Comparative Study of Fault Detection and Diagnosis for Low-Speed Ball Bearings

Master Thesis Embedded Systems

Mike Veltman
Dept. of Pervasive Systems
Faculty of Electrical Engineering, Mathematics and Computer Science
University of Twente
Enschede, The Netherlands
m.veltman-1@student.utwente.nl

***Abstract*— Fault Detection and Diagnosis (FDD) in Low-Speed Ball Bearings (LSBBs) is vital for ensuring the reliability and performance of radar systems. This comparative study examines approaches for FDD of LSBBs by collecting vibration and acoustic emission data using sensors on a test bench containing artificially induced bearing defects. Collected data was processed using pre-processing and feature extraction methods, the performance of which was evaluated using Random Forests and Principal Component Analysis.**

Results indicate that vibration sensing, at a 40 Hz sample rate with one sensor, conveys more information about LSBB defects than acoustic emissions measured between 100 and 450 kHz. Using vibration analysis, defects were successfully detected and identified. While all tested pre-processing methods performed comparably, root mean square and peak frequency magnitude were found most informative for feature extraction on bearing defects. The best-performing combination of methods was matched filter pre-processing combined with root mean square feature extraction. Using lower sample rates and fewer vibration sensors offers potential cost savings and increased computational efficiency.

Important to consider is the use of test bench data with artificial defects, which may not fully represent real-world radar systems. Therefore, data from real-world systems and bearing statuses is necessary to confirm the generalizability of the recommended approach. Despite these limitations, the findings provide valuable insights into FDD for LSBBs in radar systems, as well as in similar systems. These insights contribute to the improvement of radar system maintenance and reliability and ultimately to a safer, more secure and more efficient maritime environment.

***Index Terms*—fault detection and diagnosis, low-speed ball bearings, sensing techniques, pre-processing, feature extraction**

I. Introduction

Without bearings, many systems and machines would not be able to function. Ensuring the proper functioning of these crucial components is important in a world where industries aim for maximum efficiency, uptime and safety. Because the antennas of radar systems used for maritime traffic services rotate, bearings are a critical component. Sudden bearing failure leads to costly downtime, resulting in dangerous situations. By exploring the field of predictive maintenance, and in particular that of Fault Detection and Diagnosis (FDD) of Low-Speed Ball Bearings (LSBBs), this research addresses the concern of unexpected failure.

This research contributes to the field of predictive maintenance through the analysis and comparison of sensing techniques, pre-processing methods and feature extraction methods. This comparative study aims to find and recommend the best approach for FDD of bearing faults. The research is built on a dataset acquired from a test bench with various bearing defects running at low speeds, simulating the behaviour of LSBBs in the turning unit of a radar system.

A. Background & Context

For over 30 years, Saab Technologies B.V. has specialized in deploying radar systems worldwide to assist in maritime traffic services. Maritime traffic services aim to ensure safety, security and efficiency of transport over water [1]. Radar systems detect and identify vessels to enable tracking.

To capture radar video footage of the water surface, an electric turning unit rotates the antenna at a constant speed of 15 or 20 rpm, depending on the system requirements. The turning unit is a rotary torque motor, also referred to as the direct drive, with two bearings attached to a vertical shaft. Bearings are fundamental in rotating machinery because they allow for smooth rotation of a component. A cross-section sketch of the radar system is shown in Figure 1.

Ball bearings consist of the following components:

- Steel balls, to reduce rotational friction and support radial and axial loads.
- Inner and outer raceway with a groove, to offer the steel balls a fixed track to roll in.
- Lubricant, to reduce friction and wear, and to protect against corrosion and contamination.
- Cage, to keep the steel balls in place.
- Sealing/shielding, to preserve the lubricant inside the bearing and protect against contamination.

The bearings used in some radar systems at Saab Technologies B.V. are 6022-2RS1 bearings from SKF [2]. A partial

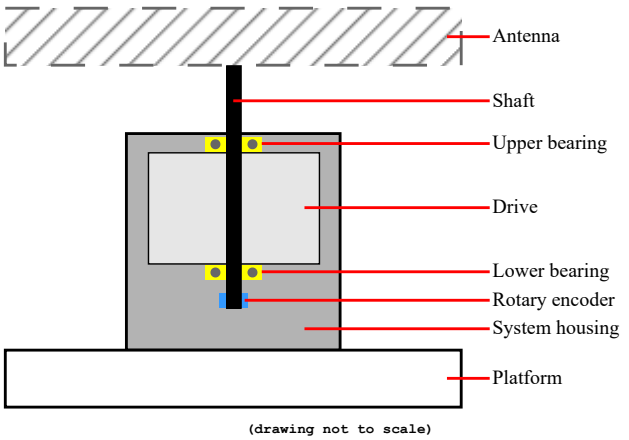


Fig. 1: Cross section of the main components of the concerned radar systems.

cross-section of the bearing is shown in Figure 2, displaying the aforementioned components. These bearings are single-row deep groove ball bearings and are able to accommodate loads in both axial and radial directions.

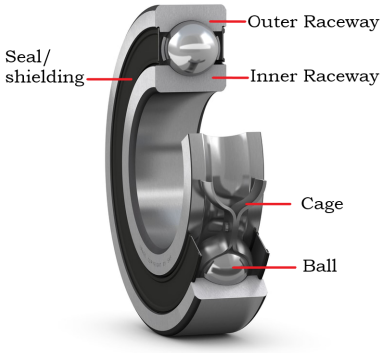


Fig. 2: Deep groove ball bearing with a seal. Adapted from [2].

Bearings are responsible for 40% to 70% of failures in rotating machinery [3]–[5]. About one-third of bearings fail due to fatigue, another third due to lubrication problems and one-sixth as a consequence of contamination [6]. Improper handling, mounting or loading of the bearing generally causes the remaining failures.

The radar systems are required to operate 24/7 and to have a lifespan of 10 years or more, with minimal maintenance, while enduring harsh environmental conditions. Environmental conditions include salt, sand or chemicals in the atmosphere, wind gusts, lightning strikes, humidity or exposure to water and extreme temperature fluctuations. These environmental conditions negatively influence the longevity of a radar system and in particular its bearings.

Repairing a radar system requires planning because parts that can break down when a failure occurs are seldom in stock, such as the drive. To make things more complicated, radar systems are usually located on remote sites on top of a tower, meaning the broken-down system has to be hoisted to ground level for repair. Equipment has to be rented and personnel, a scarce asset at the time of writing, have to be available.

Downtime of a radar system means that surveillance of maritime traffic services may not cover a specific area, increasing the chance of catastrophic accidents while also reducing search and rescue capabilities on the sea. The components requiring replacement following a failure are expensive. Coupled with the associated costs of equipment rental and hiring personnel for system repairs, the financial consequences of bearing failure are emphasized. Therefore, the importance of predicting maintenance for the bearings of turning units in radar systems cannot be overstated.

Predictive maintenance is one of four types of maintenance commonly found in literature [7], [8]:

- Reactive maintenance, at defect occurrence.
- Preventive maintenance, periodically scheduled.
- Condition-based maintenance, performed according to monitored defect features.
- Predictive maintenance, predicting defect occurrence based on monitored defect features.

These types of maintenance can also be considered as steps; where at every subsequent step maintenance is planned earlier, reducing the overall downtime of a system by preventing a catastrophic failure. The currently applied maintenance strategy is reactive maintenance with planned maintenance once after ten years. The status of bearings is currently not monitored and maintenance is performed when a failure occurs.

Predicting defect evolution in bearings is complex due to nonlinear development influenced by various factors and the low amount of energy released at early stages. The P-F curve, shown in Figure 3 illustrates the degradation of a bearing over time, where the P-F interval represents the time from detecting a Potential failure (P) up to the Functional failure itself (F). The P-F curve highlights the complexity of prediction because it shows that at any point in time, a potential failure can occur and the time to actual failure is not constant or certain. The aim of predictive maintenance is to extend this interval while accurately estimating a bearing's lifespan [9].

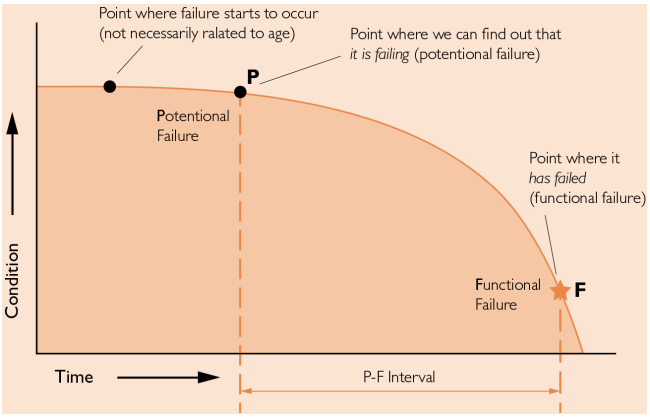


Fig. 3: P-F Curve. From [10]

B. Problem statement

A reactive maintenance strategy is currently employed, despite the critical importance of bearings in radar systems and the high costs and potential risks resulting from bearing failure. A more proactive strategy for bearing maintenance is necessary, given the harsh environmental conditions a radar system is exposed to and the unpredictable nature of bearing defects.

At present, Saab Technologies B.V. has no insight into the status of the LSBBs and does not collect relevant data. Although a predictive maintenance strategy is effective in mitigating the risk of unexpected failures, it requires the availability of historical data on the system behaviour, failure modes and failure causes.

Based on the aforementioned background, context and problems, the following problem statement has been derived:

The unpredictable nature of imminent bearing failure of LSBBs in radar systems being exposed to harsh environmental conditions, combined with a reactive maintenance strategy with no insight into the bearing status, results in high costs and safety risks when a failure occurs.

C. Research Objective

To prevent high costs and safety risks due to bearing failure of LSBBs in radar systems, one would ideally transition from a reactive to a predictive maintenance strategy. Rome was not built in a day, and likewise, a condition-based maintenance strategy precedes a predictive strategy. That is why this research focuses on condition-based maintenance by studying FDD for LSBBs.

