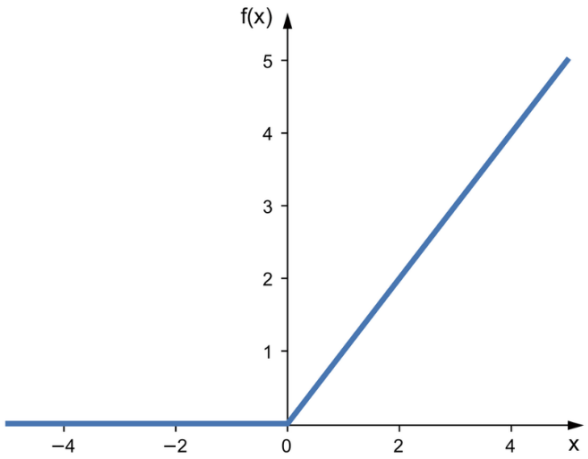


Figure 4.13: The ReLU activation function

4.2.2 Loss function

The type of loss function selected is problem dependent. For regression type problems a distance metric should be used, and for classification problems softmax cross entropy is common. Predicting the RUL is a typical regression task and the chosen loss function is the mean absolute error (MAE). MAE was selected over mean squared error (MSE) as MSE is more sensitive to outliers. MAE loss for bearing predicitons are described in eq 4.8.

$$MAE = \frac{1}{N} \cdot \sum_{t=1}^{N} |actRUL_t - predRUL_t| \tag{4.8}$$

MAE describes the average absolute error over a bearing's life. N is the total number of timesteps in a bearing's life and t is the running index corresponding to the various timesteps. $actRUL_t$ and $predRUL_t$ are the actual and the predicted RUL at timestep t.

4.2.3 Optimization

The model is trained using a gradient descent based optimization method. Although gradient descent based optimization algorithms do not guarantee a global minimum, good local minima are normally found. The model

weights are updated based on the gradients calculated from MAE loss using backpropagation. In this thesis ADAM has been used for optimization [22]. ADAM has the benefit over stochastic gradient descent that it incorporates a momentum term with an adaptive learning rate. Even though ADAM uses adaptive learning rate, it is still important to set appropriate learning rate initially. Too high learning rate will result in unstable training, whereas too low learning rate results in slow convergence.

The deep learning framework used is Pytorch, which is an open-source machine learning library in Python [23]. In order to enable batch training, the spectrograms for each of the bearings must have the same shape. Initially, the length of the spectrograms' time axis corresponded to the bearings' lifetime. The issue was solved by zero padding after EOL according to the bearing with the longest lifetime. Bearing1_1 is the longest living bearing with 2803 sampled files. It is worth noticing that the total number of timesteps is dependent on the pre-processing technique. The pre-processing techniques merging ten and ten files have one tenth of the total number of timesteps. In ten file configurations the remaining files that do not go up in a whole number of ten were discarded. In such cases random files in the middle of bearing life were selected instead of the last files as they contain unique information. As an example bearing 1_1, 3 single files had to be discarded.

4.2.4 Hyperparameter tuning

The objective of hyperparameter tuning is to optimize the network configuration to yield better predictions on the validation data and ultimately the test data. Typical hyperparameters are activation functions, selection of optimizer, learning rate, network weight initialization, number of epochs, batch size, regularization and number of features in trainable layers.

In deep learning models it is common to divide the available training samples into separate training and validation sets. However, in this case the normal approach is not as preferable because the amount of training data is limited. For this reason, cross-validation with leave-one-out has been used in the hyperparameter tuning process. How cross-validation was performed is illustrated in figure 4.14.

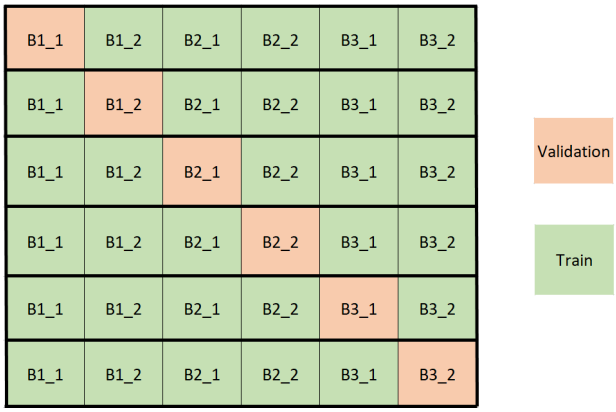


Figure 4.14: In cross-validation with leave-one-out, the model is trained using five training bearings while the sixth training bearing is used for validation.

Horizontal and vertical vibration have similar frequency spectograms. Through the cross-validation process it became clear that horizontal vibrations had better predictions overall compared to vertical, and magnitude vibrations. Hence, horizontal vibrations were selected as the utilized data axis for this thesis.

Data augmentation is normally used while training neural networks to reduce the risk of over-fitting. Typical data augmentation techniques such as random cropping, rotation and flipping do not make sense since the spectrograms are space-variant images. However, augmentations such as adding random noise and random starting points were tested. Unfortunately, neither of the augmentations had positive effects on the cross-validation predictions and were not used in the end.

Optimization of neural networks can be a time consuming process. In order to limit the numbers of hyperparameters, learning rate, training epochs and number of hidden units were selected for optimization, leaving the other hyperparameters fixed. A grid search was performed with learning rates of 0.01, 0.001 and 0.0001, and for hidden unit sizes of 256, 128 and 64. The optimal hyperparamters should ideally be determined for every pre-processing configuration individually, but because of time limitations the search was only completed for 10F STFT. Consequently, hyperparameters found from 10F STFT were utilized for all other pre-processing configurations.

The tables below show the results of the cross-validation process for the various learning rates and the number of hidden units. The MAEs displayed are taken from the epochs with lowest average MAE across all bearings.

lr	# epochs	B1_1	B1_2	B2_1	B2_2	B3_1	B3_2	avg MAE
0.01	10	11100	1890	2110	1530	6000	4240	4470
0.001	10	9700	2770	1800	1510	4740	4330	4140
0.0001	55	11300	2350	2380	1580	2960	4740	4210

Table 4.2: MAE loss on cross-validated bearings with 256 hidden units, where lr is learning rate

lr	# epochs	B1_1	B1_2	B2_1	B2_2	B3_1	B3_2	avg MAE
0.01	30	11400	1180	1770	1200	5710	4740	4340
0.001	15	10700	2290	2110	1350	3210	4520	4040
0.0001	95	12000	2180	2500	1780	2460	4880	4310

Table 4.3: MAE loss on cross-validated bearings with 128 hidden units, where lr is learning rate

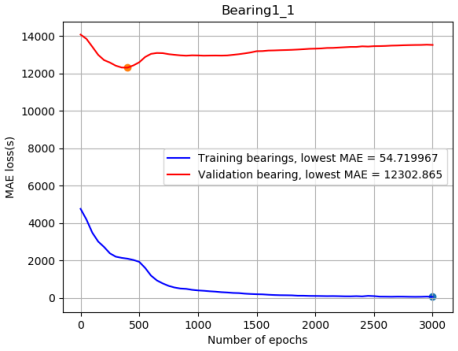
lr	# epochs	B1_1	B1_2	B2_1	B2_2	B3_1	B3_2	avg MAE
0.01	5	10500	2140	2060	1490	2390	4610	3860
0.001	40	11500	2070	2050	1580	2430	4730	4060
0.0001	345	12300	2010	2590	1270	2670	4790	4280

Table 4.4: MAE loss on cross-validated bearings with 64 hidden units, where lr is learning rate

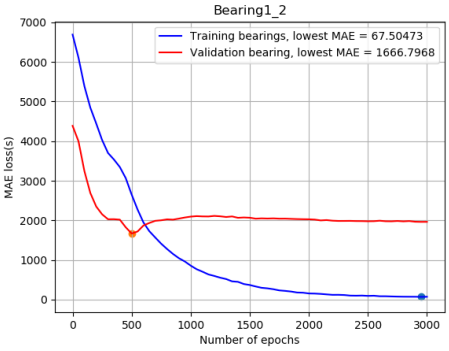
The set of hyperparameters resulting in the lowest average MAE were a learning rate of 0.01 and 64 hidden units. The lowest average MAE was obtain after five epochs only. After five epochs the model is untrained since five epochs corresponds to a total of five gradient updates. It is observed that the individual cross-validation MAEs are unstable early in the training process. For this reason it was decided to train longer even though minimum MAE was acquired earlier in the training process. Learning rates of 0.01, 0.001, and 0.0001 with 64 hidden units were considered individually since 64

hidden units had lowest MAE initially. Based on inspecting further training progress it was concluded that a learning rate of 0.0001 and 64 hidden units was the best option.

