Condition Monitoring Methods for Large, Low-speed Bearings

Martin Hemmer

UiA University of Agder

Doctoral Dissertations at the University of Agder 260

Martin Hemmer

Condition Monitoring Methods for Large, Low-speed Bearings

Doctoral Dissertation for the Degree *Philosophiae Doctor (PhD)* at
the Faculty of Engineering and Science, Specialisation in Mechatronics

University of Agder
Faculty of Engineering and Science
2020

Doctoral Dissertation at the University of Agder 260
ISSN: 1504-9272
ISBN: 978-82-7117-960-1

Printed by Media 07
Oslo

Acknowledgments

The project was made possible through the SFI Offshore Mechatronics project, funded by industrial partners, the University of Agder and the Norwegian Research Council. I would like to acknowledge the work of my main supervisor Prof. Kjell G. Robbersmyr, co-supervisors Prof. Huynh Van Khang and Dr. Tor I. Waag, and SFI work package leader Dr. Thomas J. J. Meyer. I am grateful for your guidance, encouragement and support. Among the project partners I owe a special thanks to MHWirth AS for allowing me to pursue this degree as an employee in a difficult time for the industry, and for providing real offshore equipment and test facilities for my experiments. I also greatly appreciate the help from MacGregor Norway AS, Bosch Rexroth and AG Mekanikk during development, production and assembly of the axial bearing test rig that I built as part of the project.

My colleagues at the University of Agder and in the SFI project have been a reliable source of knowledge, inspiration, encouragement and amusement. I appreciate you for making these years much more enjoyable. A particular thanks to Dr. Andreas Klausen for outstanding technical advice, collaboration and discussions. My deepest gratitude goes to my friends, family and especially my girlfriend Cecilie for the unconditional love, patience and support you have shown to help me through the times of doubt, stress and despair. I could never have done it without you.

Summary

In all industrial production plants, well-functioning machines and systems are required for sustained and safe operation. However, asset performance degrades over time and may lead to reduced efficiency, poor product quality, secondary damage to other assets or even complete failure and unplanned downtime of critical systems. Besides the potential safety hazards from machine failure, the economic consequences are large, particularly in offshore applications where repairs are difficult. This thesis focuses on large, low-speed rolling element bearings, concretized by the main swivel bearing of an offshore drilling machine. Surveys have shown that bearing failure in drilling machines is a major cause of rig downtime. Bearings have a finite lifetime, which can be estimated using formulas supplied by the bearing manufacturer. Premature failure may still occur as a result of irregularities in operating conditions and use, lubrication, mounting, contamination, or external environmental factors. On the contrary, a bearing may also exceed the expected lifetime. Compared to smaller bearings, historical failure data from large, low-speed machinery is rare. Due to the high cost of maintenance and repairs, the preferred maintenance arrangement is often condition based. Vibration measurements with accelerometers is the most common data acquisition technique. However, vibration based condition monitoring of large, low-speed bearings is challenging, due to non-stationary operating conditions, low kinetic energy and increased distance from fault to transducer. On the sensor side, this project has also investigated the usage of acoustic emission sensors for condition monitoring purposes.

Roller end damage is identified as a failure mode of interest in tapered axial bearings. Early stage abrasive wear has been observed on bearings in drilling machines. The failure mode is currently only detectable upon visual inspection and potentially through wear debris in the bearing lubricant. In this thesis, multiple machine learning algorithms are developed and applied to handle the challenges of fault detection in large, low-speed bearings with little or no historical data and unknown fault signatures. The feasibility of transfer learning is demonstrated, as an approach to speed up implementation of automated fault detection systems when historical failure data is available. Variational autoencoders are proposed as a method for unsupervised dimensionality reduction and

feature extraction, being useful for obtaining a health indicator with a statistical anomaly detection threshold. Data is collected from numerous experiments throughout the project. Most notably, a test was performed on a real offshore drilling machine with roller end wear in the bearing. To replicate this failure mode and aid development of condition monitoring methods, an axial bearing test rig has been designed and built as a part of the project. An overview of all experiments, methods and results are given in the thesis, with details covered in the appended papers.

Publications

The following listed articles have been published or submitted for publication in peer reviewed conference proceedings and journals. The version presented in this thesis differs only in formatting and minor errata.

Paper A Martin Hemmer, Tor I. Waag and Kjell G. Robbersmyr. A Review of Methods for Condition Monitoring of Large, Slow-rotating Bearings. Presented at: *30th Conference on Condition Monitoring and Diagnostic Engineering Management.* University of Central Lancashire, Preston, UK

Paper B Martin Hemmer and Tor I. Waag. A Comparison of Acoustic Emission and Vibration Measurements for Condition Monitoring of an Offshore Drilling Machine. *Proceedings of the Annual Conference of the Prognostics and Health Management Society 2017*, 278-285, 2017. ISBN 978-1-936263-26-4.

Paper C Martin Hemmer, Kjell G. Robbersmyr, Tor I. Waag, Rolf Albrigtsen, Torfinn Pedersen, Thomas J. J. Meyer and Chloë Vercammen. Rib-Roller Wear in Tapered Rolling Element Bearings: Analysis and Development of Test Rig for Condition Monitoring. *Engineering Assets and Public Infrastructures in the Age of Digitalization.* Springer, in press. Presented at: *13th World Congress on Engineering Asset Management.* Stavanger, Norway

Paper D Martin Hemmer, Huynh V. Khang, Kjell G. Robbersmyr, Tor I . Waag and Thomas J. J. Meyer. Fault Classification of Axial and Radial Roller Bearings Using Transfer Learning through a Pretrained Convolutional Neural Network. *Designs*, 2(4), 56, 2018. doi: 10.3390/designs2040056.

Paper E Martin Hemmer, Andreas Klausen, Huynh V. Khang, Kjell G. Robbersmyr and Tor I . Waag. Health Indicator for Low-speed Axial Bearings using Variational Autoencoders. Under review at: *IEEE Access.*

Paper F Martin Hemmer, Andreas Klausen, Huynh van Khang, Kjell G. Robbersmyr, and Tor I. Waag. Simulation-driven Deep Classification of Bearing Faults from Raw

Vibration Data. Accepted for publication in: *International Journal of Prognostics and Health Management (IJPHM)*. Accepted date: 15.08.2019.

The following article was written and published during the time of this project, but is not included in the thesis.