The goal of this research is to collect, process and analyse data, compare analysis results and give recommendations for an effective sensing and processing combination for FDD of LSBBs in operation under both axial and radial loads. This goal is accomplished by exploring, analysing and comparing the effectiveness of sensing techniques, pre-processing methods and feature extraction methods based on evaluation metrics from existing literature. In this study, the combination of sensing and processing is referred to as the approach.

The results will contribute to bearing FDD research of systems with similar properties. In particular, the results will be valuable to organizations that are aiming to reduce maintenance costs and maximize the uptime of their radar systems to ensure safety, security and efficiency in their respective field.

D. Research Questions

The context provided by the research objective, problem statement and the analysis of related work, enables the formulation of the research questions. Identifying the most effective approach to FDD for LSBBs is the focus of the main research question:

What is the most effective fault detection and diagnosis approach for low-speed ball bearings used in radar systems?

To answer the main question, it was split into multiple sub-questions (SQ):

SQ1 What evaluation methods and metrics can be used to assess and compare the effectiveness of different fault detection and diagnosis approaches?

SQ2 What are the common faults and failure modes of low-speed-ball bearings used in radar systems?

SQ3 Which sensing techniques are most effective for fault detection and diagnosis of low-speed ball bearings?

SQ4 Which pre-processing methods are the most effective in improving the quality of collected data for fault detection and diagnosis of low-speed ball bearings?

SQ5 Which feature extraction methods are the most effective for fault detection and diagnosis of low-speed ball bearings?

E. Outline

The following section is Section II, which identifies gaps and provides insights from related work. The processes of collecting, processing and analysing data are discussed in Section III, Section IV and Section V respectively. Next, Section VI presents the results of this study. In Section VII the results are interpreted before summarizing the key findings and conclusions in Section VIII. Finally, Section IX gives recommendations and suggests future research possibilities.

II. Related Work

This section discusses the literature and previous studies in the context of FDD of LSBBs. Existing literature helped in formulating the sub-questions for this study. Through analysis of related work, SQ1 and SQ2 are addressed, covering evaluation methods and metrics used to assess different FDD approaches, as well as common LSBB defects in radar systems.

A. Defects

To understand how sensing techniques can be used to detect bearing defects, what can be measured on defects is first discussed.

Faults serve as a general term including not only defects but also any other system malfunction or failure. For instance, defects such as cracks in the bearing can lead to a fault, eventually resulting in a failure. The different types of defects can be categorized into Single Point (SP) or Generalized Roughness (GR) defects [4]. These types of defects are caused by various factors [11]–[13].

An SP defect, also known as spalling, is generally caused by improper handling before installation or electrical current passing through the bearings from lightning.

The GR defect is also known as general wear or micro-spalling. GR defects result from metal-to-metal contact, caused by misalignment, or lubricant that is starved or contaminated. GR defects generally lead to an increase in friction. The area around indentations in the raceways, created by over-rolled contaminants, becomes subject to cyclic stress resulting in surface fatigue and metal breakage. From this point damage progresses [14].

Both types of defects can be caused by misalignment of the shaft and excessive loads. Even though radial bearings that are analysed in this study can also handle some axial load, such load remains a weak point [6].

Because all rotating machines vibrate, mechanical problems are generally accompanied by an increase in vibration levels [6]. Defects are usually first manifested in changes in peaks and later in the overall energy of vibrations [15]. High and low-frequency acoustic emission waves in the ultrasonic spectrum travel through a structure when a defect forms or increases in size [16]. Due to increased friction from a bearing defect, an increase in temperature, sound noise levels and stator current may occur [17]. Lastly, temporary rotational resistance is caused depending on the size of a spall. This may result in an inconstant rotational speed [18].

B. Sensing Techniques

Vibration analysis is the industry standard for bearing FDD [17], [19]. This is due to its proven sensitivity to detect defects from vibration patterns and non-intrusive nature allowing real-time monitoring.

The existing vibration datasets mentioned before are generated by three accelerometers in orthogonal axes (XYZ). These sensors commonly measure vibrations between 0 Hz and 20 kHz. The literature did not mention the accuracy of their results when using less than three sensors. This either indicates that all three dimensions are equally important for accurately characterizing the defect vibrations, or that this could be an area of study with an aim to reduce the cost and complexity of analysis. However, one paper did mention researching lower sample rate accuracy for low-speed systems to reduce computational time and memory space [20]. A lower sample rate might prove beneficial for radar systems in remote locations where the data bandwidth to a base station is limited.

Acoustic Emission (AE) analysis is another, but less common, sensing technique measuring elastic waves in the ultrasonic spectrum [15]–[17], [21]. High-frequency AE waves are generally more sensitive to small developing defects, whereas low-frequency AE waves can propagate further and may be more effective at detecting larger defects and the current status of the system.

Piezoelectric sensors attached to a system are able to detect these high-frequency AE waves above 100 kHz [15]. These waves are generated due to changes in the material properties of a bearing, such as when a defect forms or develops [16]. For early detection of a defect to increase the P-F interval, detecting high-frequency AE waves may prove more effective than vibration analysis for LSBBs [16], [22]–[24].

Ultrasound microphones are able to detect low-frequency AE waves in the structure or in the air, depending on the noise and strength of the wave from a defect. Ultrasound sensing may bridge the gap between an AE sensor and vibration sensor working ranges between 20 kHz and 100 kHz [15]. These microphones can both be placed on the monitored structure or more conveniently at a distance such that the waves are transferred via air to the sensor.

Whereas papers based on vibration analysis often only look at an increased value in time domain features, AE analysis may be able to indicate the defect type and size better [21]. Another advantage of AE analysis is its high SNR in a noisy environment [17]. Disadvantages are that an AE sensing system is expensive, produces vast amounts of data because of its high sample rate and requires specialized expertise for setup and management [15], [17]. Two papers researched both AE and vibration analysis but did not compare the performance of both techniques [21], [25].

Other ways of monitoring a bearing are via stator current, bearing temperature, lubrication status or the rotary encoder instantaneous acceleration [13], [17], [25], [26]. Specifically for the direct drive which does not contain a gearbox, monitoring the stator current might prove useful. By increased friction due to a defect, different power usage may be required as opposed to a bearing in perfect conditions. Another paper looked at the effectiveness of using the rotary encoder for bearing FDD, but did not compare its performance to more common sensing techniques such as vibration analysis or AE [18].

Combining data from several sensing techniques can prove useful, as some techniques are better at indicating certain specific defects than others. An example of this is that the stator current might be able to detect a bearing cage defect, in contrast to a rotary encoder-based sensing technique [5], [18].

C. Datasets

Lack of data is a common theme in the field of predictive maintenance for bearings. Published papers have been re-using already existing run-to-failure datasets from test benches [3], [21], [27]–[33]. The issue with these datasets is that the test benches on which the data is collected vary greatly from the properties of the radar systems with LSBBs. One paper also expresses concerns that most research is based on the same dataset [7]. Common properties of existing datasets researching predictive maintenance or FDD of bearings are:

- Vibration data, collected by three orthogonally placed accelerometers.
- Lab environment with minimal environmental noise, but with mechanical noise from a gearbox.
- Artificially introduced SP defects.
- High rpm, often 2,000 rpm or higher.
- Radial loads, due to a horizontally placed shaft.
- Loads between 4 and 70 kN.

Because the lifespan of a bearing is longer than the scope of most research projects, degradation is accelerated by operating at a high rpm and introducing artificial defects. Artificial defects are usually SP defects, whereas another realistic defect type in a radar system is a GR defect [4], [6], [11]. Developed methods may not perform as well with data collected by field experiments where a bearing has undergone realistic wear.

The radar systems which this research focuses on, do not contain a gearbox and rotate at low rpm (15 or 20) with low mixed direction loads (2.5 kN from the antenna weight). Wind gusts generate random changes in load in both axial and radial directions. This means that the high rpm and high radial load

of the laboratory datasets are different from the speeds and loads in radar systems. The load zone of a horizontal shaft of existing datasets is more predictable [16]. This difference in properties might render existing methods less effective. At lower rpm and lower load, less energy is released making the signal-to-noise ratio (SNR) low and a defect harder to detect, especially in its incipient stage [12], [23].

One solution to a lack of data is to create a realistic model of a system and simulate it [25], [34]. This requires advanced mechanical engineering insights and expertise, which is not widely available. Papers that use generated data from simulations recommend acquiring data from a real-life scenario [7], [25], [34].

Some papers do collect data from field experiments, like a wind turbine [20], [26]. But the approaches in these papers then lack pre-processing, use a different type of bearing such as a slewing bearing, have higher speed or load, varying rpm, or already have an advanced model available simulating the system to train ML algorithms [4], [34]

D. Pre-Processing

The purpose of pre-processing is to enhance the ability to detect a defect by either transforming data, manipulating data or both.

Transforming data is a pre-processing type where data is converted between domains with minimal loss of information. This means that a piece of data can be transformed back to its original form and still being mostly similar. One example is transforming angular positions to angular velocity or acceleration.

Another example is by transforming data between time, frequency or time-frequency domain. Between the time and frequency domain, the transformation can be done by applying a Fourier Transform (FT), or its inverse. Some time-frequency transformation methods include Short-Time Fourier Transform (STFT) or the Wavelet Transform (WT) [11], [35]. STFT allows for a trade-off between time and frequency resolution by analysing a signal in small windows of time, whereas Wavelet Transform uses a variable time-frequency resolution that depends on the scale of the wavelet function.

Manipulating data is commonly used to remove noise. Existing literature applies filters, such as a high-pass filter, low-pass filter, band-pass filter, wiener filter or matched filter in order to remove noisy frequencies [13], [16], [36], [36], [37]. Basic averaging of multiple signals was also applied using a tachometer pulse to indicate the start of a rotation [14].

FDD techniques often use the frequency domain for analysing a bearing but can suffer from frequency smearing, which happens when the rotational speed fluctuates too much. A promising pre-processing method against frequency smearing is Time Synchronous Averaging (TSA) [22], [34]. TSA can be applied to a measurement where tachometer pulses were recorded in parallel to compute the instantaneous angular velocity. This method looks to be promising against frequency smearing, especially in the frequency domain [38].