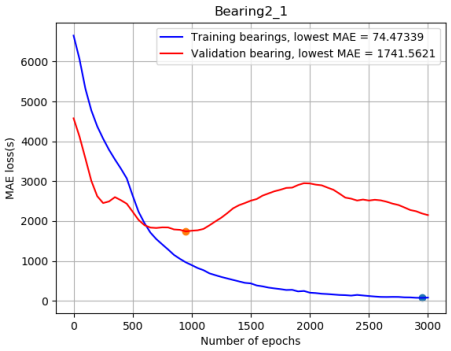
The figures below show cross-validation MAE loss for the chosen hyperparameters as a function of epochs.



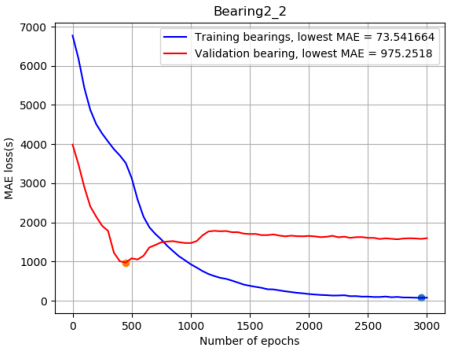
(a) MAE loss bearing1_1



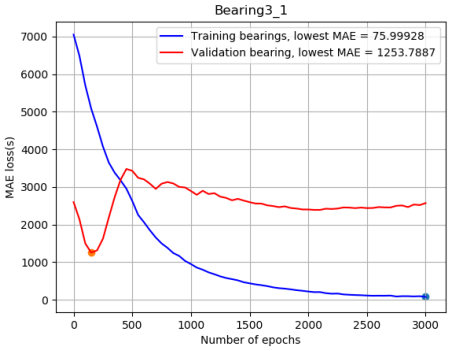
(b) MAE loss bearing1_2



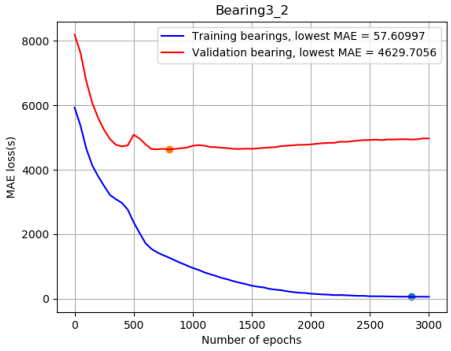
(c) MAE loss bearing2_1



(d) MAE loss bearing2_2



(e) MAE loss bearing3_1



(f) MAE loss bearing3_2

Figure 4.15: MAE loss of cross-validated bearings with learning rate of 0.0001 and 64 hidden units.

Since the cross-validation losses for all bearings do not follow a monotonic downward trend until a certain point, it was decided that 2500 epochs was the best and safest option. At 2500 epochs training loss and cross-validation loss have converged and are stable.

Figures 4.16-4.21 display the corresponding predictions and the 10F STFT spectrograms of the cross-validation bearings after 2500 epochs.