Paper G T. I. Waag, M. Hemmer. Observation and processing of instantaneous frequency variations during bearing tests. *Engineering Assets and Public Infrastructures in the Age of Digitalization.* Springer, in press. Presented at: *13th World Congress on Engineering Asset Management*, Stavanger, Norway

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

AE	Acoustic emission
AR	Autoregressive
BOP	Blow out preventer
BPFI	Ball pass frequency inner race
BPFO	Ball pass frequency outer race
BSF	Ball spin frequency
CBM	Condition based maintenance
CI	Condition indicator
CM	Condition monitoring
CNN	Convolutional neural network
CVAE	Conditional variational autoencoder
CWT	Continuous wavelet transform
DBN	Deep belief network
DL	Damage level
DT	Decision tree
ELBO	Evidence lower bound
fc	fully-connected
FFT	Fast Fourier transform
FTF	Fundamental train frequency
FWER	Family-wise error rate
GAN	Generative adversarial network
HE	Healthy
HI	Health indicator
hp	Horsepower
IR	Inner race
KL	Kullback-Leibler
k-NN	k-nearest neighbor
LSTM	Long short-term memory
ML	Machine learning
O&G	Oil and gas
OR	Outer race
PACF	Partial autocorrelation function
PdM	Predictive maintenance
P_D	Probability of detection
P_{FA}	Probability of false alarm
PHM	Prognostics and health management
REB	Rolling element bearing
RE	Rolling element

RMS	Root mean square
RNN	Recurrent neural network
rpm	Revolutions per minute
SAE	Sparse autoencoder
SANC	Self-adapting noise cancellation
SC	Scratched
SK	Spectral kurtosis
SNR	Signal-to-noise ratio
SOM	Self-organising map
SVM	Support vector machine
TL	Transfer learning
TSA	Time synchronous averaging
VAE	Variational autoencoder

Variables

a	Scratch model, roller radius
A	Scratch model, roller center
α	Significance
b	Scratch model, bearing radius
B	Scratch model, bearing center
C	Dynamic load rating
D	Bearing pitch diameter
d	Roller diameter
D_{KL}	Kullback-Leibler divergence
$\boldsymbol{\epsilon}$	VAE sampled noise vector
$F(\cdot)$	Cumulative distribution function
$F^{-1}(\cdot)$	Inverse cumulative distribution function
F_{ax}	Axial force
F_{norm}	Normal force
f_r	Shaft frequency
F_s	Sample frequency
F_{seat}	Seating force
$\boldsymbol{h}$	Exponential decay function
J	Latent space dimension
k	Bearing life equation exponent
L_{10}	Bearing basic rating life
$\mathcal{L}_{KL}$	KL loss
$\mathcal{L}_R$	Reconstruction loss
$\mathcal{L}_{VAE}$	VAE loss
M	Minibatch size
$\boldsymbol{m}_1$	IR or RE modulating function
$\boldsymbol{m}_2$	FTF modulation function
$\boldsymbol{\mu}$	VAE latent varaible mean values
N	VAE features per datapoint
n_r	Number of rollers
$\boldsymbol{p}$	Impulse train
P	Equivalent dynamic load
p	AR model order
$p_d(\mathbf{x})$	VAE distribution of $\mathbf{x}$
ϕ	VAE encoder parameters
φ	Bearing contact angle

P	Scratch model, particle position
ψ	Scratch model, angle
P_x	Particle position, x coordinate
P_y	Particle position, y coordinate
$p(\mathbf{z})$	Target latent variable distribution
q	Number of hypothesis tests
$q(\mathbf{z})$	Aggregated latent variable distribution
$\boldsymbol{r}$	Position vector
γ	Scratch model relation: $(a+b)/b$
$\boldsymbol{\sigma}^2$	VAE latent varaible variances
t_h	Duration of $\boldsymbol{h}$
θ	VAE decoder parameters
t_r	Record duration
t_s	Stride time
t_w	Duration of rolling window
$\boldsymbol{w}$	White noise vector
$\boldsymbol{w}_{bp}$	Bandpass-filtered white noise
$\hat{\mathbf{x}}$	VAE data vector recontruction
$\mathbf{x}$	VAE data vector
$\boldsymbol{x}_F$	Simulated faulty data
$\boldsymbol{x}_{HE}$	Healty data
$\boldsymbol{x}_S$	Simulated fault data
$\boldsymbol{x}'_S$	Simulated fault data, not normalized
Y	Random variable, function of observed of CIs
Y_0	Subset of Y, healthy data only.
$\mathbf{z}$	VAE latent vector

Chapter 1

Introduction

1.1 Background

The bearing is a fundamental component in all types of rotating machinery. Bearings constrain degrees of freedom, transfer load and reduce friction of moving parts in a system. Beyond this, the variety of applications has sparked the development of a great number of bearing types and sizes. Grouped by principle of operation, plain bearings, rolling element bearings (REBs), fluid bearings and magnetic bearings are some of the most common types. This thesis considers rotary REBs, which use rolling elements (REs) to reduce the friction between a shaft and a housing in relative rotation. A generic bearing consists of four main parts: Outer race (OR), cage, inner race (IR) and REs. A ball bearing with an OR fault is shown in Fig. 1.1, together with the RE diameter d, pitch diameter D and RE contact angle φ.

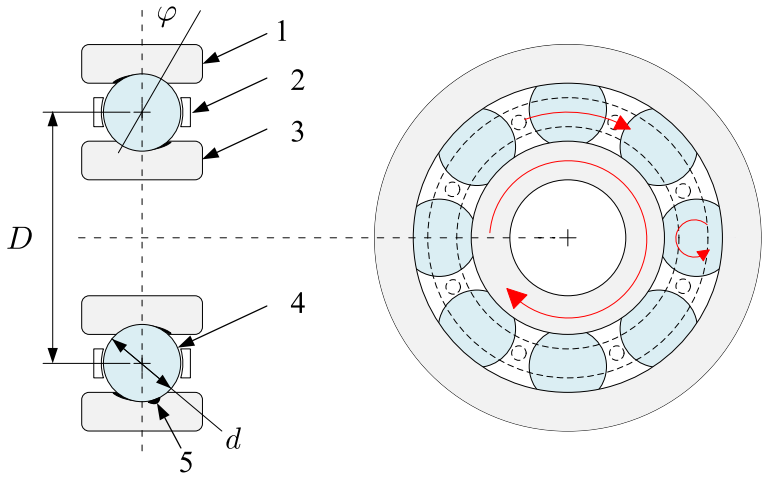


Figure 1.1: Diagram of ball bearing with rotating IR and a single-point OR fault. **1)** OR, **2)** Cage, **3)** IR, **4)** RE, **5)** OR fault