Other advanced pre-processing methods, like wavelet denoising, were also proposed [30]. Wavelet denoising is used in image and audio processing but can be applied in the analysis of vibration signals for bearings for removing unwanted noise. One paper effectively applies wavelet denoising before applying TSA [39]. Applying order tracking has been looked at by [11], but was mainly used to separate interference from the gearbox for identifying a defect which a direct drive does not contain. Nevertheless, related work usually sticks to more basic forms of pre-processing to remove gearbox noise or applies no pre-processing at all.

By not performing pre-processing some publications utilize raw data, meaning that the data potentially contains unwanted noise [3], [31]. One paper states that pre-processing is important to increase the accuracy and timeliness of FDD [35].

E. Feature Extraction

Feature extraction is the process of deriving descriptive defect information from a large dataset. The purpose of feature extraction is to make the data easier to be interpreted by a human, algorithm or ML model, while still containing relevant information to what it tries to describe. This is called dimensionality reduction. The benefits of dimensionality reduction by using features as training data for an ML model are increased accuracy, lower complexity and less computation time for training and classifying.

Vibration signals have a high dimensional nature, making the feature extraction stage after the pre-processing stage necessary, and is often seen as the most crucial part of FDD [11], [31], [40]. Papers that apply feature extraction to signals from sensing techniques mentioned in Section II-B, do so in the time, frequency or time-frequency domain [19].

Most common features used in related work in the time domain are for example RMS, kurtosis, skewness, peak value and crest factor [13], [16], [21]. Frequency domain features applied in literature are characteristic defect frequencies, peak frequency and peak frequency magnitude [5], [11], [20], [36], [41]. Time-frequency domain analysis is usually done by eye or image processing. Most time-frequency domain features produce multi-dimensional plots that result in information loss [42]. More advanced feature extraction methods are for example entropy-based, measuring the randomness or disorder in a signal [15], [20].

Often, papers try to perform real-time diagnosis [27], [41]. Increasing diagnosis speed comes at the cost of the quality of the diagnosis itself. The low speed and load nature of a radar system makes real-time diagnosis unnecessary.

F. Data Analysis

Reviewed literature uses Random Forests (RF) to classify defects or to compute the Remaining Useful Life (RUL) of a bearing based on historical data [27], [34], [43]. RF can also be applied to determine the best pre-processing and feature extraction method. This can be done by looking at how much a certain combination of pre-processing method and feature

extraction method was useful for determining a defect type, also known as the predictor importance.

Principal Component Analysis (PCA) is a method of linear dimensionality reduction and, like RF, can also be used to indicate feature importance. In related work, it has been used mainly for dimensionality reduction where ML was applied for computing the RUL [42], [44].

Support Vector Machines (SVM) and Autoencoders are additional methods that could be used for classification, regression, and dimensionality reduction. However, they may not be as effective for predictor importance as RF or PCA. SVM is primarily a binary classifier and does not inherently provide measures of feature importance [44]. Autoencoders, on the other hand, are unsupervised neural networks used for dimensionality reduction or denoising which do not naturally indicate feature importance either. [29], [42]

G. Predictive Maintenance

Currently, existing work predicts maintenance for bearings based on estimating the RUL of a bearing, which is the time left until the bearing reaches a critical point at which it breaks [3], [30], [42], [45].

The RUL can be computed by training a model on large amounts of historical data of multiple defects. In literature, there are three main categories of RUL prediction: [7], [45]:

- Physical model-based: a physical model of the system is created and used to predict its behaviour by simulating different operating conditions.
- Knowledge-based: expert knowledge and experience applied to observations and analysis of previous failures and then used to identify possible defects in the system.
- Data-driven: statistical models and machine learning algorithms are used to analyse large amounts of generated data from the system.

The data-driven category is useful when there is a large amount of data available and when there are no physical models or expert knowledge available to understand the system's behaviour.

The data from the data-driven category comes from an implemented FDD approach. This data is collected over time and contains the development of features leading to different defects. The health indicator can be constructed based on this FDD data and is commonly used to train a RUL prediction model on [30], [33], [37], [44]. To construct the health indicator, a model needs to be trained on FDD data of multiple run-to-failure scenarios.

Machine Learning (ML) has gained popularity for RUL prediction, because of its ability to identify patterns in complex data. Techniques like recursive neural networks (RNN), deep neural networks (DNN), and long short-term memory networks (LSTM) are commonly used, with LSTM showing promise due to its ability to manage long-term dependencies [31], [42].

Using ML for computing the RUL does come with challenges, such as optimal feature selection and efficient feature compression. A recent paper manually selected features that are used to predict the RUL out of a larger set, instead of selecting the optimal features based on a method that is more prone to human error [44]. Solving these challenges prevents information redundancy and thus prevents model training difficulty and overfitting [42]. Some papers state that their RUL prediction approach is promising, but fail to compare their results to others [43]. Another problem existent in RUL prediction is being too optimistic in predicting, and therefore, too late [44]. It is more desirable to have a pessimistic RUL to replace a bearing before it breaks down.

The main gap in related work is that the most suitable sensing techniques, pre-processing and feature extraction methods for FDD of LSBBs are not clear for a system with the properties of a radar system. To accomplish this, generating a new dataset is especially important.

The following three sections explain the process of going from research objectives to results. Acquisition of datasets with a test bench using sensors, pre-processing and feature extraction methods and analysis methods.

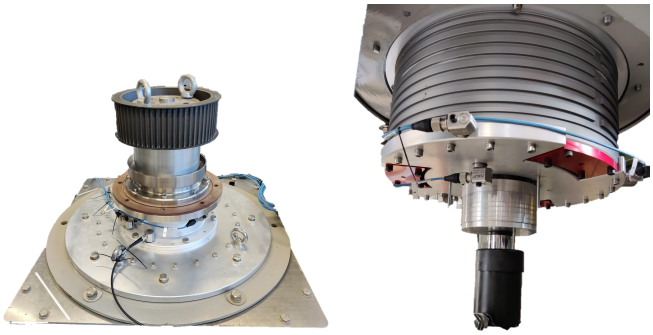
III. Data Collection

Considering the scope of the research project, simulating a radar system or acquiring data from an actual radar system was not possible. In cooperation with CHL Netherlands B.V., a test bench was adapted to collect representative data. The sensing techniques used to acquire data are discussed later in this section.

A. Test Bench

As seen in Figure 4, the main difference between the test bench and a real radar system is the lack of a 250 kg weighing and 6-meter long antenna. In Figure 4a it can be seen that a flanged pulley is located where normally an antenna would be placed. On Figure 4b the bottom side of the test bench, where the drive is located, is shown during the replacement of the bottom bearing. This difference was resolved by applying a pre-load, which is internal mechanical load, to the bearing. For example by tightening up the spring and bolts holding the bearing in place vertically by an extreme amount. Furthermore, the drive used in the test bench is the same type as is used in the radar systems, a DD2 direct drive. It was set up to rotate at 0, 15, 20 and 60 rpm. Noise from wind could not be simulated in the lab. A secondary drive and toothed belt could provide rotational resistance. However, it was decided not to use this method because it would generate substantial amounts of artificial unrealistic noise in the data.

The types of bearings defects that were simulated are a SP defect and GR defects. The SP defect was created by using heavy machinery to damage the surface of the outer raceway. Because the bearing could not be disassembled and assembled, the outer raceway was the only part of the bearing to which damage could be applied. An attempt to generate a GR defect was made by removing all lubricant from the bearing, causing more friction and metal-to-metal contact during rotation. By mixing sand from a CNC cutting machine with glue and injecting it into the bearing, the bearing was contaminated with



(a) Top side (b) Bottom side

Fig. 4: Test bench (pictures taken during bearing replacement)

particles, leading to two more datasets of a GR defect. One GR defect dataset was recorded 10 minutes after adding sand, the other after letting the system run overnight for 18 hours at a speed of 60 rpm. Data from a perfect bearing was also collected as a reference. The GR defect with sand is shown in Figure 5a, which indicates the groove in which the ball bearing ran and the contaminant that was pushed to the sides over time. In Figure 5b we can see the SP defect damage that was inflicted on the outer race.

To summarize, the bearings statuses of which data was collected are:

1) Perfect bearing
2) SP defect, local damage to the outer raceway.
3) GR defect, without lubricant.
4) GR defect, without lubricant with contaminant after 10 minutes of running.
5) GR defect, without lubricant with contaminant after 18 hours of running.

(a) GR defect with contaminant (b) SP defect

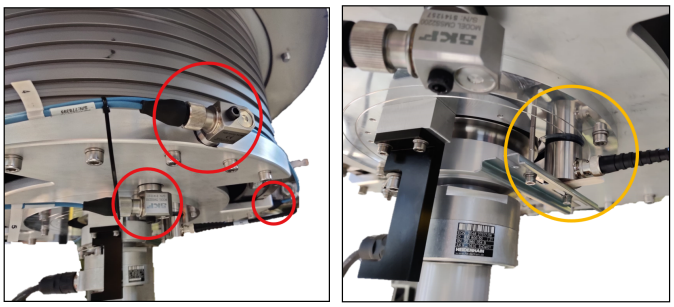
Fig. 5: Bearing defects

The lower bearing was replaced with a bearing containing the defects as it is the most accessible of the two. This is the reason the data analysis is applied to data acquired from the lower bearing.

Data was collected when the system was fully turned off and when the drive was set to rotate at 0, 15 and 20 rpm. The reason for collecting data when the system is turned off is to see what the default noise level is in the data. Data from 0 rpm can be useful because it might show the noise introduced by the drive without rotating. This noise can come from the motor controller (PID control) which uses power to keep the drive at 0 rpm. Both 15 and 20 rpm are speeds seen in radar systems for maritime traffic management. The remainder of this study focuses on the dataset collected at 15 rpm.