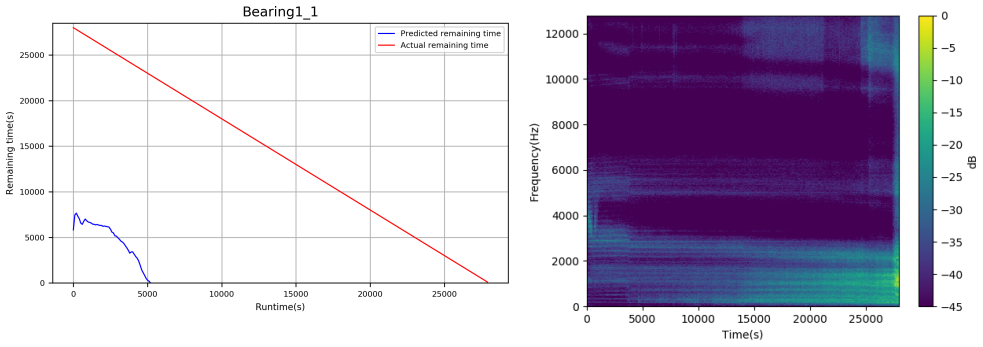


Figure 4.16: 10F STFT cross-validation and spectrogram of bearing 1_1

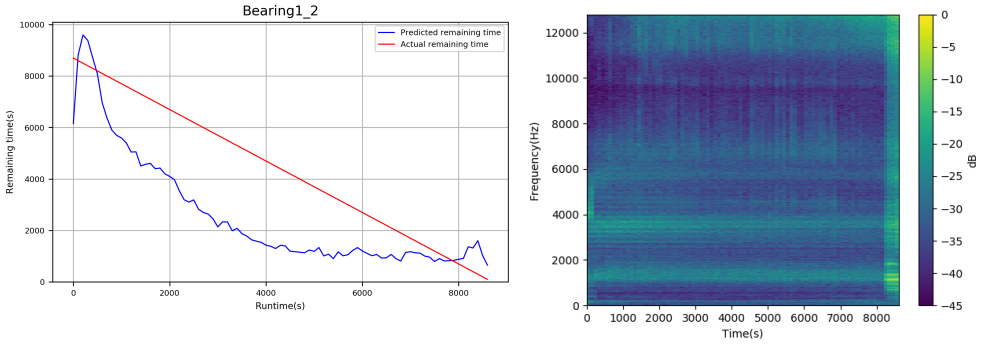


Figure 4.17: 10F STFT cross-validation and spectrogram of bearing 1_2

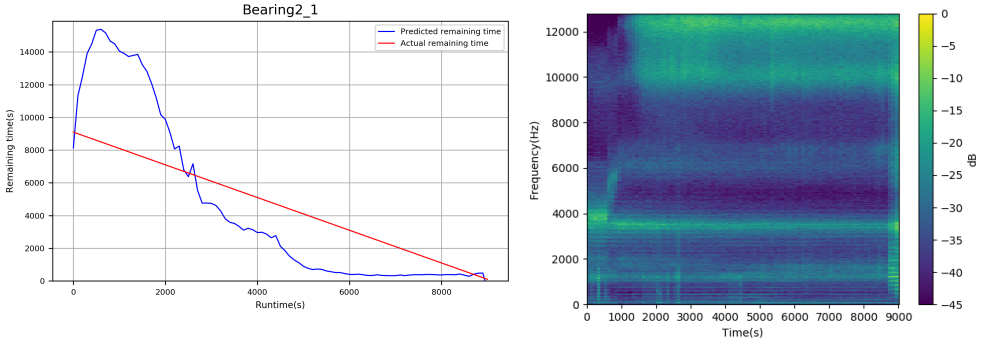


Figure 4.18: 10F STFT cross-validation and spectrogram of bearing 2_1

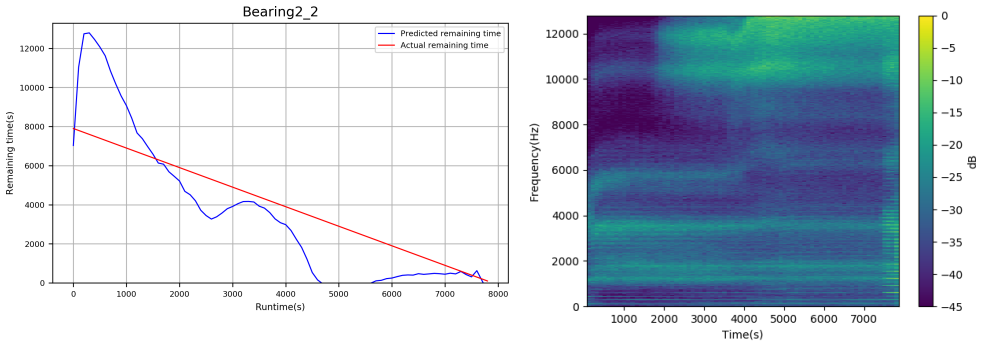


Figure 4.19: 10F STFT cross-validation and spectrogram of bearing 2_2

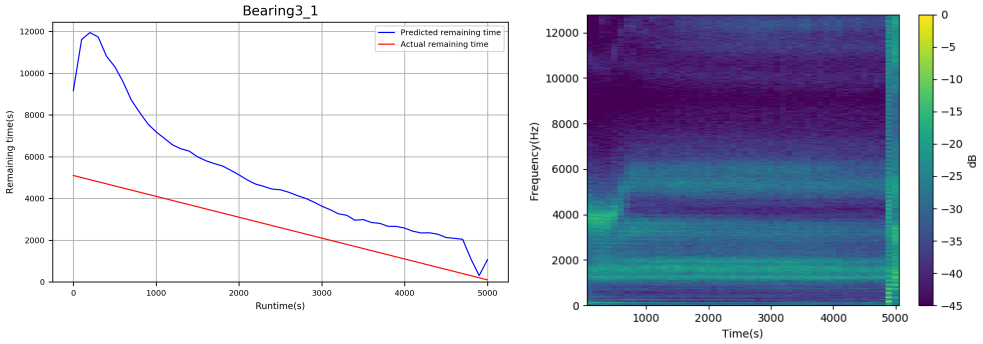


Figure 4.20: 10F STFT cross-validation and spectrogram of bearing 3_1

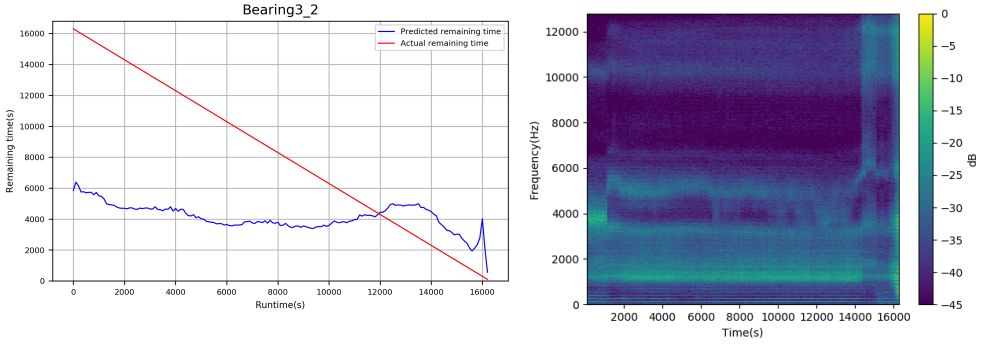


Figure 4.21: 10F STFT cross-validation and spectrogram of bearing 3_2

Based on the figures above it is difficult to identify what type of decisive attributes the network has learned. From the validation predictions it is

possible to observe the connections between input and prediction at given timesteps. This is best visualized by observing predictions in the last stages of bearing life where most validation predictions react to the change of sudden vibration intensity. The predictions for bearing1_1 are way off, and might be explained by a different degradation pattern compared to the other bearings. Even though bearing1_1 has a normal expected degradation, the other bearing have not. Consequently, bearing1_1 has degradation characteristics that is not present in the other bearings, and these characteristics are unknown to the network when validating bearing1_1.

The final models used for testing were in the end trained by all of the six training bearings together with the optimal hyperparameters found from the cross-validation.

4.2.5 RNN model with CNN as feature extractor

The LSTM cell is a fully connected layer the way it connects to it's input features. For that reason, the LSTM cell has no scale or translational invariance. The three different operational conditions have different shaft frequencies resulting in different resonance peaks in the frequency spectrum. If the main useful degradation information is located within the shaft frequency and corresponding resonance, then the LSTM lacks the ability to generalize across the operational conditions. To further improve on the limitation seen from the LSTM, a CNN module has been tested as a feature extraction layer. The CNN structure is not scale invariant, but at least have the property of being translational in-variant.

The RNN model including the CNN extraction layer is illustrated in Figure 4.22.

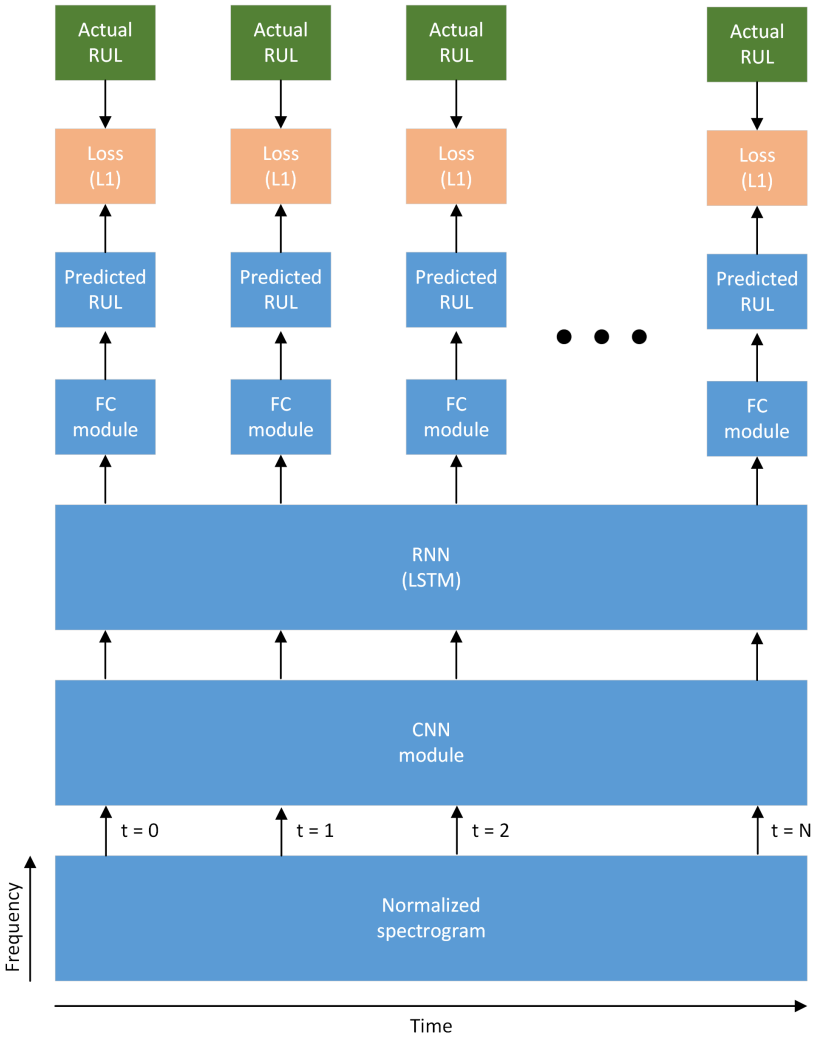


Figure 4.22: Overview of the CNN+RNN neural network

The design of the CNN module is found in table 4.5 below. The weights and biases are initialized using He initialization.