Load is transferred from one race to another through the REs. The REs are kept evenly separated by a cage, which is not a load-bearing component. A lubricant is used to form an elastohydrodynamic fluid film between the moving parts, which further reduces friction and wear by preventing metal-to-metal contact. REBs are broadly characterized by the RE shape, and the direction of load that can be accommodated. Ball bearings use spherical rollers, which have a point-like, elliptical contact area with the bearing races. Roller bearings have a line-like, rectangular contact area. Line contact provides a better load distribution and load capacity as compared to ball bearings. However, this increases friction and reduces the ability to handle high rotational speeds. Loads acting on the bearing can be decomposed to radial and axial directions. Depending on the main direction of intended load, bearings are also classified as radial or axial. In practice, many bearing types can accommodate combined loads. Tapered roller bearings have the same main components, as shown in Fig. 1.2. OR and IR may also be named cup and cone. Due to the tapered roller shape, a retaining rib on the OR is required to keep the REs in place. In the rib-roller contact area, shown in red, mating surfaces are sliding rather than rolling. The bearing shown is of type ISB 29230, and has been used in the experimental work of this project to investigate condition monitoring (CM) methods for roller end wear detection.

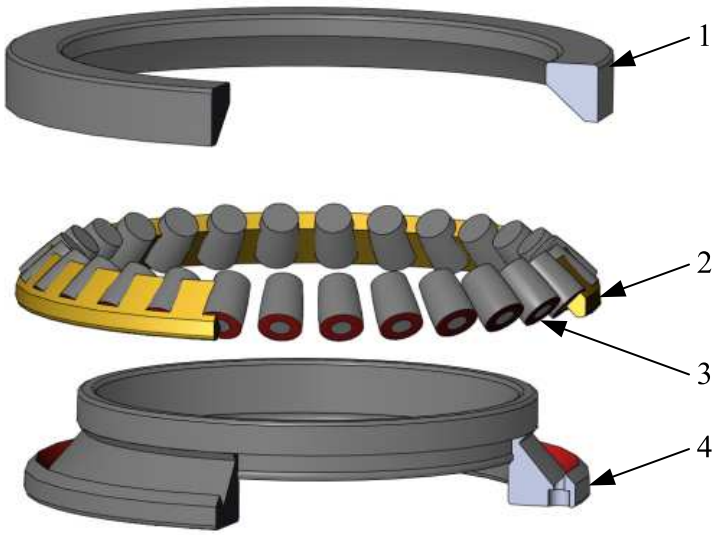


Figure 1.2: Cross-sectional diagram of axial roller bearing type ISB 29230. Rib-roller contact area in red. **1)** Outer race (cup), **2)** Cage, **3)** Roller, **4)** Inner race (cone)

Even with static external loads, the internal material stress is cyclic due to load transferring through the moving REs. Under otherwise nominal operating conditions, the cyclic loading will induce fatigue cracks in the material and eventually cause particles to break off from the contact surfaces. This phenomenon is called spalling. The basic rating life,

L_{10}, is defined in ISO 281:2007 [1] and is calculated based on basic dynamic load rating C, equivalent dynamic bearing load P and an exponent $k = 3$ for ball bearing and $k = \frac{10}{3}$ for roller bearings as given in Eq. (1.1).

$$L_{10} = \left(\frac{C}{P}\right)^k \tag{1.1}$$

This means that in a larger population of identical bearings, 10 % are expected to fail before L_{10} life. The remaining 90 % are expected to exceed the basic rating life. For any individual bearing, the actual lifetime may be both longer and shorter than the rating life. Manufacturers also provide models to correct factors like variable speed and load, contamination, lubricant condition and required reliability. In reality, the parameters for lifetime calculation are hard to determine exactly, adding more uncertainty to the estimate. Beyond fatigue, failure can have a wide range of other root causes. Table 1.1 lists major failure modes with sub-types, as defined in ISO 15243 [2]. In addition to failure modes, the standard comprises fault characteristics, terminology and failure root causes. Other failure modes also result in localized surface damage. This thesis refers to such damage as single-point faults, as they exhibit the same periodic fault signature.

Table 1.1: Common bearing failure modes

Failure mode	**Sub-type**
Fatigue	Surface initiated fatigue
	Sub-surface initiated fatigue
Wear	Abrasive wear
	Adhesive wear
Corrosion	Moisture corrosion
	Frictional corrosion
Electrical erosion	Excessive voltage
	Current leakage
Plastic deformation	Overload deformation
	Debris indentation
	Handling indentation
Fracture and cracking	Forced fracture
	Fatigue fracture
	Thermal cracking

As the bearings in rotating machinery will degrade over time, some maintenance is

eventually required to fulfil performance requirements. Typical maintenance actions include relubrication, shaft alignment, modification of mounting, fit or clearance, and bearing replacement. However, maintenance comes at a cost. Machinery and production must be stopped, parts must be repaired or replaced, and skilled manpower must be employed. This cost, combined with the consequence of failure and available knowledge of machine condition, determines when maintenance is performed. A corrective maintenance strategy does not attempt to avoid failure, and maintenance is carried out when the system no longer functions as required. This approach is applicable when the consequence of failure is acceptable compared to maintenance cost. A preventive maintenance strategy aims to avoid failure by performing maintenance actions after a given time (calendar based), number of operating hours or other experience-based measures. No strategy considers machine health, and maintenance is performed either too early or not early enough. A more advanced approach is to estimate the machine condition, and perform maintenance when required. Condition based maintenance (CBM), predictive maintenance (PdM), and prognostics and health management (PHM) are all maintenance regimes that include CM to reduce unplanned downtime and extend component lifetime. Ideally, a damage is detected at an early stage, allowing time for planning and execution of maintenance. Therefore, CM methods to determine bearing health are of value for the industry. The work presented in this thesis focuses on single-point surface damages from fatigue and abrasive wear in axial bearings.