B. Sensing Techniques

As mentioned in Section II, vibration analysis is the industry standard, making it a straightforward decision to analyse this sensing technique. The other sensing technique to which it was compared is AE. Mainly because related work showed that it should be effective for LSBBs due to the low amount of energy present in the system at low rpm and load. The stator current and rotary encoder data have not been used as sensing techniques because of the influence of the PID control, affecting the collected data. Besides that, these techniques were difficult to integrate and collect data from within the scope of the research project. However, collecting stator current data from the DD2 might prove to be an effective sensing technique if the influence of PID control and environmental noise on the data are studied in more detail, to compansate for these influences with pre-processing.



(a) Three vibration sensors (b) One AE sensor

Fig. 6: Vibration and AE sensors on the bottom side of the test bench

Vibration data was collected using SKF's CMSS 2200 general-purpose sensor in combination with their IMx-8 Data Acquisition System (DAS) [46], [47]. The sensor is an accelerometer with a sensitivity of $100mV/g$, a ±3 dB frequency range of 0.7 to 10,000 Hz and a resonant frequency above 22 kHz. The vibration sensors could remain in place whenever changes to the system, such as replacing the bearing, were made. Three sensors were placed per bearing, as shown in Figure 6a, each in an orthogonal direction to measure all vibrations in all directions. The setup and configuration of the sensors were done using SKF's @ptitude Observer Monitor software. The X and Y-axis sensors, located in the radial direction of the bearing, were set up to collect data at a sample rate of 1,280 samples per second. In Figure 7 the time domain data for all measurements of the X-axis sensor can be seen. The recordings consist of 16,384 samples, equal to 12.8 seconds or a little over 3 rotations at 15 rpm. The vibration sensors do not have a sensitivity to a specific frequency.

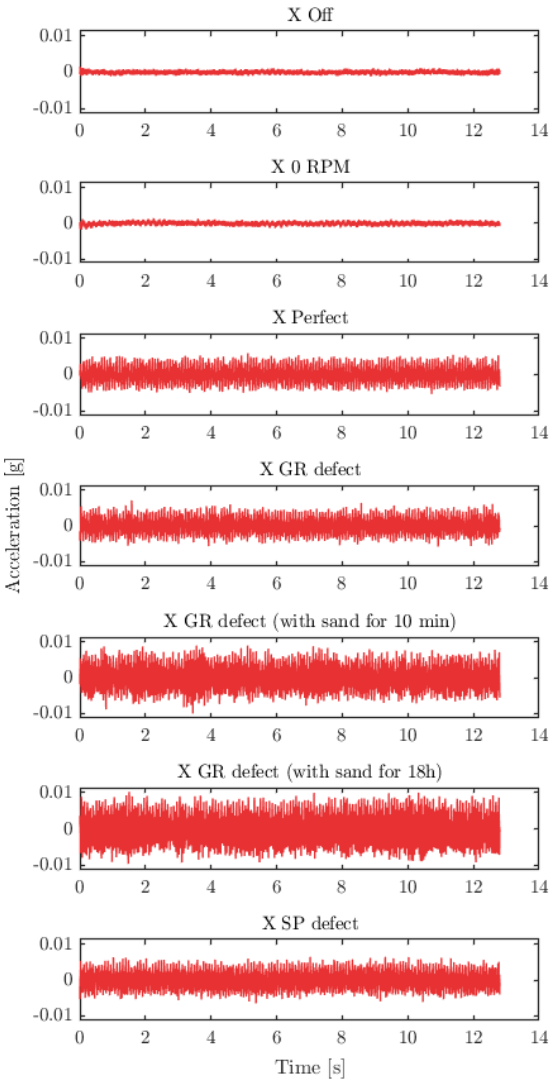


Fig. 7: Time domain data for all measurements of the X-axis sensor.

Changes in configuration negatively affected the usability of the Z-axis data. As a result, the analysis is focused on the X and Y-axis data, which were collected consistently throughout the experiment. Nonetheless, the collected data still provides valuable insights into the vibration characteristics of the bearings and makes a comprehensive analysis of the system's performance possible.

For logging the rpm together with the vibration data, an IG5597 inductive sensor was used and connected to the IMx-8 DAS. In the @ptitude Observer Monitor software the detection of a metal protrusion by this sensor was converted to rpm. The vibration monitoring software was configured to only capture vibration data whenever the rpm value was constant and not zero.

For collecting AE data, the VS150-RSC sensor in combination with the linWave 1002 DAS from Vallen Systeme GmbH was utilized [48], [49]. The VS150-RSC sensor is a bidirectional piezoelectric sensor with a built-in pre-amplifier, able to capture inaudible ultrasonic waves between 100 and 450 kHz. Its resonant frequency, where it performs best, is at 150 kHz. One side of the sensor contains a ceramic wear plate, over which silicon grease is smeared to act as a coupling agent. This makes sure that the transfer of the acoustic waves to the sensor is done with minimal loss. One sensor per bearing is sufficient to capture the information about bearing defects. Figure 6b shows one of the AE sensors connected to the bottom side of the bearing near the lower bearing.

The AE sensor had to be removed and placed again every time the lower bearing was replaced since it was connected to the structure holding the bearing in place. To make sure the process of placing the AE sensor did not affect the measurements a pencil break test was performed. Every time something changed on the test bench and at the start and end of a day of collecting data the calibration test was done. A pencil break acts as a stimulus for acoustic waves, standing out above the other measurements. Multiple pencil breaks are performed around the AE sensor for every calibration test, at about 5 cm. When the average results of a calibration test fall within 3 dB of one another, the sensor is as sensitive as it was in the previous calibration test. This was the case for all measurements.

The software used to acquire data from the linWave DAS is VisualAE. In this software, it is possible to set the configurations of the DAS and analyse the data. The main parameters to be configured are the high-pass and low-pass filter settings in kHz, the hit threshold, and the length of a measurement when the hit threshold was reached. This configuration collects the data continuously and only stores it when a hit threshold was reached, which is called a hit in AE terms. The configuration parameters used are HPF at 100 kHz, LPF at 600 kHz and the threshold at 40 dbAE with a reference voltage of $1\mu V$ referred to the pre-amplifier input U_r. The dbAE scale is used to measure the intensity of acoustic emissions in decibels because this is a way of expressing the ratio between a sound wave and electrical energy. Expressing intensity of sound in decibels based on voltage is computed using Equation 1. 0 dbAE is the intensity of sound when the voltage U is equal to the reference U_r of $1\mu V$, meaning 40 dbAE corresponds to $100\mu V$. With an HPF of 100 kHz, the direct drive control frequencies of 8 and 16 kHz are automatically filtered.

$$dBAE = 20 \log \frac{U}{U_r} \tag{1}$$

The AE sensing technique excels in detecting the occurrence or development of a defect, as opposed to vibration analysis, which is fit for detecting if a defect is currently present in the bearing. When an AE sensing technique did not detect any hits during a recording, it is safe to say no defect occurred or developed in that time period. Nothing can be said about the presence of a defect. The remainder of this study discusses the analysis of vibration data because AE data did not contain hits, as will be restated in Section VI.

IV. Data Processing

Due to the absence of hits collected in AE data, the focus of pre-processing and feature extraction is on vibration data. An attempt was made to apply these techniques to AE data too, which confirmed that no hits were collected and thus no information was overlooked.

A. Pre-Processing

Before data is pre-processed, it has been prepared by labelling the datasets per defect and sensor, making sure there are no empty values in a dataset and the alternating current hum of 50 Hz is removed by using a Butterworth notch filter. A Butterworth filter is also referred to as a maximally flat magnitude filter because it is designed to have a frequency response that is as flat as possible in the passband. In short, a Butterworth filter should have uniform sensitivity for wanted frequencies while completely rejecting unwanted frequencies. After preparing the datasets, three methods of pre-processing were applied individually, creating three additional datasets.

The first is a system filter, where the noisy frequencies of a system with a perfect bearing were analysed. Then using a Butterworth notch filter the noisy frequencies are attenuated. The frequencies that exist in the perfect bearing spectrum which are filtered out are 11, 22, 24, 33, 39, 44, 48, 61, and 77 Hz. Frequencies coming from the direct drive control are 8 and 16 kHz but are not recorded due to the sample rate of 1,280 Hz of the vibration sensors.

The second is a matched filter. This involves making a template from a certain window of the perfect bearing dataset and convolving its conjugated time-reverse with the unknown signal. The result is a signal with the signal-to-noise ratio (SNR) maximized when the template perfectly described the known unwanted signal. Such a filter is commonly used in radar systems for comparing the received signal with the known emitted signal. The template window size was set at the sample rate, meaning a template of one second.

Wavelet denoising is the third pre-processing method, concentrating signal features into a few large-magnitude wavelet coefficients. The smaller value coefficients are typically noise. They are removed without affecting the signal quality. After that, the remaining coefficients are used for the inverse wavelet transform to reconstruct the wavelet-denoised signal. The default level of wavelet decomposition is used, which is based on the size of the input dataset. For a dataset size of 5,120, the level is nine and for a size of 1,280, the level is seven. If computing the wavelet denoise takes too much time because the dataset is too large, the level of wavelet decomposition can be reduced to lower than the default. This was not necessary in this case.

Finally, the raw dataset without pre-processing was retained to compare to the pre-processing methods.

B. Feature Extraction

Selected features from the literature are extracted in the time and frequency domain. Changes in the health status of a bearing may be seen in changes in these features extracted from vibration signals.

The selected time domain features are now discussed with their respective formula. Where N is the size of dataset x over which the feature is extracted. In more detail, x_i is the individual data point in the dataset and $\overline{x}$ is the sample mean of the dataset.

The RMS value, calculated using Equation 2, is a common way to represent the effective amplitude of a vibration signal over a given time period, with higher values indicating higher levels of vibration or energy.