	Kernel	Stride	Padding	Channels
Input:				
Conv2d	(11,1)	(4,1)	(1,0)	32
ReLU	-	-	-	-
Conv2d	(5,1)	(1,1)	(2,0)	32
ReLU	-	-	-	-
Max pool	(3,1)	3,1)	-	-
Conv2d	(3,1)	(2,1)	(1,0)	32
ReLU	-	-	-	-
Max pool	(3,1)	3,1)	-	-
Conv2d	(3,1)	(2,1)	(1,0)	32
ReLU	-	-	-	-

Table 4.5: CNN module

4.3 Score function of the Prognostic challenge

The accuracy of the prognostic challenge was calculated by the difference in time error between predicted RUL and actual RUL of bearings at a specific points in time for each test bearing. Underpredictions and overpredictions were not treated equally as overpredicitions were considered to be more severe. Score of each configuration was determined based on the average accuracy across all test bearings. The score function differs between underpredictions and overpredicition, even though the network weights are trained according to MAE loss. MAE loss is symmetric by default, and does not evaluate late or early predictions like the score function does. From the figure 4.23 the scoring function does not distinguish between late (-40%) and very late predictions as they yield 0 score either way. This is also in contrast to MAE where the feedback is linear. The following equations describe how scoring was performed.

$$\%Er_t = 100 \cdot \frac{actRUL_t - predRUL_t}{actRUL_t} \tag{4.9}$$

$\%Er_t$ is the error between actual RUL and predicted RUL at the given time-point

If overpredicting, $\%Er_t$ becomes negative and accuracy will given as:

$$A_t = exp^{-ln(0.5)\cdot(Er_i/5)} \tag{4.10}$$

if underpredicting, $\%Er_t$ becomes positive and accuracy will given as:

$$A_t = exp^{+ln(0.5)\cdot(Er_t/20)} \tag{4.11}$$

A_t is the bearing accuracy

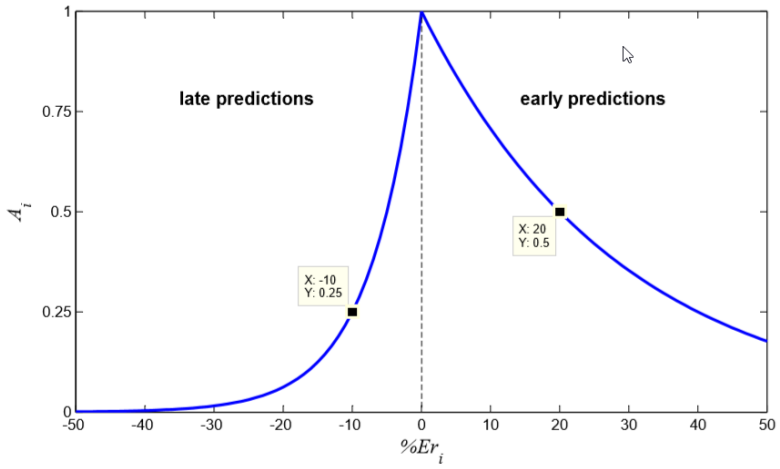


Figure 4.23: Scoring function used in prognostic challenge [12]

5. Results

5.1 Overview of configurations

The proposed configurations in this thesis have a basis in a LSTM RNN, but differ in terms of signal processing methods, and neural network design. The table below gives an overview of configuration used for testing purposes.

Configuration	**# merged files**	**pre-processing**	**CNN**	**RNN**
1F STFT	1	STFT	-	LSTM
10F STFT	10	STFT	-	LSTM
10FPC STFT	10	PC STFT	-	LSTM
1F CWT	1	CWT	-	LSTM
10F STFT CNN	10	STFT	True	LSTM

Table 5.1: Overview of configurations

5.2 Test predictions

Table 5.2 shows the RUL results for the test bearing with the 10F STFT configuration.

Test_bearing id	Condition	Actual RUL point(s)	Predicted RUL point(s)
1_3	1	5730	17748
1_4	1	339	55
1_5	1	1610	1664
1_6	1	1460	508
1_7	1	7570	8706
2_3	2	7530	4477
2_4	2	1390	7032
2_5	2	3090	1863
2_6	2	1290	820
2_7	2	580	4920
3_3	3	820	796

Table 5.2: Test bearings and RUL points to estimate (the RUL points used in the score function)

The following figures show actual RULs (red line), predicted RULs (blue line), actual RUL to estimate (black point), predicted RUL to estimate (green point), and spectrogram for the different test bearings.