Damage on internal surfaces will interact with other bearing parts during operation. Given enough energy through load and speed, the interaction will excite transient structural resonance in the support structure. In single-point faults, the damage is localized at a small area, causing the transients to occur periodically and the interaction becomes impact-like. The impact rate is a function of shaft frequency f_r, bearing geometry and fault location. Identifying the presence of these frequencies in a time series is the key to diagnose a single-point bearing fault [3]. Transients are produced at ball pass frequency outer race (BPFO), ball pass frequency inner race (BPFI) and ball spin frequency (BSF), describing faults in the OR, IR and RE, respectively. The fundamental train frequency (FTF) describes the rotational frequency of the cage, which mainly appears as an amplitude modulation frequency in RE faults. The calculation of characteristic fault frequencies inn ball and cylindrical roller bearings is shown in Eq. (1.2) - (1.5), where n_r is the number of rollers.

$$BPFO = \frac{n_r f_r}{2}\left(1 - \frac{d}{D}\cos\varphi\right) \tag{1.2}$$

$$BPFI = \frac{n_r f_r}{2}\left(1 + \frac{d}{D}\cos\varphi\right) \tag{1.3}$$

$$BSF = \frac{Df_r}{2d}\left(1 - \left(\frac{d}{D}\cos\varphi\right)^2\right) \tag{1.4}$$

$$FTF = \frac{f_r}{2}\left(1 - \frac{d}{D}\cos\varphi\right) \tag{1.5}$$

Single-point damage can occur in all types of REBs, whereas roller end abrasive wear only occurs in roller bearings, an example being the spherical tapered roller bearing shown in Fig. 1.2. This thesis considers abrasive wear, which is characterized by material removal from a surface inflicted by another body with greater hardness. In REBs, this can occur from external particle contamination or existing wear debris combined with insufficient lubrication. Abrasive wear on roller ends has been observed in top drive main bearings, being an unusual and less researched failure mode than fatigue damages.

This thesis considers large, low-speed machinery, defined as having an operating speed below 600 rpm (10 Hz) [4, 5, 6]. However, there is no unified definition of neither low-speed nor large size. Large, low-speed bearings are of particular interest, as both aspects render additional challenges to bearing CM. Signal-to-noise ratio (SNR) describes the ratio of power between the fault-induced signal of interest and background noise. Considering a single-point damage as in Fig. 1.1, the RE impacts the fault as it rolls over. At low speed, there is less kinetic energy in the system. Thus, the impact energy is lower, which leads to a lower amplitude structural resonance. Background noise independent of speed can then camouflage the signal of interest. Large size also means that the oscillation has to propagate a larger distance from the fault to the transducer. As the support structure has some inherent damping, vibrations will be attenuated before reaching the transducer. Characteristic fault frequencies are proportional to shaft rate, and thus both lower and closer in the frequency domain at low speed. Separation of low frequency components requires longer time series data, as spectral resolution is inversely proportional to acquisition time. However, this increases the risk of non-stationary operating conditions during measurements [5]. For example, wind turbines experience speed fluctuations under operations due to wind shear and tower wind shadow [7]. A single-point fault may therefore be less detectable, as frequency content from the same fault is then spread between several bins in the spectrum, making it more difficult to identify peaks.

In tapered roller bearings, the lubricant has to form a fluid film between rib and roller end to avoid wear. Subject to the wear, the internal clearances in a bearing will be altered, which may lead to a catastrophic failure [8]. Low speed and high load result in a thinner oil film, increasing the risk of boundary lubrication conditions [9]. Unlike single-point surface damages, the wear does not necessarily generate periodic impacts. Therefore, CM techniques that are suitable for single-point surface damage may not be applicable to

detect the roller end wear. A better understanding of the failure mode is important for both academia and industry, as discussed in Section 3.1, Papers B and C.

1.2 Motivation for Top Drive Study

Top drives are mainly used for handling and supplying torque to the drill string on oil and gas (O&G) drilling rigs. A top drive generally has a maximum speed below 300 rpm (5 Hz), but the actual drilling speed varies. During drilling, the weight of the drill string is supported by a large axial bearing. These bearings have an outer diameter up to around 750 mm, depending on the capacity of the top drive. Top drives are in the critical path of drilling operations, where a failure results in non-productive time on the rig and loss of income. Together with rotary tables and associated equipment, top drives contribute to 60 % of total maintenance costs of drilling rigs, with a cost of over 165,000 USD per incident [10]. Another survey showed that top drives were responsible for 13 % of rig downtime, only surpassed by subsea blow out preventers (BOPs), as shown in Table 1.2 [11]. The same survey showed that bearing failures were the most common causes of downtime.

Table 1.2: Drilling rig downtime grouped by equipment failure

Equipent	**Downtime**
Subsea BOP	46 %
All other	16 %
Top Drive	13 %
Riser	10 %
Mud System	6 %
Surface BOP	5 %
Thruster	4 %

A combination of reactive and calendar-based preventive maintenance has been the standard in the O&G standard for years. Upon routine visual inspection, the roller end wear has been observed in top drive main bearings. The statistics show that unexpected failure still occurs, and there is also a clear trend in the industry towards a condition based and predictive maintenance regime. Improved methods for CM of top drive bearings can contribute to large economic gains. However, online, non-intrusive methods for roller end wear detection have not been developed or missing in literature. Similar large, slow-rotating REBs are also found in other heavy industrial equipment such as paper and

steel mills, wind turbines and cranes. Thus, any research on roller end wear detection is significant for CBM in industry beyond the O&G sector.

1.3 State of the Art

This section introduces state-of-the-art methods for obtaining, processing and evaluating CM data for large, low-speed bearings with single-point faults or roller end wear.

Detection and diagnosis of either fault type require CM data, which refers to any measurable quantity that contains machine health information, being extracted as a condition indicator (CI). Given the challenges of low SNR and speed fluctuation discussed in Section 1.2, both CM data type and processing methods are crucial to detect and diagnose faults. For bearings, lubricant, temperature, vibration, ultrasound and acoustic measurements are the main categories of CM data [12, 4, 3]. Mechanical vibration is commonly measured using accelerometers, but displacement sensors are also effective for low-speed applications [13, 14]. Acoustic emissions (AE) are high frequency stress waves from rapid releases of energy in a material, such as sub-surface crack growths [15] or wear [16]. It was shown early that AE could detect faults before vibration methods [17, 18], and this technology is particularly efficient at low speed [13, 19, 20, 21, 22, 23, 24]. AE measurements have also been shown to reveal particle contamination in fluid film bearings [25] and wear detection in tribological systems [26, 27]. The main drawback of using AE equipment is the high cost as compared to other solutions.

Lubricant analysis is an alternative to acceleration and AE measurements for wear detection. Offline analysis of periodical oil samples are still the industry standard, but are still reactive or routine maintenance. The on-line oil monitoring is relatively new and requires additional sensors and equipment [28, 29, 30]. Therefore, a study of using both acceleration and AE measurements for top drive wear detection is very useful to determine the best sensor or data options for the fault diagnosis of large, low-speed bearings.