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}|x_i|^2} \tag{2}$$

The kurtosis value indicates the length of the tails of a signal distribution, calculated using Equation 3. An increase in the number of outliers can show the development of a defect.

$$K = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^4}{\left(\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2\right)^2} \tag{3}$$

The asymmetry of a signal distribution is indicated with the skewness value using Equation 4. The asymmetry can be influenced by a defect.

$$Sk = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^3}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2}\right)^3} \tag{4}$$

Equation 5 indicates the highest value of the input signal, which is the peak value. Defects are usually first manifested in changes in the peaks of a signal, and later in the overall energy of the signal.

$$PV = \max(\mathrm{x_i}) \tag{5}$$

When the peak value is divided by the RMS (level of vibration), an early warning of a defect can be seen. This is called the crest factor and is computed with Equation 6.

$$Crf = \frac{\mathrm{PV}}{\mathrm{RMS}} \tag{6}$$

Frequency domain features are also extracted, for example, the characteristic defect frequency amplitude. When an SP defect occurs, energy is released at a certain frequency when the rpm is constant. This frequency depends on the location of the SP defect (inner/outer raceway, ball or cage). The ball pass frequency of the outer race (BPFO), one of four bearing characteristic defect frequencies, can be calculated using Equation 7 [11]. Here N_b is the number of balls, B_d is the ball diameter in millimetres, P_d is the pitch diameter in millimetres and β is the contact angle in radians between the ball and bearing.

$$BPFO = \frac{N_b}{2}\cdot\frac{RPM}{60}\left(1+\frac{B_d}{P_d}\cdot\cos\beta\right) \tag{7}$$

Given that $N_b = 14$, $RPM = 15$, $B_d = 18.62$, $P_d = 139.54$ and $\beta = \frac{1}{4} \times \pi$, and given that the SP defect that is simulated with the test bench is an outer raceway defect, $BPFO = 1.58$ Hz was computed.

The peak frequency is another frequency domain feature, which gives the frequency at which the amplitude is highest. This feature may be interesting when a defect is present at a certain frequency, but not at the computed characteristic defect frequencies. The peak frequency magnitude is the magnitude at that frequency. The result of a defect, either SP or a GR defect, may be seen in a rise in the peak frequency magnitude. For example when there is metal-to-metal contact, the natural frequencies already present in a system with a perfect bearing will increase in amplitude.

Time-frequency domain features are not analysed. Because the time-frequency domain is displayed in an image-like plot and requires image processing to extract features. Commonly, a time-frequency plot is analysed manually by eye.

AE-specific features also exist, describing the properties of a specific hit. These features are dependent on the occurrence of a hit and can only then be extracted. Examples of these hit-specific features are: rise time, time duration, counts and shape.

V. Data Analysis

Vibration data was processed in MATLAB. Random Forest (RF) was chosen as the main analysis method for predictor importance, because of its computation efficiency and its ability to capture complex relationships between input data and a defect. Principal Component Analysis (PCA) was chosen as a means of validation. To not get confused with feature extraction methods, the input data for RF and PCA are called RF and PCA features, which are all possible combinations of pre-processing and feature extraction methods, labelled with the bearing status.

A. Random Forest

RFs can be used for classification problems, which apply to the current research, and regression problems. A classification problem is about predicting a discrete class label based on an observation. An observation contains a set of RF features, also known as the predictors, and a corresponding target variable, also known as the class label, that the model tries to classify. RF was used for determining RF feature importance, also known as predictor importance, by looking at how much an RF feature contributes to classifying a bearing status.

An RF model is trained on random RF features per observation in the training set and consists of multiple decision trees. A decision tree is a visual representation of a set of conditions that decide how to split RF features of observations into smaller subsets to classify a label. The classification of each tree is averaged to give the final classification of the RF model. The interaction curvature split prediction test was used for the model because it gives accurate unbiased predictor importance results. This test analyses the curvature of the predicted labels for each RF feature and then identifies which RF features have an impact on the model. The test is unbiased because it considers all possible split points meaning it is able to identify important predictors and interactions in the presence of many irrelevant ones. Split points are the RF features that are used during training to split a node in a decision tree.

To train the RF model on more than one value per vibration dataset, a moving window with a size equal to one rotation was used. This is equal to 5120 samples at 15 rpm. Larger moving window sizes did not affect the accuracy, which was already 100% at a window size equal to one rotation. All datasets of all bearing statuses are combined and shuffled and 80% of that dataset is stored as the training set, 10% as the testing set and the remaining 10% as the validation set. The datasets are balanced to prevent majorities or minorities from certain defects. Balancing was done by determining the amount of the smallest defect category samples and then reducing the number of samples in the other defect categories to match that amount.

The RF feature subset size on which the model was trained is the square root of the amount of the total amount of RF features: $\sqrt{64} = 8$. The minimum amount of leafs a tree can have is set at 1. Increasing the minimum amount of leaves has no impact on the accuracy. The max depth is configured by setting the max number of splits to seven to prevent the tree from overfitting. It must be mentioned that without a split number limit, most trees would not be larger than 15 splits. Classification accuracy is the percentage of accurate classifications made with the test set as input to the model. The amount of trees is set at 128, to give all RF features a chance to be used in a decision tree in order to extract their predictor importance. It must be noted that increasing the number of trees in the forest has little to no influence on classification accuracy and will result in higher training and computation time. It was observed that after 20 trees, the classification accuracy reaches 100%.

Out-of-bag predictor importance was applied to get the best-performing pre-processing method and feature combinations. The out-of-bag set consists of the RF features that were not used during training.

The predictor importance is computed by measuring the decrease in accuracy when predictions are made using out-of-bag samples while randomly permuting the value of a specific RF feature. RF features with larger decreases in accuracy indicate greater importance in the model's prediction process, providing insights into the relative significance of each feature without the need for a separate validation set. It is not possible to give predictor importance for one type of defect. The reason is that one RF feature that discriminates defect A against defect B is also doing the opposite, meaning that the importance of that RF feature is equal for both defects.

The RF models are trained for various sample rates on all combinations of pre-processing and feature extraction methods for the X and Y sensors individually and combined. 1,280 Hz is the highest sample rate that was analysed. Lower sample rates were analysed by down-sampling the datasets. The lowest sample rate for which the accuracy is still 100% is used for

further analysis of predictor importance.

During data analysis, it was observed that the overall RF predictor values rise when the training dataset size decreases, even with equal RF features. This may have to do with the dataset becoming less complex, resulting in decision trees becoming less robust with fewer split points. The decision trees then rely more on a subset of features for classification, leading to an increase in predictor importance values for those features. Relative normalization is applied to compensate for this effect by dividing the individual importance by the total importance for all pre-processing methods or all feature extraction methods. This ensures that the normalized values are expressed as a relative representation of the original values in terms of their contribution to the total.

B. Principal Component Analysis

As mentioned in Section II, PCA can be used not only for dimensionality reduction but also for indicating predictor importance. The PCA feature set is the same as for RF predictor importance because the RF training set was used for PCA analysis.

For dimensionality reduction, PCA extracts the most relevant information while minimizing the impact of less important PCA features. Principal components are new variables that represent linear combinations of the input PCA features. PCA obtains these principal components and ranks them based on the amount of variance they capture, which is indicated by their eigenvalues. A subset of the highest-ranking principal components can reduce the dimensionality of the data to simplify further analysis and computational efficiency. That is why dimensionality reduction is often used in ML applications.

For indicating predictor importance, the coefficients provide valuable insights. The coefficients reflect the relationship between the original PCA features and the principal components. The largest absolute coefficients suggest a strong influence of a particular PCA feature on the corresponding principal component. One can identify the PCA features that contribute the most to the principal components and thus see the relative predictor importance in the dataset.

Before applying PCA, the data has to be standardized because PCA is sensitive to the scales of the input features, unlike RF. Thus all features have to be in the same scale to prevent bias towards features with larger scales. This is done by subtracting the mean of each feature from its values and then dividing it by its standard deviation, also known as computing the z-score. After standardization, all columns with zero variance are removed to prevent the matrix from being linearly dependent.

MATLAB's PCA function from the Statistics and Machine Learning Toolbox was used. First, the coefficients for all PCA features were computed per principal component as a matrix. Each column contains coefficients for one principal component in the order of descending component variance. By taking the absolute of the coefficients and then taking the total of all principal components per PCA feature, the total weight per PCA feature is determined. The PCA features were sorted based on these total weights to identify the most important pre-processing method and feature extraction method combination, as was done with RF too.

To check if the PCA results are valid, the cumulative sum of the explained variance by each principal component is plotted. The explained variance indicates how much of the total variance in a dataset is explained by a particular PCA feature. A scree plot should have a clear elbow point that shows the optimal number of principal components, which it does as can be seen in Figure 8. In this context, the elbow point is the point where the explained variance begins to level off. The scree plot shows a high proportion of the total variance explained by the first couple of principal components, the first nine components are already able to explain at least 80% of the total variance in the data. The best-performing principal component has been used to train an RF model again, resulting in 100% accuracy. When briefly looking at the performance without dimensionality reduction when using only one RF feature to train, 100% accuracy was not achieved. This indicates the PCA can effectively reduce dimensionality.

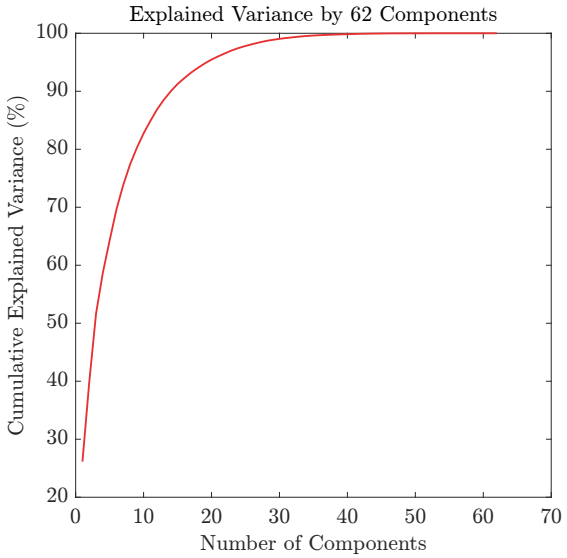


Fig. 8: Visible elbow point in the cumulative explained variance.

The results of the RF and PCA analysis will be presented for the pre-processing and feature extraction methods separately, after selecting a suitable sample rate by looking at the accuracy per sample rate. In a heat map, the results of each combination of the methods are shown for RF predictor importance. After that, having selected the best-performing pre-processing and feature extraction methods and the suitable sample rate, the RF accuracy is again computed versus the window size on which the RF model is trained.