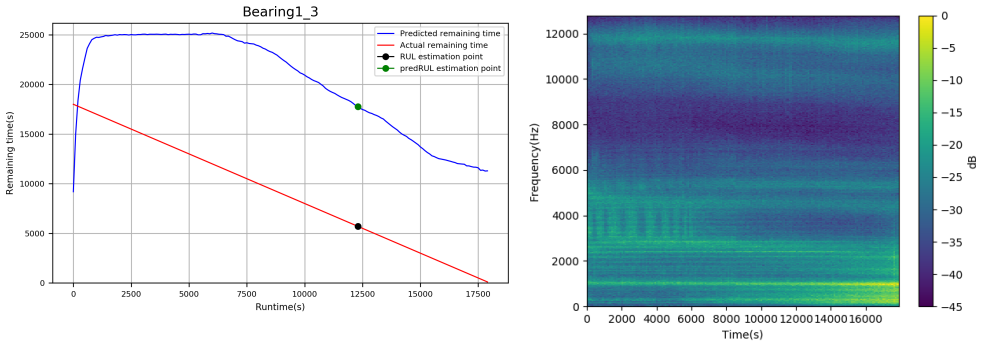


Figure 5.1: Prediction and spectrogram of test bearing1_3

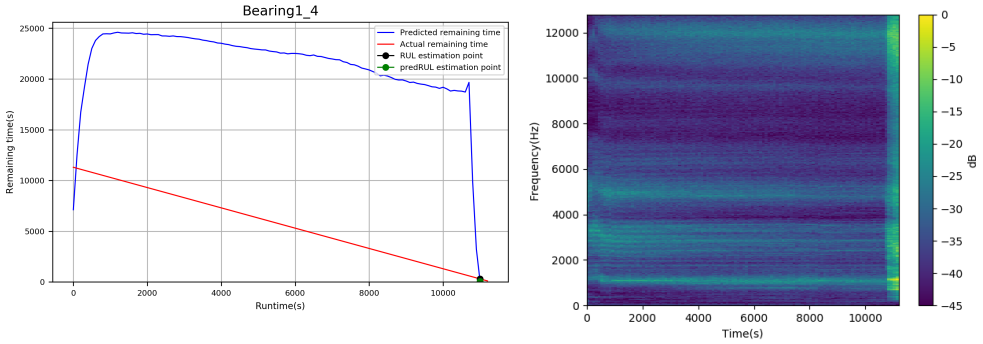


Figure 5.2: Prediction and spectrogram of test bearing1_4

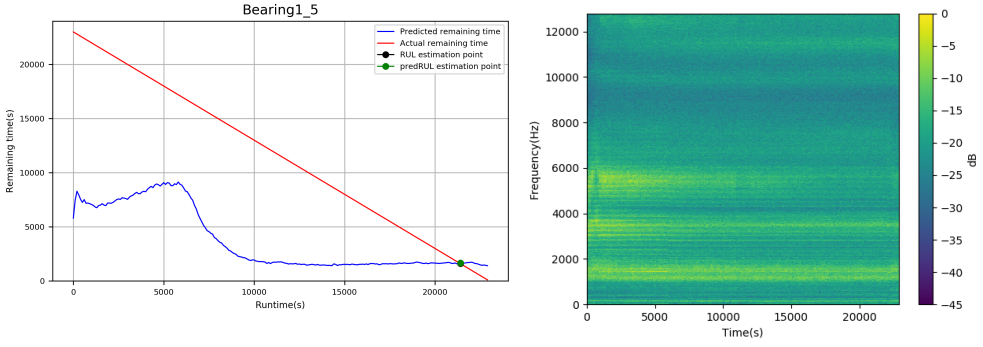


Figure 5.3: Prediction and spectrogram of test bearing1_5

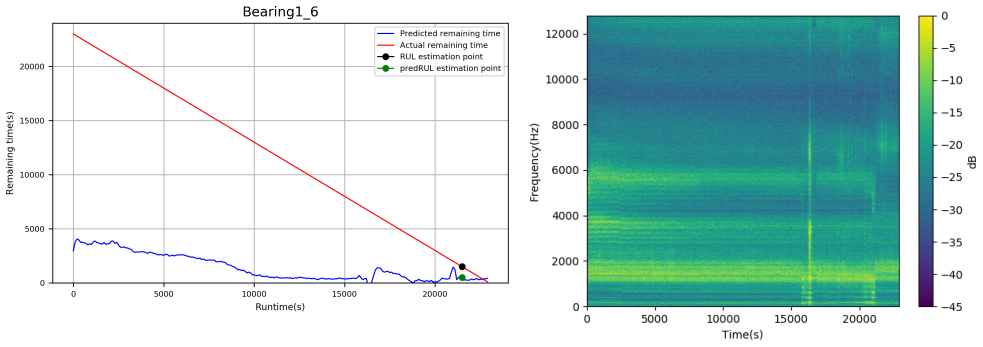


Figure 5.4: Prediction and spectrogram of test bearing1_6

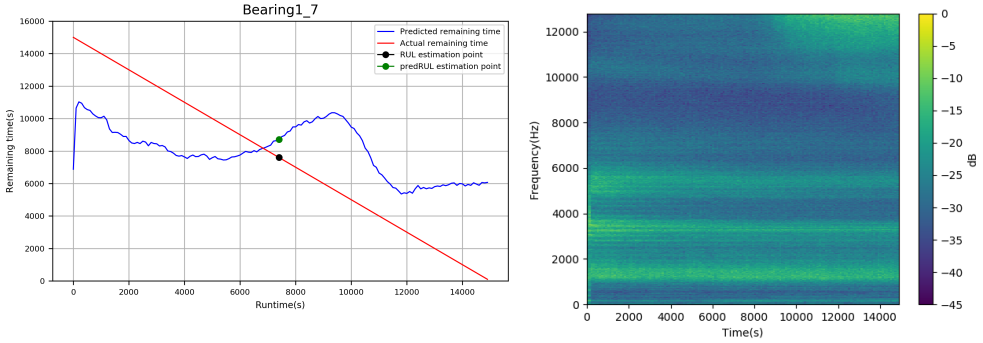


Figure 5.5: Prediction and spectrogram of test bearing1_7

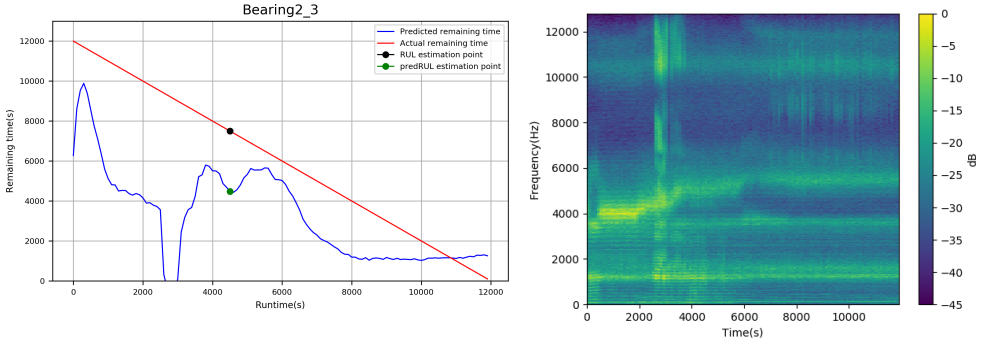


Figure 5.6: Prediction and spectrogram of test bearing2_3

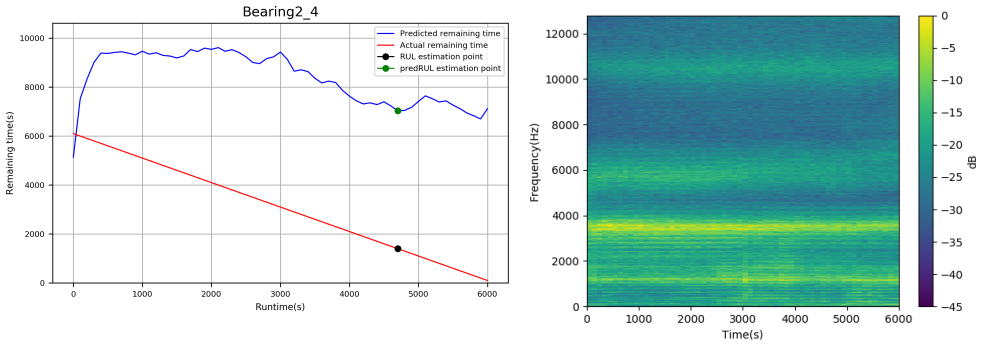


Figure 5.7: Prediction and spectrogram of test bearing2_4

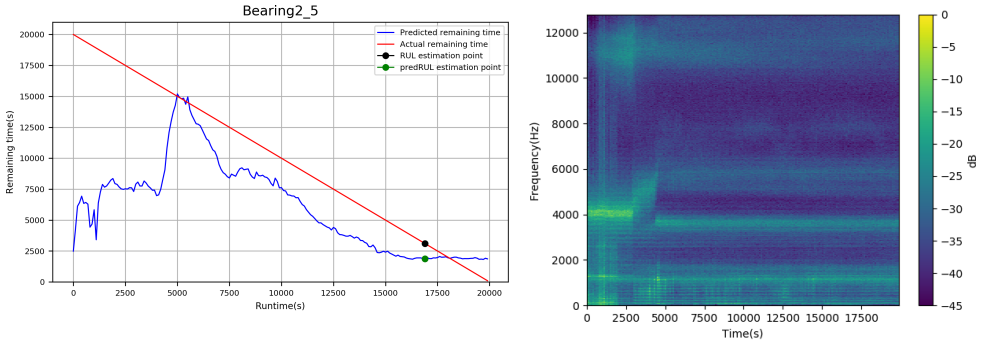


Figure 5.8: Prediction and spectrogram of test bearing2_5

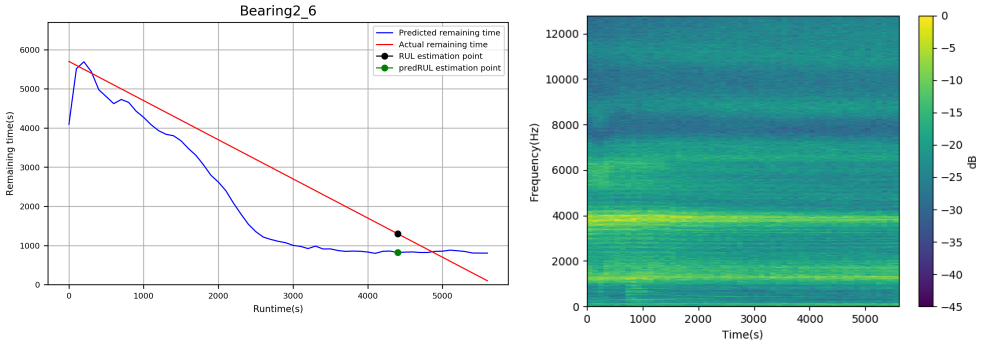


Figure 5.9: Prediction and spectrogram of test bearing2_6

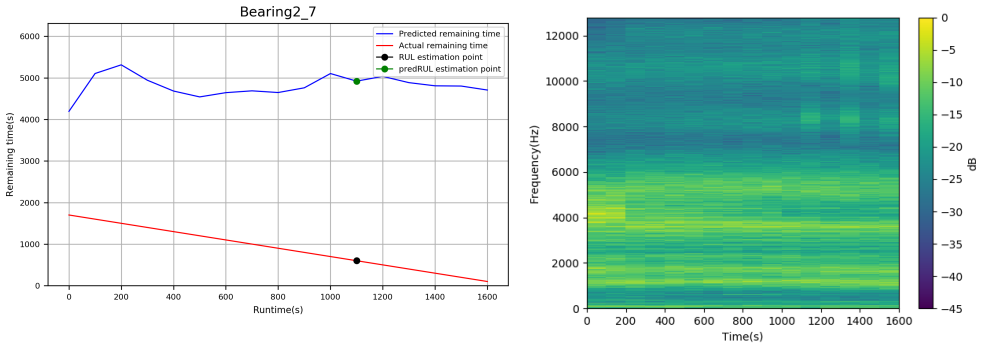


Figure 5.10: Prediction and spectrogram of test bearing2_7

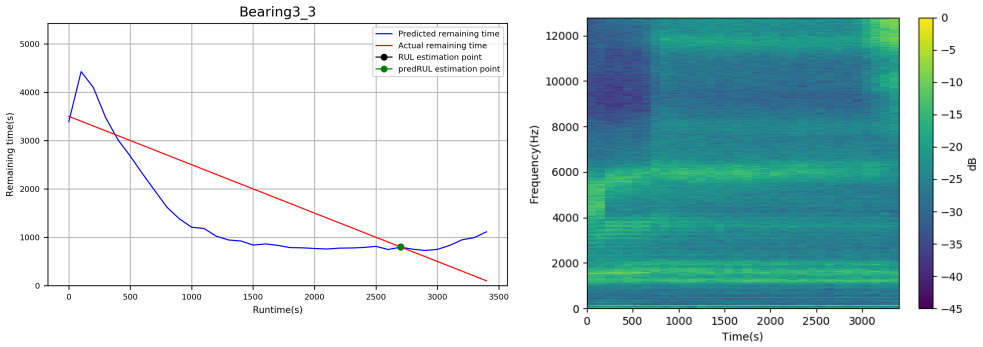


Figure 5.11: Prediction and spectrogram of test bearing3_3

The following table shows MAE, and score for different bearings.

Test bearing	MAE(s)	Score
1_3	11319	2.353e-13
1_4	15482	0.055
1_5	7920	0.627
1_6	10133	0.105
1_7	3316	0.125
2_3	2911	0.245
2_4	5309	3.640e-25
2_5	4094	0.253
2_6	826	0.283
2_7	3910	8.817e-46
3_3	678	0.903
avg	**5990**	**0.236**

Table 5.3: Performance of 10F STFT on test bearings after 2500 epochs

Table 5.4 shows average MAE, and average score for different configurations on all test bearings.

Configuration	Avg MAE(s)	Avg score
1F STFT	4965	0.214
10F STFT	5990	0.236
10FPC STFT	4850	0.080
1F CWT	5551	0.128
10F STFT CNN	5425	0.089

Table 5.4: Overall performance on test bearings after 2500 epochs

5.3 Expanded training data

The results for the different configurations are not impressive with respect to the obtained MAE for the various test bearings. The explanation for the limited performance may relate to the complexity of the various different bearing wear mechanisms and the limited amount of training bearings available. To investigate if increasing the set of training bearings will improve the performance a test using some of the test bearings for training purpose

was performed. In this test the 11 test bearings were split two groups of 5 and 6 bearings while the remaining test bearings were used for training together with the original 6 training bearings. The model used for the experiment was 10F STFT using the hyper parameters established with the standard 6 training bearings. The table 5.5 shows the resulting MAE for the test bearings when increasing the amount of train bearings. As can be observed MAE for the individual bearings are affected but the average MAE for all bearings remains almost constant. Table 5.5 show MAE loss with and without expanded training data.