Fault detection can be achieved by CI values, where a change in the CI values indicates a fault. RMS of vibration measurements is widely used as a CI, and velocity RMS limits for different types of industrial rotating machinery are defined in ISO 10816-3 [31]. Peak-to-peak and kurtosis (the 4th statistical moment) are also commonly used. Moreover, further processing techniques may be necessary to perform diagnosis, i.e. determining not only the presence, but also the type of fault. For diagnosis of single-point defects, the envelope analysis has proven effective for decades [32]. Digital demodulation is efficiently done using fast Fourier transform (FFT) in the calculation of the Hilbert envelope [33]. Demodulation of a frequency band instead of the full signal will isolate the fault-induced

vibration and thus improve diagnosis capability by increasing SNR. Spectral kurtosis (SK) is used for identifying a suitable frequency bands for demodulation by identifying bands with high kurtosis [34]. A fault-induced impact creates more extreme values compared to healthy (HE) operation, giving a higher kurtosis value. The fast kurtogram [35], autogram [36] and adaptive kurtogram [37] are proposed as improvements of the kurtogram. The Hilbert-Huang transform [38, 39, 40] and wavelet denoising [41, 42] are also successfully applied to low-speed bearing fault detection, and more useful for non-stationary signals.

In top drives and other complex mechanical systems, measured vibrations are corrupted by noises from other sources, such as gears and motors. This type of vibration is directly linked to shaft rotation whereas REB vibrations from a localized surface damage are essentially random due to slip of the RE and cage [43]. Cyclostationary signal processing methods have also proven very effective for uncovering periodicities in low-speed applications [44, 7, 45]. Vibration from this type of REB damage can be modeled as second order cyclostationary signal, while noise from gears are first order cyclostationary [46, 3, 47]. The distinction between discrete shaft-locked and random vibration component allows isolating the bearing vibration signal using self-adaptive noise cancellation (SANC), time synchronous averaging (TSA) [48, 6], cepstral editing [49, 50] or other methods [51]. The problem of non-stationary operation can be addressed by order tracking, i.e. recording the shaft position and resampling the signal from time domain to angular domain [52, 3, 53]. If an encoder or tachometer is not available, an estimate of shaft rotation can be obtained through analytic signal phase demodulation or ridge detection in time-frequency domain [54]. While the aforementioned methods are able to detect and diagnose many bearing faults that occur in industry, including some large, low-speed applications, there are still aspects to be addressed. The majority of methods rely on the assumption that a fault results in periodic transients that can be detected and used to diagnose a bearing fault. The results of this thesis suggest that this is not the case for roller end wear. Additionally, implementing a CM system that requires a lot of human interaction in the analysis of a large amount of data, increases the financial investment. Instead, automated systems provided by model-based, data-driven or hybrid algorithms [55, 56, 57], reduce the time consumption on analysis and requirements to trained personnel. However, both rule-based decision systems and physical models are very hard to obtain in real, complex systems. Data-driven statistical or machine learning (ML) algorithms are therefore attractive as automatic CM systems [58]. The transition to industry 4.0, with more sensors and available data [59], combined with rapid increases in computer hardware performance, data storage capacity and availability of high-level programming languages, makes ML more accessible and encourages the development of data-driven CM methods.

This project has mainly focused on two areas within ML, namely supervised classification and feature learning for anomaly detection. In supervised classification, the goal is to assign the correct class to an observation based on experience from previous, labeled observations. Support vector machine (SVM) [60, 61, 62, 63, 64], decision tree (DT) [65, 66, 67], k-nearest neighbor (k-NN) [68, 69] have all been shown effective in supervised bearing fault classification. Normally, a set of features are extracted from the raw data as input to the classification algorithm. Proper selection or design of features requires expert domain knowledge and experience. Transfer learning (TL) is an ML approach that remedies this by utilizing previous experience from one problem and transferring it to another. With TL, one avoids starting from scratch in every new problem, and is in theory not limited to problems in the same domain [70]. TL has been applied to bearing fault diagnostics, where it has been used to transfer knowledge across different feature domains [71], operating conditions [72] and across different machines [73]. However, the considered knowledge transfer in the existing works are still within the same field. A study on feature transfer to bearing fault detection from another field is missing in literature. Further, limited or restricted available data from HE and damaged faulty bearings is a pervasive problem for development of CM methods [7]. Acquiring data through real machinery measurements or laboratory test rigs is therefore very important in developing methods that are less reliant on historic data.

Deep learning based approaches such as autoencoders [74], convolutional neural networks (CNN) [75, 76, 77, 78, 79], deep belief networks (DBN) [80, 81, 82, 83], self-organizing maps (SOMs) [84], and various other deep network architectures [85, 72, 86, 87, 88] can also be used to either learn new features from raw data, or perform dimensionality reduction of existing features before classification. In bearing CM, raw data consists of sequences, each containing thousands to millions of data points. Temporal order and distance matters, and it may be necessary to consider long term dependencies, such as periodic transients from single-point faults. CNNs can solve this with multiple stacked convolutional and pooling layers with large kernels. This quickly becomes computationally demanding, and reduces temporal resolution. Text, speech and translation problems have successfully applied recurrent neural networks (RNN), which are able to retain information through a sequence. However, the ability is reduced with distance. Efforts to counteract this, such as long short-term memory (LSTM) networks [89], attention mechanisms [90, 91] and dilated hierarchical architectures have been applied [92] to capture long term dependencies in sequences. Despite some of the feature learning algorithms are classified as unsupervised, representative training data from a faulty condition is still required to perform classification. Feature learning on simulated raw vibration data with long term dependencies may replace actual fault data in the training phase, but such a

piece of work is hard to find in literature.

Proper selection and effective usage of the aforementioned ML algorithms requires expert knowledge in the development phase. If detection of a fault is more important than diagnosis, methods that allow identification of anomalous data can be applied. Detecting incipient faults in bearings does not necessarily mean that the bearing needs an immediate replacement. As an extension of fault detection, it is desirable to establish methods that allow health estimation and threshold setting for measurement values with control of statistical confidence level. Methods for threshold setting based on CIs with known probability distribution under HE condition have been proposed and applied to low-speed bearings [93, 7], utilizing statistical hypothesis testing. This approach for threshold setting allows a probability of false alarm control, but is limited to CIs with known, well-defined distributions. A generalization to arbitrarily distributed CIs is therefore valuable for detection of anomalous behaviour. This can be achieved through generative models, which is an ML approach aimed at learning unknown relations between observed data and an underlying, well-defined probability distribution. Generative adversarial networks (GANs) [94], variational autoencoders (VAEs) [95] and adversarial autoencoders [96] are all generative models that are successful in image and sequence generation. The VAE has also been used [97] for dimensionality reduction and feature learning an subsequent supervised classification of bearing fault diagnostics. However, using VAEs to quantify anomalies and calculate a HI has not been found in the recent research on bearing fault diagnostics.