VI. Results

In this section, the results and findings of the study are presented. Since no hits were recorded, no results can be presented on AE analysis. Additional analysis using pre-processing, feature extraction and training an RF model on

continuously recorded AE data was inconclusive. That is why this chapter is aimed at the performance of vibration analysis.

To find out if analysis with a lower sample rate would also be able to detect a defect, the vibration datasets of both sensors individually and combined are down-sampled from 1,280 Hz to 640, 320, 160, 80, 40, 20 and 10 Hz. Low-pass filtering is done before decimating the signal to prevent aliasing. The results can be seen in Figure 9.

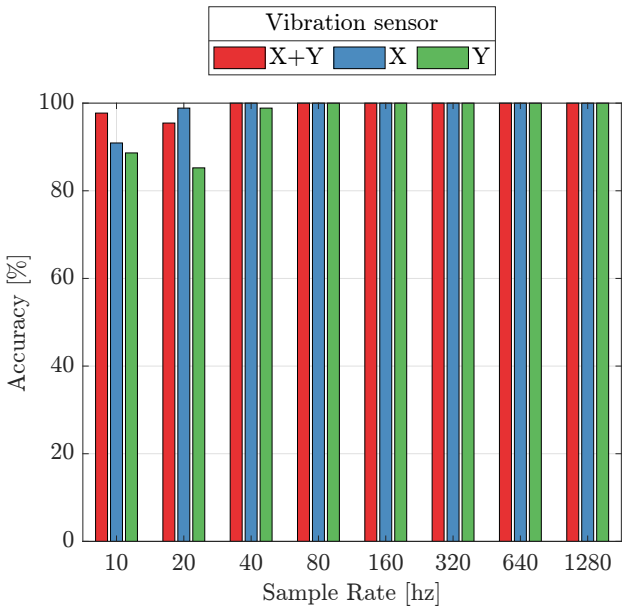


Fig. 9: RF Accuracy per sample rate for vibration X, Y and X+Y axis sensors.

We can see that 80 Hz is the last result where all datasets give 100% RF accuracy. 40 Hz shows the first decreased performance for the Y-axis sensor. 20 Hz shows a decreased performance for all three datasets. The remainder of the analysis of the results is done with a dataset of 40 Hz sample rate using both X and Y sensors to make the results more generalizable.

In Figure 10, the RF predictor importance of the pre-processing methods for all feature extraction methods is shown. The PCA coefficients are plotted to confirm the RF results. The results are normalized to prevent the effects of different dataset sizes on the RF predictor importance and PCA coefficients. The 25% line indicates the average score when all methods are equally as important. Results above this line mean this method has a higher than average importance, and vice versa for below the line. We can see that no pre-processing performs slightly better than the others, but no major differences in performance are measured. Adding noise to the signal after feature extraction methods were selected, gave the same results where not one pre-processing method stood out above the rest

The same is plotted for the feature extraction methods in Figure 11. This time with a horizontal baseline at 12.5%, the outcome of one divided over eight methods. RMS and peak frequency magnitude perform best, followed by the peak value

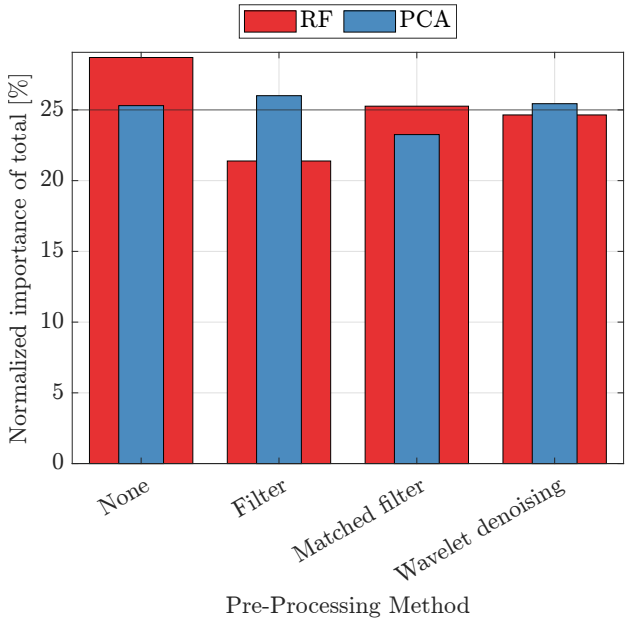


Fig. 10: Normalized RF predictor importance and PCA coefficients of pre-processing methods.

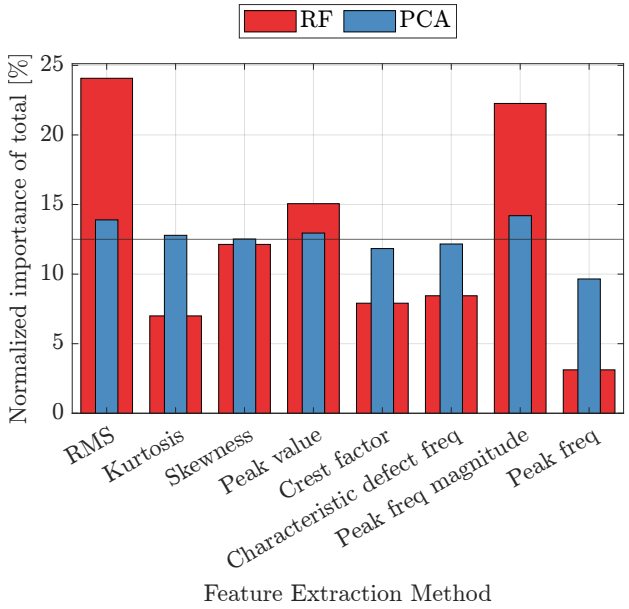


Fig. 11: Normalized RF predictor importance and PCA coefficients of feature extraction methods.

and skewness. Kurtosis, crest factor and characteristic defect frequency score low, with peak frequency scoring worst of all feature extraction methods. PCA again validates the RF predictor importance results, except for kurtosis, where the PCA coefficient indicates that kurtosis might contain more information than RF predictor importance is indicating.

To get a better insight into the performances of pre-processing and feature extraction method combinations, Figure 12 displays a heat map based on the RF predictor importance. Red indicates higher predictor importance, whereas

blue colour indicates lower performance. In the top row and rightmost column, the mean performance of each method is also shown. These mean results are similar to RF predictor importances in Figure 10 and Figure 11. In the top right, the baseline value of 3.125% can be seen. The baseline is the result of 100% divided by the 4 tested pre-processing and 8 feature extraction methods. Summing all method importances results in 100%. The best-performing combination

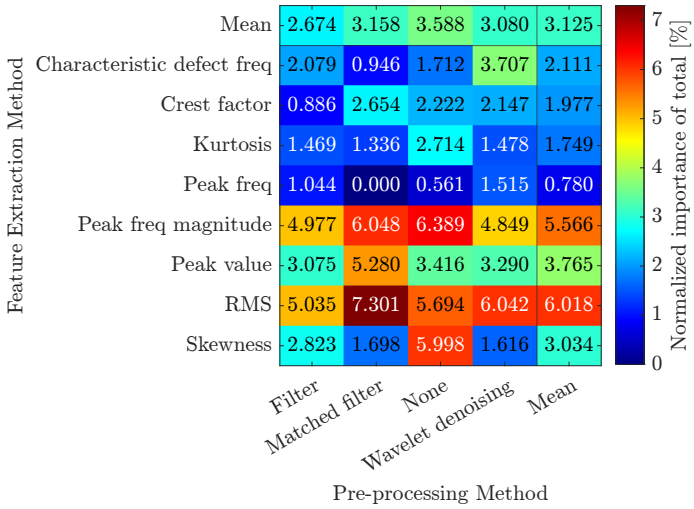


Fig. 12: Heat map of RF predictor importance to compare individual combinations of pre-processing and feature extraction methods.

is a matched filter before the RMS feature. Interestingly, the performance of no pre-processing before the skewness feature performs significantly better than the other pre-processing methods for the same feature. Also noteworthy is how the peak value performs best with matched filter pre-processing, but worse with other pre-processing methods. The same goes for applying wavelet denoising before extracting the characteristic defect frequency amplitude. Applying the matched filter before the peak frequency feature gives the worst performance.

By selecting the two best-performing feature extraction methods from the results, RMS and peak frequency magnitude, the RF accuracy was computed for different sample rates and window sizes on which the RF model is trained. No pre-processing was applied, as none stood out from the results. The outcome of the computations is shown in Figure 13, where yellow is a higher accuracy and blue/purple means lower accuracy. It can be seen that 100% accuracy is reached for 80 Hz and a smaller window size of 1280 samples. 100% Accuracy is also reached for a sample rate of 40 Hz at a window size of 5160 samples, which is equal to one rotation of the shaft. This is similar to the results shown in Figure 9. The mesh plot also shows that there is an increase in accuracy for 10 and 20 Hz when the window size increases.

VII. Discussion

This section interprets the results and addresses challenges in FDD of LSBBs. Sub-questions 3, 4, and 5 are addressed

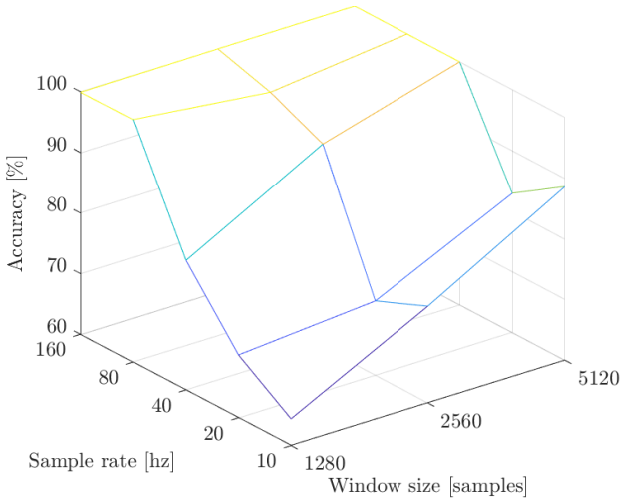


Fig. 13: Mesh plot of RF accuracy with sample rate versus window size.

by analysing the results and drawing conclusions based on the findings. The effectiveness of different sensing techniques for detecting defects in low-speed ball bearings are discussed (SQ3), as well as the most effective pre-processing methods for improving data quality (SQ4), and the most effective feature extraction methods for detecting and identifying defects in LSBBs (SQ5).