Test bearing	MAE (6 train)	MAE (11 train)	MAE (12 train)
Bearing1_3	11319	x	6174
Bearing1_4	15482	x	5176
Bearing2_6	826	x	1032
Bearing2_7	3910	x	10755
Bearing3_3	678	x	1928
Bearing1_5	7920	9999	x
Bearing1_6	10133	8577	x
Bearing1_7	3316	3573	x
Bearing2_3	2911	4855	x
Bearing2_4	5309	3945	x
Bearing2_5	4094	7847	x
Avg MAE	MAE (6 train)**5990**	MAE (11+12 train)**5805**	

Table 5.5: MAE of expanded training data

6. Discussion and conclusion

Different pre-processing techniques and neural network models are in this thesis rated by both MAE and the score function from the prognostic challenge. In general the correlation between MAE and score is poor. Consequently, high score does not prove an accurate overall performance. Figure 5.3 of Bearing1_5 is an example. Generally the predicted and actual RUL points specified in the prognostic challenge matches significantly better than at an average point of time. This may partly relate to that the RUL points specified in the prognostic challenge are usually located in later stages of the bearings' life. The average MAE is also to some extent dominated by bearings with long life.

The individual bearing predictions are similar for the different configurations (pre-processing techniques and network models). Predictions across all configurations are shown in Appendix 6.1. The similarity in predictions between the various configurations indicates that the different networks find similar solutions trying to match the vibration spectrum with RUL.

Most predictions follows a monotonic downward RUL trend, even in areas where the frequency information does not seem to evolve, example figure 5.8. This is possibly due to a feature learned by the RNN network, that RUL decreases with time steps and not really derived from the specific vibration frequency spectrums.

All training bearings show a sudden rise of broad-band vibrations close to EOL. The same phenomena is not present with the test bearings except for bearing1_4. It is interesting to observe that the prediction of test bearing1_4 reacts to the sudden broad band vibrations close to EOL, while the prediction on the the other test bearings at EOL do not have a similar sudden change.

The difference in EOL behaviour between the training bearings and the test bearings is surprising. It may indicate that the running conditions have not been equivalent, or a different bearing batch has been introduced, or that the bearings have not been grouped randomly between train and test bearings in the challenge.

For neural networks to predict RUL effectively based on vibration data only there needs to be a connection between the vibration data magnitude or frequency content and RUL in some form. Assuming the required connection exists, a network will also need training data that represent the various wear and defect mechanisms in order to generalize.