As seen from the literature review, several shortcomings in CM of large, low-speed bearings are identified and addressed in this thesis. The contributions are further described in the following section.

1.4 Contributions

Over the course of the project, 6 papers have been published, accepted for publication or submitted to scientific conferences and journals. The content of these papers are the basis of this thesis. Contributions are focused towards the challenges of current CM methods for large, low-speed bearings identified in Section 1.3.

An initial literature review in Paper A identified CM methods aimed at large, low-speed bearings. The review also revealed that the vast majority of methods are aimed at failure modes producing weak, periodic transients in a time series, using advanced signal processing to enhance and isolate the signal of interest. The roller end wear failure mode has received little attention in research so far, and after performing testing on a top drive with a worn main bearing in Paper B, it was concluded that the fault signature from

roller end wear was not likely to exhibit periodic behaviour. A better understanding of this failure mode was deemed necessary for development of online detection and diagnosis methods. Therefore, a test rig for axial bearings was purpose-built at the University of Agder, designed to apply roller end damage during operation. The test rig is further described in Paper C. As it was established that conventional CM methods are not aimed at roller end wear detection, the remaining work was to utilize the feature extraction and classification capabilities of ML to develop automated CM methods suitable for single-point as well as roller end damage.

Paper D demonstrates that a CNN pretrained on images is capable of directly extracting features for binary classification of bearings in HE and damaged conditions. This results in a fast and simple implementation of a supervised, data-driven classifier. This study uses vibration data from single-point damage and AE data from roller end scratching in the test rig. As available data from large, low-speed applications is very limited, Paper E presents an approach to identify anomalous behavior and calculate a HI without requiring historic fault data. Autoregressive (AR) model coefficients of top drive acceleration measurements and test rig AE measurements were used as input features. As the AR coefficients did not have well-defined distributions, a VAE was trained to infer the relation between observed data and underlying Gaussian latent variables, which are assumed to generate the observed data. A HI was then calculated from the latent variables. Paper F also addresses the issue of limited historic fault data, but considers single-point damages alone. Instead of using historic data, a vibration model is developed to generate simulated signals. A deep 1D CNN architecture with dilated convolutions, designed to capture long-term dependencies in raw vibration data, learns general features to classify the faults.

1.4.1 Paper A: A Review of Methods for Condition Monitoring of Large, Slow-rotating Bearings

Summary: The paper is a brief introduction to state-of-the-art methods for bearing vibration modeling, data acquisition, processing and decision-making for large and low-speed applications. Top drive downtime and maintenance optimization are introduced as a motivation for research. Challenges with regards to speed and size are discussed, based on the assumption of single-point damage.

Contributions: The paper provides a review of CM methods in the context of large and low-speed bearings.

This paper has been published as:
Martin Hemmer, Tor I. Waag and Kjell G. Robbersmyr. A Review of Methods for Condition Monitoring of Large, Slow-rotating Bearings. Presented at: *30th Conference on Condition Monitoring and Diagnostic Engineering Management.* University of Central Lancashire, Preston, UK

1.4.2 Paper B: A Comparison of Acoustic Emission and Vibration Measurements for Condition Monitoring of an Offshore Drilling Machine

Summary: A top drive, that was taken out of operation for maintenance, showed arc-shaped scoring on the roller ends. The purpose of the experiments was to investigate the applicability of acceleration and AE-based measurement systems to this failure mode. The bearing was subject to an additional damage on the roller ends and then run with less lubricant. For reference, the worn bearing was then replaced with a HE one. For each level of damage, measurements were taken at different speeds. It was expected to observe one or more of the bearing fault frequencies, but this was not successful. The AE system was able to detect an increase in RMS, but without a characteristic frequency, the diagnosis was not possible by using signal processing. Similar results were obtained using vibration measurements, but with an inconsistent RMS increase. The process of replacing the worn bearing also changed the vibration characteristics of the machine, which actually led to an increase in acceleration RMS after overhaul. The characteristics of roller end wear are presented as an explanation for the lack of fault frequency components. The paper highlights the need of a better understanding of the failure mode characteristics.

Contributions: Online detection and diagnosis of roller end wear is challenging for current CM methods that rely on the detection of characteristic fault frequencies. An explanation for this behavior is proposed on the basis of observed roller end wear.

This paper has been published as:
Martin Hemmer and Tor I. Waag. A Comparison of Acoustic Emission and Vibration Measurements for Condition Monitoring of an Offshore Drilling Machine. *Proceedings of the Annual Conference of the Prognostics and Health Management Society 2017*, 278-285, 2017. ISBN 978-1-936263-26-4.

1.4.3 Paper C: Rib-Roller Wear in Tapered Rolling Element Bearings: Analysis and Development of Test Rig for Condition Monitoring

Summary: A model describing arc-shaped roller end scoring is proposed, based on the assumption that roller end wear is created by particles caught between the roller end and retaining rib in tapered axial bearings. The paper documents the functionality of a test rig developed to investigate CM methods for roller end wear in low-speed axial bearings. The test rig can accommodate bearings up to 240 mm in outer diameter. A permanent magnet motor with encoder allows a precise control of shaft rotation. The main feature of the test rig is that roller end scratches can be generated in a controlled environment during operation. It was also developed with oscillating motion testing in mind.

Contributions: This work studies how particles in the rib-roller contact area can create the characteristic arc-shaped scoring that has been observed on roller ends. A test rig for replicating the scoring is developed.

This paper has been published as:
Martin Hemmer, Kjell G. Robbersmyr, Tor I. Waag, Rolf Albrigtsen, Torfinn Pedersen, Thomas J. J. Meyer and Chloë Vercammen. Rib-Roller Wear in Tapered Rolling Element Bearings: Analysis and Development of Test Rig for Condition Monitoring. *Engineering Assets and Public Infrastructures in the Age of Digitalization.* Springer, in press. Presented at: *13th World Congress on Engineering Asset Management.* Stavanger, Norway

1.4.4 Paper D: Fault Classification of Axial and Radial Roller Bearings using Transfer Learning Approach

Summary: AE data from scratch formation in an axial tapered roller bearing and vibration data from OR and RE damage in a radial roller bearing was collected from two different test rigs. Wavelet-based spectrograms of waveform segments were calculated and stored as images. A pretrained version of the AlexNet CNN architecture was evaluated for use as a feature extractor and binary classifier of faulty and HE spectrograms. Three approaches were evaluated: extracted features used directly to train an SVM, dimensionality reduction of the feature using a sparse autoencoder before training an SVM, and fine-tuning of the AlexNet architecture to perform classification. The is an advantage as compared to building and training classifiers from scratch. The method requires less feature knowledge than conventional signal processing-based analysis. It also reduces the

need for skilled analysts and speeds up implementation and analysis time.