A. Sensing Technique

Opposed to AE data, vibration data did contain information on defects as was shown in the results and is thus the most effective. The high RF defect classification accuracy implies that there were differences between the defects that can be captured by accelerometers and are distinguishable.

In the context of current global trends of edge computing to enhance efficiency and lower cost, it was also shown that one sensor could capture sufficient information, instead of the more commonly used three sensors in existing work. Besides that, a low sample rate of 40 Hz proved to be the lowest sample rate with high RF accuracy, again indicating that vital defect information is captured in the lower frequency spectrum.

After analysis of the pre-processing and feature extraction methods, the best-performing methods were selected and the results of comparing sample rate and window size are interpreted. The lower the sample rate, the larger the window size needs to be and vice versa. This can be explained by that a lower sample rate acts as a low-pass filter. To extract low-frequency data at higher accuracy, longer measurements are required, hence the larger window size. For sample rates below 40 Hz, an increase in window size also increases the RF accuracy, meaning that bearing defect information is stored in lower-frequency components.

A limitation of the collected data is its resolution at lower frequencies being low due to the sample rate and dataset size. This poses the necessity for further detailed research to bearing defect-related information stored in the lower frequencies. In-

creasing the accuracy in the lower frequencies can be achieved with longer recordings at a low sample rate.

With a low amount of sensors, low sample rate and limited required window size the amount of data that needs to be collected and processed is reduced, also reducing the cost of hardware. Remote radar sites have a limited data bandwidth which is mainly used for radar footage so using less data is also advantageous in this regard. This opens up research possibilities to use cheaper off-the-shelf hardware to accomplish bearing FDD. Simplicity of the system while also reducing cost, increases scalability options for multiple installations. With data collected from more systems, RUL prediction capabilities can also be improved. One disadvantage of vibration analysis is that a sensor must physically be connected to the structure.

A limitation of this study is that test bench data may not fully represent a real radar system and corresponding bearing defects because of differences in conditions and operating environment. For the initial comparison of FDD approaches, valuable insights are provided by the test bench in a controlled environment.

B. Pre-Processing Methods

In contradiction to existing research, the importance of pre-processing is not shown in this work as no pre-processing method stands out in performance. Applying no pre-processing performed as well as the other pre-processing methods, possibly because the raw vibration data already contains sufficient information about a defect or the system contains a low amount of noise. When data is collected from a real radar system with realistic noise, pre-processing might prove useful. This coincides with the aforementioned limitation of using a test bench to collect data.

C. Feature Extraction Methods

From the results of the feature extraction methods, we can see that amplitude-related features perform well. RMS and peak frequency magnitude stand out as the most effective. This means that the energy and dominant frequency components of the vibration signal contain information about the bearing status. The peak frequency magnitude performing well can also be interpreted as the amplitude of the natural frequencies in a system conveying information on the bearing. This is further demonstrated by the amplitude-based peak value feature performing third best.

Less bearing defect-related information can be found in features related to signal distribution because skewness performed below average and kurtosis was second to worst. The feature that performs worst is the peak frequency, meaning that the dominant frequency is not prone to changes when the bearing status changes.

It was expected that the characteristic defect frequency feature should perform better than the results indicate, because it focuses on changes in the amplitude given by the SP defect on the outer raceway. The low performance could be due to it only helping with detecting an SP defect and does not contain information on GR defects. Other reasons for the low performance are due to the sample rate being too low resulting in too wide frequency bands or that the Z-axis vibration sensor was the one recording most information on this defect.

The performances of pre-processing and feature extraction methods are confirmed by PCA. Only the kurtosis feature extraction method performance differs between RF and PCA. Two possible factors contributing to this difference are that PCA assumes high-variance features are more informative, while RF considers predictive power, resulting in variations in feature importance rankings. Combined with that kurtosis's performance in PCA may be due to capturing unique variance-contributing information, but its significance in the RF model may be minimal when similar information is captured by other features.

The selected feature extraction methods were proven to be the appropriate ones because the RF accuracy is still high when analysing the sample rate and window size effects for these features. The high accuracy of the RF means the model does not generalize well to datasets of other situations. Nevertheless, the predictor importance is indicative of factors that contribute to such performance and can be used.

D. Combinations Of Methods

The heat map in Figure 12 shows the importance of comparing the performance of a certain pre-processing method with a certain feature extraction method. One cannot assume that a certain pre-processing method works well for all feature extraction methods. This is emphasized by applying no pre-processing before skewness, matched filter before peak value and wavelet denoising before characteristic defect frequency feature extraction.

The combination of a matched filter pre-processing before extracting the RMS feature was found to be the most effective. When only looking at the two best-performing feature extraction methods, RMS and peak frequency magnitude, the matched filter seems to perform best. Another two amplitude-related features that performed better with the matched filter than other pre-processing methods are the peak value and crest factor. This indicates that the matched filter was effective in filtering out expected noise based on the template signal of a system with a perfect bearing to maximize the SNR, making amplitude-related feature extraction methods perform better. Coincidentally, the matched filter also performs worst in combination with the peak frequency feature. When looking at the data behind this combination, the same frequency was seen which logically gives zero information on different types of defects.

Having examined the performances of sensing techniques and processing methods, and their importance in the context of the research objectives and questions, the insights gained from this analysis form the basis for drawing conclusions and highlighting the implications of the findings.

VIII. Conclusion

The primary objective of this research was to identify the most effective approach for FDD of LSBBs used in radar

systems. The findings derived from this study provide valuable insights that address the main research question.

This study shows that different LSBB defects result in distinguishable differences in the vibrations of a system with similar properties to a radar system, as demonstrated by vibration analysis. It was also shown that it is possible to use one vibration sensor instead of three orthogonal sensors, with a low sample rate of 40 Hz, while still extracting relevant information on bearing defects.

After a comprehensive assessment, it was observed that no pre-processing was necessary and that amplitude-related feature extraction methods, such as RMS and peak frequency magnitude, generally performed best in detecting defects. Particularly, the combination of applying a matched filter before extracting the RMS feature is the most effective in the context of this research. Although other pre-processing and feature extraction methods also performed well, one should keep in mind that not every combination of methods achieves the same level of performance.

The insights gained from this research contribute to the broader field of radar system maintenance, as they enable radar system maintainers to effectively monitor the health of bearings and implement condition-based maintenance strategies. Condition-based maintenance helps ensure the reliable operation of radar systems, ultimately contributing to a safer, more secure and more efficient maritime environment.

IX. Recommendations

Being the final section of this thesis, recommendations are made for future research possibilities based on the findings and knowledge gained throughout this research.

Collecting data from radar systems will allow future research to use more realistic data than that of a test bench, which is why that is the main recommendation of this study. Maintenance events need to be recorded if the system changes because a change in data can then be explained by certain maintenance events. Once enough data is collected on various radar systems with various bearing statuses and defects, condition-based maintenance can be implemented and steps towards predictive maintenance can be made.

With that in mind, future research recommendations towards predictive maintenance that advance and extend the current research findings are made:

- One recommendation is to study the effectiveness of pre-processing techniques in real-world settings. Raw data seemed to contain enough information to detect defects. However, data from a real in-use system instead of a test bench might contain more noise, requiring pre-processing. For example, because a matched filter is based on a template at one specific speed, speed fluctuations might affect its pre-processing performance. To compensate for rotational speed fluctuations from wind gusts affecting feature domain features, TSA may be an effective pre-processing method.
- This also poses the necessity to further research in which information is stored in the lower frequencies. Increasing the accuracy in the lower frequencies can be achieved with longer recordings at a low sample rate to also prevent the hardware buffer from overflowing. Such research could be accommodated with research towards feature extraction methods not handled throughout this study.
- ML networks for health indicator construction and RUL prediction are getting more popular, a study to which networks can best be applied for the context of the research would be useful. LSTM appears to be a particularly promising ML network because of its ability to capture long-term dependencies in time-series data.

Other, more exploration-oriented, future research recommendations are also made:

- If the PID control influence and environmental noise can be filtered out, stator current might prove as a useful sensing technique since the direct drive does not contain a gearbox. This gives a near-linear relation between stator current and rotational resistance, making it more suitable for detecting bearing defects. Stator current has the advantage that its data can be collected from all functioning radar systems with changes to the software, meaning no hardware needs to be purchased and installed.
- The last recommendation is to explore the potential of utilizing inexpensive off-the-shelf hardware as opposed to expensive proprietary hardware. Because this research shows that a low sample rate and low amount of vibration sensors work for FDD, it could be possible to use cheaper and more available hardware. This can offer opportunities for cost reduction, wider accessibility scalability and adaptability to pave the way for broader implementation. With these factors in mind, this future work should focus on edge computing and non-contact sensing.

Acknowledgment

This work was supported by Saab Technologies B.V. I would like to express my gratitude for the financial assistance provided, which contributed to the success of this research.

I am very grateful to my weekly supervisor, ir. Maarten Tietge. Our discussions have been essential in shaping my ideas and maintaining my motivation.

My appreciation goes to my university supervisor, prof.dr. Paul Havinga, later joined by dr.ir. Berend-Jan van der Zwaag, for their guidance and supervision throughout this project.

Special thanks to Bas Withagen, for his cooperation in making the test bench work for collecting data.

My thanks are extended to the colleagues at Saab Technologies B.V. for sharing their perspectives, often during lunch walks.

I would also like to thank the students I have had the pleasure of knowing during my time at the university, for their friendship and shared learning experiences.

Finally, I am grateful to my partner, parents and friends for their unwavering support and encouragement throughout this endeavour.

REFERENCES

[1] "Maritime Traffic Management — Security — Saab."

[2] "6022-2RS1 - Deep groove ball bearings — SKF."