The RUL predictions obtained using the various pre-processing techniques and neural networks may indicate limited connection between the vibration data and RUL. The results indicate that the improved frequency resolution obtained using 10F STFT, 10FPC STFT and 1F CWT does not improve the predictions over 1F STFT. This can be interpreted as if there is limited correlation between the detailed bearing vibrations frequencies (eq 2.3-2.6) and RUL.

From the validation process it became clear that the unique characteristics of bearing 1_1 not present in the remaining training data were problematic for the network. Assuming significant vibration spectrum and RUL correlation exists in the data set, it is reasonable to expect improved test bearing MAE performance when increasing the training set. Expanding the training set from 6 to 11/12 did however not yield any clear improvements in MAE in this data set.

As stated by the pronostic challenge the bearings do not follow the theoretical frameworks and L_{10}-life. It is therefore possible that the bearings in the prognosis depict different random frequency behaviours without specific recognizable patterns. Similar RUL studies on bearings subject to normal wear patterns may hopefully indicate a stronger correlation between vibration spectrum and RUL.

Further improvements of the network could be achieved by better use of vertical vibrations as the network is potentially missing data. There are also indications that training separate networks for each operational condition could yield better results. A reason for this hypothesis is that the spectrograms for the individually operational conditions are more similar.

The indication is further supported by the fact that both winners of the prognostic challenge made separate models for the individual operational conditions. Alternatively, the frequency spectrums could be normalized by the shaft frequency for the operational conditions or the condition could be encoded into the model directly. An example is adding a particular value as a feature into the LSTM cell for each case.

6.1 Conclusion

This thesis demonstrates alternative pre-processing techniques to achieve spectrograms with improved frequency resolution to be used for bearings RUL prognosis. The techniques utilized include FFT of merged files, FFT of merged files in phase and wavelet transforms. Finally the effect of the different pre-processing techniques were tested on different neural network models trained to estimate RUL. The findings indicate that the improved frequency resolution did not improve the RUL prognosis for the bearings used in the thesis. The bearings utilized in the thesis had not been degraded in a manner that represent normal wear of bearings and did not follow the theoretical frameworks and L_{10}-life. The pre-processing techniques and neural networks tested may possibly be more effective assuming normal wear conditions. Accurate RUL predictions were found to be challenging with this dataset. The obtained score was comparable to other participants in the prognostic challenge.

Bibliography

[1] M. mechanics. (2018) 5 causes of motor failure and how to prevent them. [Online]. Available: http://www.maintenance-mechanics.com/5-causes-motor-failure-prevent/

[2] T. Lin, K. Yu, and J. Tan, *Condition Monitoring and Fault Diagnosis of Roller Element Bearing*, 05 2017.

[3] D. Wang and K. Tsui, "Statistical modeling of bearing degradation signals," *IEEE Transactions on Reliability*, vol. 66, no. 4, pp. 1331–1344, Dec 2017.

[4] S. Porotsky and Z. Bluvband, "Remaining useful life estimation for systems with non-trendability behaviour," in *2012 IEEE Conference on Prognostics and Health Management*, June 2012, pp. 1–6.

[5] E. Sutrisno, H. Oh, A. S. S. Vasan, and M. Pecht, "Estimation of remaining useful life of ball bearings using data driven methodologies," in *2012 IEEE Conference on Prognostics and Health Management*, June 2012, pp. 1–7.

[6] J. S. Lal Senanayaka, H. Van Khang, and K. G. Robbersmyr, "Autoencoders and recurrent neural networks based algorithm for prognosis of bearing life," in *2018 21st International Conference on Electrical Machines and Systems (ICEMS)*, Oct 2018, pp. 537–542.

[7] D. Bogh, J. R. Crowell, and D. Stark, "Bearings for ieee 841 motors," *IEEE Transactions on Industry Applications*, vol. 39, no. 6, pp. 1578–1583, Nov 2003.

[8] M. mechanics. (2017) What is l10 life and why does it matter? [Online]. Available: https://www.linearmotiontips.com/what-is-l10-life-and-why-does-it-matter/

[9] Bruel and Kjær. Detecting faulty rolling-element bearings. [Online]. Available: https://www.bksv.com/media/doc/BO0210.pdf

[10] S. McInerny and Y. Dai, "Basic vibration signal processing for bearing fault detection," *Education, IEEE Transactions on*, vol. 46, pp. 149 – 156, 03 2003.

[11] ASTbearings. Radial ball bearings - overview. [Online]. Available: https://www.astbearings.com/radial-ball-bearings-overview.html

[12] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, "Pronostia: An experimental platform for bearings accelerated degradation tests," 06 2012, pp. 1–8.

[13] T. I. R. Society and F.-S. Institute. (2012) Ieee phm 2012 prognostic challenge. [Online]. Available: https://github.com/Tommylushixiang/phm-ieee-2012-data-challenge-dataset/blob/master/IEEEPHM2012-Challenge-Details.pdf

[14] Understanding and diagnosing the three types of imbalance. [Online]. Available: https://www.skf.com/group/services/services-and-solutions/introduction-to-condition-monitoring/understanding-and-diagnosing-three-types-of-imbalance.html

[15] M. Cohen, "A better way to define and describe morlet wavelets for time-frequency analysis," *NeuroImage*, vol. 199, pp. 81–86, 2019.

[16] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *CoRR*, vol. abs/1803.01271, 2018. [Online]. Available: http://arxiv.org/abs/1803.01271

[17] K. Cho, B. van Merrienboer, Çaglar Gülçehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *CoRR*, vol. abs/1406.1078, 2014. [Online]. Available: http://arxiv.org/abs/1406.1078

[18] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: continual prediction with lstm," in *1999 Ninth International Conference*

on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470), vol. 2, Sep. 1999, pp. 850–855 vol.2.

[19] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," *30th International Conference on Machine Learning, ICML 2013*, 11 2012.

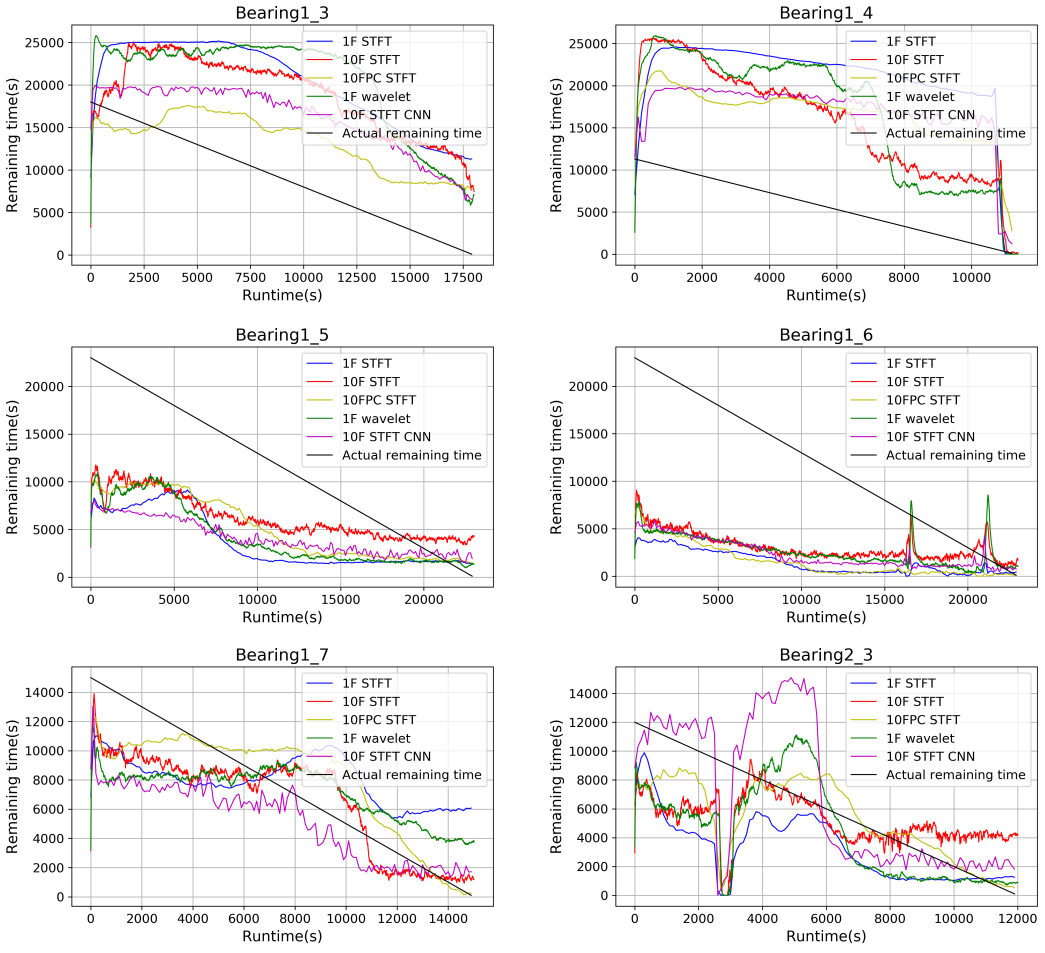
[20] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *CoRR*, vol. abs/1502.01852, 2015. [Online]. Available: http://arxiv.org/abs/1502.01852