Contributions: Within this work, a pre-trained neural network can effectively extract features from spectrograms for fault detection in bearings subject to single-point faults and roller end scratch formation.

This paper has been published as:
Martin Hemmer, Huynh V. Khang, Kjell G. Robbersmyr, Tor I . Waag and Thomas J. J. Meyer. Fault Classification of Axial and Radial Roller Bearings Using Transfer Learning through a Pretrained Convolutional Neural Network. *Designs*, 2(4), 56, 2018. doi: 10.3390/designs2040056.

1.4.5 Paper E: Health Indicator for Low-speed Axial Bearings using Variational Autoencoders

Summary: Vibration data from a top drive and AE data from an axial bearing test rig was collected under degrading machine health conditions. Both datasets contain HE data and several stages of roller end wear. An AR model was fitted to segments of the data, and the coefficients were used as features for training a variational autoencoders (VAEs). In a VAE, an encoder network infers parameters of a latent generative random process from the observed training data, where the target latent distribution is assumed to be zero-mean multivariate Gaussian with identity covariance matrix. The approximately Gaussian distribution of the latent variables is proposed to calculate a HI with probability of false alarm control. Further, conditional VAEs (CVAEs) are used to train a single model conditioned on the mode of operation or different speeds in this case.

Contributions: VAEs is applied for unsupervised learning of a low-dimensional latent vector from observed data, which again is used for calculating a machine HI. The HI was shown to react to roller end wear and poor lubrication using data from a real top drive and artificial roller end wear from a test rig.

This paper has been submitted as:
Martin Hemmer, Andreas Klausen, Huynh V. Khang, Kjell G. Robbersmyr and Tor I . Waag. Health Indicator for Low-speed Axial Bearings using Variational Autoencoders. Under review at: *IEEE Access*.

1.4.6 Paper F: Simulation-driven Deep Classification of Bearing Faults from Raw Vibration Data

Summary: A deep 1-D CNN was trained to detect bearing faults using simulated fault data for training. A cyclostationary bearing vibration model is used to generate the faulty bearing signals, based on bearing specifications under some assumptions about the mechanical system. By using a hierarchical dilated architecture, high resolution raw vibration data can be analysed without preprocessing. The method is tested on a seeded fault dataset, and two run-to-failure datasets, including a low-speed test at 250 rpm. The network results in 100 % fault diagnostic accuracy for OR and IR faults, but misclassifies all RE faults as IR. In run-to failure tests, faults detection performance is similar to references in the literature, but the fault type was not consistent.

Contributions: This work proposes a deep 1-D CNN architecture capable of diagnosing bearing faults from raw, simulated vibration data. The use of generated training samples extends the applicability of ML to classification problems, where historic fault data is not available for training.

This paper has been published as:
Martin Hemmer, Andreas Klausen, Huynh van Khang, Kjell G. Robbersmyr, and Tor I. Waag. Simulation-driven Deep Classification of Bearing Faults from Raw Vibration Data. Accepted for publication in: *International Journal of Prognostics and Health Management (IJPHM)*. Accepted date: 15.08.2019.

1.5 Outline

The rest of this thesis is structured as follows. Chapter 2 summarizes experiments performed as a part of the study, with reference to relevant papers. Other datasets obtained from other sources utilized in the thesis are also briefly described. In chapter 3, the developed methods and associated results are outlined. Section 3.1 describes the roller end wear failure mode, while CM methods for roller end wear and single-point defects are covered in Sections 3.2 and 3.3, respectively. Chapter 4 draws overall conclusions and suggests further work.

Chapter 2

Experiments and Datasets

This chapter outlines the experiments performed under the course of this project, in addition to experiments performed by others, which were utilized in the publications. Further details are given in the respective papers. Table 2.1 summarizes which experiments are used in the different papers.

Table 2.1: Overview of datasets used in papers.

Paper	Experiment
A	-
B	1
C	1
D	2, 3A
E	1, 3B
F	4, 5, 6

Experiment 1 was performed on a top drive, which is the foundation of Paper B, described in Section 2.1. Section 2.2 describes experiment 2, containing localized damage in a radial bearing that was acquired using a test rig at RWTH Aachen University. Development of an in-house axial bearing test rig is covered in Section 2.3, which is based on Paper C. This test rig has been used to acquire two datasets, described in Section 2.3.1 and 2.3.2. Experiments 4-6, used in Paper F, are not performed as part of this project. Experiments 4 and 5 are publicly available, being useful for benchmarking the developed methods. Experiment 6 is an accelerated lifetime test performed on an existing in-house ball bearing test rig . This type of test was not possible on the developed axial bearing test rig within the project timeframe.

2.1 Experiment 1: Top Drive Test

Tests were performed on a top drive that was taken out of operation for maintenance. A similar top drive is shown in Fig. 2.1 [98], alongside a schematic drawing of the test setup, highlighting the main components. Measurements were taken in the axial direction of the main bearing. As the measurements were taken in a workshop, only the shaft self weight was applied.