[3] M. Xia, T. Li, T. Shu, J. Wan, C. W. De Silva, and Z. Wang, "A Two-Stage Approach for the Remaining Useful Life Prediction of Bearings Using Deep Neural Networks," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 3703–3711, 6 2019.

[4] J. Ben Ali, B. Chebel-Morello, L. Saidi, S. Malinowski, and F. Fnaiech, "Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network," *Mechanical Systems and Signal Processing*, vol. 56, pp. 150–172, 5 2015.

[5] M. B. Koura, A. H. Boudinar, A. F. Aimer, and M. E. A. Khodja, "Induction Motor Bearing Faults Diagnosis Using Stator Current and Vibration Analysis," *Periodica polytechnica Electrical engineering and computer science*, vol. 65, pp. 344–351, 10 2021.

[6] "Bearing damage and failure analysis," tech. rep., SKF, 2017.

[7] A. Heng, S. Zhang, A. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," 4 2009.

[8] D. Faraj, *Using Machine Learning for Predictive Maintenance in Modern Ground-Based Radar Systems*. PhD thesis, 2021.

[9] J. A. Sainz and M. A. Sebastián, "Methodology for the maintenance centered on the reliability on facilities of low accessibility," in *Procedia Engineering*, vol. 63, pp. 852–860, Elsevier Ltd, 2013.

[10] "Maximizing the P-F Interval Through Condition-Based Maintenance - Applications - Maintworld."

[11] A. Rai and S. H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribology International*, vol. 96, pp. 289–306, 4 2016.

[12] W. Caesarendra, P. B. Kosasih, A. K. Tieu, C. A. S. Moodie, and B. K. Choi, "Condition monitoring of naturally damaged slow speed slewing bearing based on ensemble empirical mode decomposition," *Journal of Mechanical Science and Technology*, vol. 27, pp. 2253–2262, 8 2013.

[13] Kim Yong-Hanand Tan, Andy C C, Mathew Josephand, and Yang Bo-Suk, "Condition Monitoring of Low Speed Bearings: A Comparative Study of the Ultrasound Technique Versus Vibration Measurements," in *Engineering Asset Management* (Mathew Josephand Kennedy and Jimand Ma Linand TanAndyand Anderson Deryk, eds.), (London), pp. 182–191, Springer London, 2006.

[14] P. D. McFadden and M. M. Toozhy, "Application of synchronous averaging to vibration monitoring of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 14, no. 6, pp. 891–906, 2000.

[15] B. Van Hecke, Y. Qu, and D. He, "Bearing fault diagnosis based on a new acoustic emission sensor technique," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, vol. 229, pp. 105–118, 4 2014.

[16] J. D. J. D. McIntyre, R. B. K. N. Rao, D. G. Sleeman, and University of Sunderland. Centre for Adaptive Systems., *Acoustic Emission in Monitoring Extremely Slow Rotating Rolling Bearing*. Coxmoor Publishing, 1999.

[17] W. Zhou, T. G. Habetler, and R. G. Harley, "Bearing Condition Monitoring Methods for Electric Machines: A General Review," in *2007 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, pp. 3–6, 2007.

[18] L. Renaudin, F. Bonnardot, O. Musy, J. B. Doray, and D. Rémond, "Natural roller bearing fault detection by angular measurement of true instantaneous angular speed," in *Mechanical Systems and Signal Processing*, vol. 24, pp. 1998–2011, Academic Press, 2010.

[19] G. Georgoulas and G. Nikolakopoulos, "Bearing fault detection and diagnosis by fusing vibration data," *IECON Proceedings (Industrial Electronics Conference)*, pp. 6955–6960, 12 2016.

[20] Y. Peng, W. Qiao, and L. Qu, "Compressive Sensing-Based Missing-Data-Tolerant Fault Detection for Remote Condition Monitoring of Wind Turbines," *IEEE Transactions on Industrial Electronics*, 2021.

[21] D. Aníbal and S. Núñez, *Diagnosis of Low-Speed Bearings via Vibration-Based Entropy Indicators and Acoustic Emissions*. PhD thesis.

[22] M. He and D. He, "Deep Learning Based Approach for Bearing Fault Diagnosis," *IEEE Transactions on Industry Applications*, vol. 53, pp. 3057–3065, 5 2017.

[23] N. W. Nirwan and H. B. Ramani, "Condition monitoring and fault detection in roller bearing used in rolling mill by acoustic emission and vibration analysis," in *Materials Today: Proceedings*, vol. 51, pp. 344–354, Elsevier Ltd, 2021.

[24] M. Elforjani and S. Shanbr, "Prognosis of Bearing Acoustic Emission Signals Using Supervised Machine Learning," *IEEE Transactions on Industrial Electronics*, vol. 65, pp. 5864–5871, 7 2018.

[25] M. Zhao and J. Lin, "Health assessment of rotating machinery using a rotary encoder," *IEEE Transactions on Industrial Electronics*, vol. 65, pp. 2548–2556, 3 2018.

[26] L. F. Villa, A. Reñones, J. R. Perán, and L. J. De Miguel, "Angular resampling for vibration analysis in wind turbines under non-linear speed fluctuation," 8 2011.

[27] D. Wu, S. Liu, L. Zhang, J. Terpenny, R. X. Gao, T. Kurfess, and J. A. Guzzo, "A fog computing-based framework for process monitoring and prognosis in cyber-manufacturing," *Journal of Manufacturing Systems*, vol. 43, pp. 25–34, 4 2017.

[28] S. Tyagi and S. K. Panigrahi, "An improved envelope detection method using particle swarm optimisation for rolling element bearing fault diagnosis," *Journal of Computational Design and Engineering*, vol. 4, pp. 305–317, 10 2017.

[29] D. Juodelyte, V. Cheplygina, T. Graversen, and P. Bonnet, "Predicting Bearings Degradation Stages for Predictive Maintenance in the Pharmaceutical Industry," pp. 3107–3115, Association for Computing Machinery (ACM), 8 2022.

[30] L. Liu, X. Song, K. Chen, B. Hou, X. Chai, and H. Ning, "An enhanced encoder–decoder framework for bearing remaining useful life prediction," *Measurement: Journal of the International Measurement Confederation*, vol. 170, 1 2021.

[31] X. Li, Q. Ding, and J. Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering and System Safety*, vol. 172, pp. 1–11, 4 2018.

[32] R. Liu, B. Yang, and A. G. Hauptmann, "Simultaneous bearing fault recognition and remaining useful life prediction using joint-loss convolutional neural network," *IEEE Transactions on Industrial Informatics*, vol. 16, pp. 87–96, 1 2020.

[33] M. Yan, L. Xie, I. Muhammad, X. Yang, and Y. Liu, "An effective method for remaining useful life estimation of bearings with elbow point detection and adaptive regression models," *ISA Transactions*, vol. 128, pp. 290–300, 9 2022.

[34] C. Sobie, C. Freitas, and M. Nicolai, "Simulation-driven machine learning: Bearing fault classification," *Mechanical Systems and Signal Processing*, vol. 99, pp. 403–419, 1 2018.

[35] J. Zarei, M. A. Tajeddini, and H. R. Karimi, "Vibration analysis for bearing fault detection and classification using an intelligent filter," *Mechatronics*, vol. 24, no. 2, pp. 151–157, 2014.

[36] V. R. K. Ramachandran, A. S. Ramirez, B. J. Van Der Zwaag, N. Meratnia, and P. Havinga, "On-node Signal Processing to Reduce the Power Consumption of Wireless Sensor Nodes for Vibration Monitoring," in *IEEE ISSNIP 2014 - 2014 IEEE 9th International Conference on Intelligent Sensors, Sensor Networks and Information Processing, Conference Proceedings*, IEEE Computer Society, 2014.

[37] D. Chen, Y. Qin, Y. Wang, and J. Zhou, "Health indicator construction by quadratic function-based deep convolutional auto-encoder and its application into bearing RUL prediction," *ISA Transactions*, vol. 114, pp. 44–56, 8 2021.

[38] E. Bechhoefer and M. Kingsley, "A Review of Time Synchronous Average Algorithms," *Annual Conference of the PHM Society*, vol. 1, no. 1, 2009.

[39] C. Mishra, A. K. Samantaray, and G. Chakraborty, "Rolling element bearing defect diagnosis under variable speed operation through angle synchronous averaging of wavelet de-noised estimate," *Mechanical Systems and Signal Processing*, vol. 72-73, pp. 206–222, 5 2016.

[40] K. Kaplan, Y. Kaya, M. Kuncan, M. R. Micnaz, and H. M. Ertunç, "An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis," *Applied Soft Computing Journal*, vol. 87, 2 2020.

[41] G. Betta, C. Liguori, A. Paolillo, and A. Pietrosanto, "A DSP-based FFT-analyzer for the fault diagnosis of rotating machine based on vibration analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, pp. 1316–1321, 12 2002.

[42] L. Ren, Y. Sun, J. Cui, and L. Zhang, "Bearing remaining useful life prediction based on deep autoencoder and deep neural networks," *Journal of Manufacturing Systems*, vol. 48, pp. 71–77, 7 2018.

[43] D. Bruneo and F. De Vita, "On the use of LSTM networks for predictive maintenance in smart industries," in *Proceedings - 2019 IEEE International Conference on Smart Computing, SMARTCOMP 2019*,

pp. 241–248, Institute of Electrical and Electronics Engineers Inc., 6 2019.

[44] T. Gao, Y. Li, X. Huang, and C. Wang, "Data-driven method for predicting remaining useful life of bearing based on bayesian theory," *Sensors (Switzerland)*, vol. 21, pp. 1–17, 1 2021.

[45] T. Zonta, C. A. da Costa, R. da Rosa Righi, M. J. de Lima, E. S. da Trindade, and G. P. Li, "Predictive maintenance in the Industry 4.0: A systematic literature review," *Computers and Industrial Engineering*, vol. 150, 12 2020.

[46] "CMSS 2200 - Sensors — SKF."

[47] "Multilog on-line systems IMx 8 and IMx-16Plus — SKF."

[48] "VS150-RSC - Vallen Systeme."

[49] "WaveLine Products - Vallen Systeme."