[21] B. Ding, H. Qian, and J. Zhou, "Activation functions and their characteristics in deep neural networks," in *2018 Chinese Control And Decision Conference (CCDC)*, June 2018, pp. 1836–1841.

[22] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 12 2014.

[23] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

Appendix


Bearing1_3
Bearing1_4
Bearing1_5
Bearing1_6
Bearing1_7
Bearing2_3
1F STFT
10F STFT
10FPC STFT
1F wavelet
10F STFT CNN
Actual remaining time
Remaining time(s)
Runtime(s)

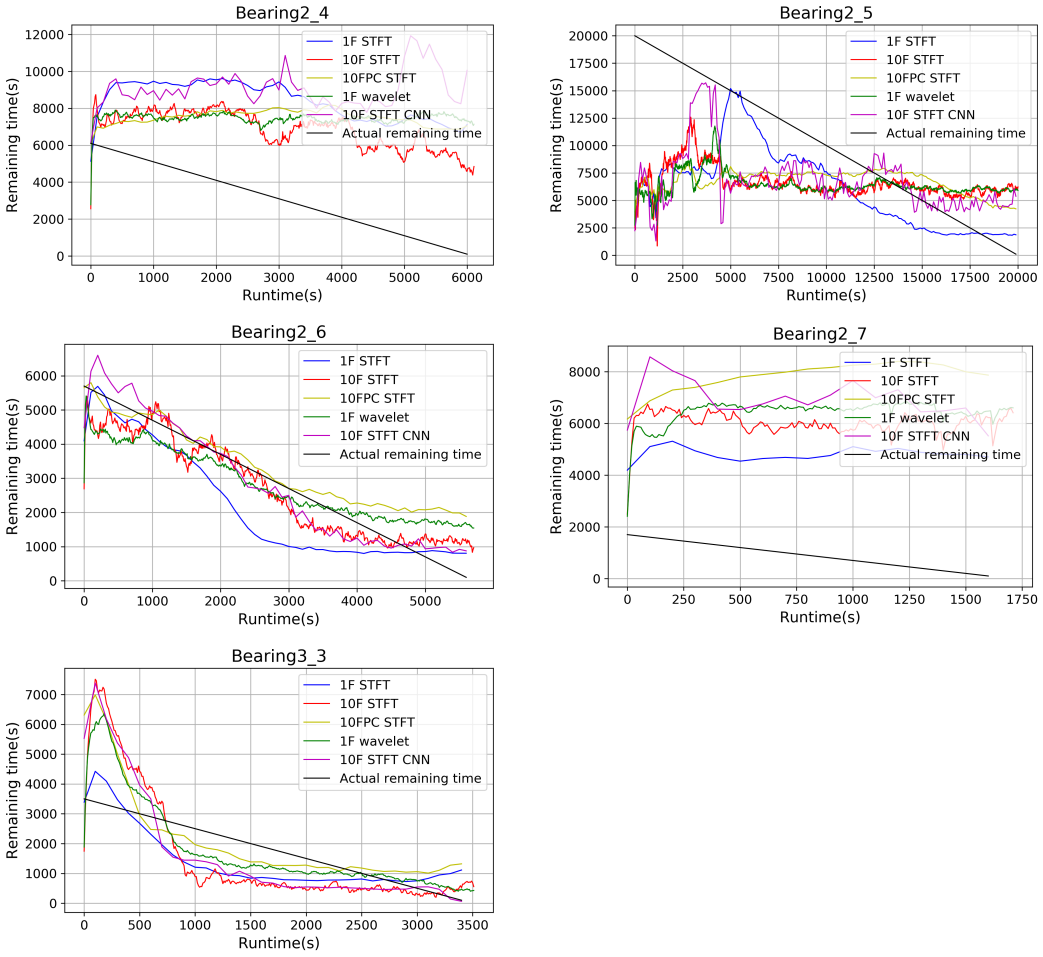


Figure 6.1: Average MAE on test bearings for all configurations

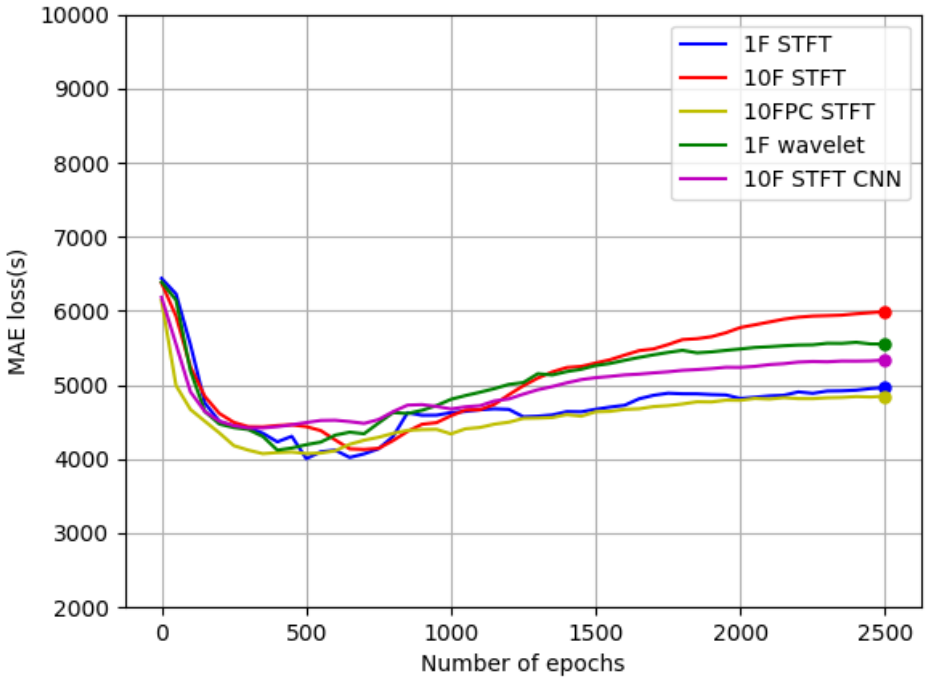


Figure 6.2: Test MAE loss for different configurations as a function of training epochs