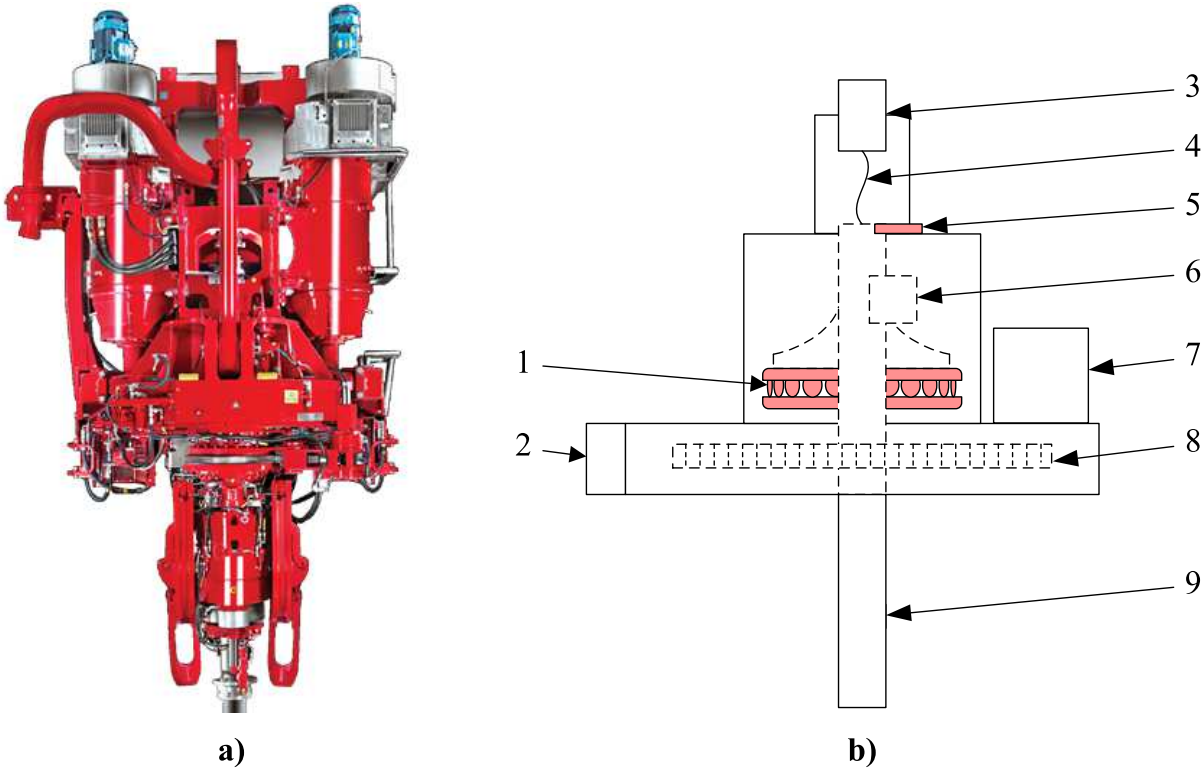


Figure 2.1: **a)** MHWirth top drive, similar to the one used in tests **b)** Schematic drawing of top drive test setup. 1: Main bearing, 2: Circulating pump, 3: Encoder, 4: Flexible coupling, 5: Axial sensor location, 6: Radial sensor location, 7: Hydraulic motor, 8: Drive gear, 9: Shaft.

The bearing has visible wear on the roller ends, which were discovered during visual inspection of the bearing as shown in Fig. 2.2. The wear has a characteristic arc-shape, being further described in Paper B and C.

After recording data from the worn bearing, artificial indentations were made using a chisel-like device with a carbide tip while the bearing was mounted inside the top drive. Boroscope images of the bearing before and after the artificial damage are shown in Fig 2.3. The lubricant was then drained below the indicated minimum level before repeating the measurement. Finally, the worn bearing was replaced with a new one, giving the HE

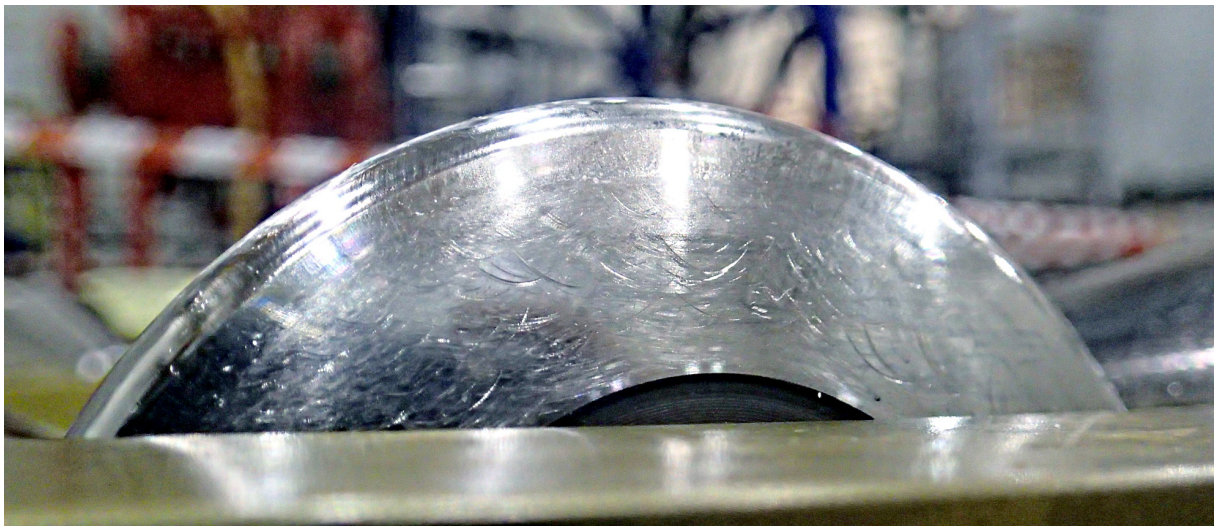


Figure 2.2: Damaged top drive bearing with arc-shaped scratches on the roller end.

reference. The machine was mounted in a temporary fixture with approximately 10 kN axial load from the shaft self weight. Data was collected from 50 rpm to 250 rpm in 50 rpm intervals.

a) **b)**

Figure 2.3: Artifically damaged top drive bearing. **a)** Boroscope image of roller end scratches on top drive main bearing. **b)** Boroscope image of artificially damaged bearing.

Paper B and E did not use the same data, therefore the naming is inconsistent. Table 2.2 gives an overview of naming used in the papers. The term *Damage level* (DL) will be used for the remaining of the thesis.

Vibration and AE measurements were performed in radial and axial directions to evaluate the measurement systems and technologies. The AE system used an analog demodulation circuit before digital sampling to reduce data volume [99].

Table 2.2: Experiment 1 naming matrix.

Document	Healthy	Worn	Damaged	Damaged + Poor lubrication
Thesis	DL0	DL1	DL2	DL3
Paper A	HL0	HL1		HL2
Paper D		DL1	DL2	DL3

2.2 Experiment 2: Radial Bearing Test

Experiment 2 was carried out at the Institute for Machine Elements at the RWTH Aachen University, using the test setup shown in Fig. 2.4a. Data from this experiment is used in Paper D. The test bearing is a purely radial roller bearing of type NU220 ECP.

Triaxial accelerometers are mounted on the test bearings inner and outer housing. Sampled data was also collected with the AE system as in dataset 1. Shaft rotation, torque, temperature and cage rotation were also recorded, but have not been used in papers. The test rig is further described in Paper D. The test bearing is shown in Fig 2.4, with the temperature sensor, AE sensor and accelerometer locations.

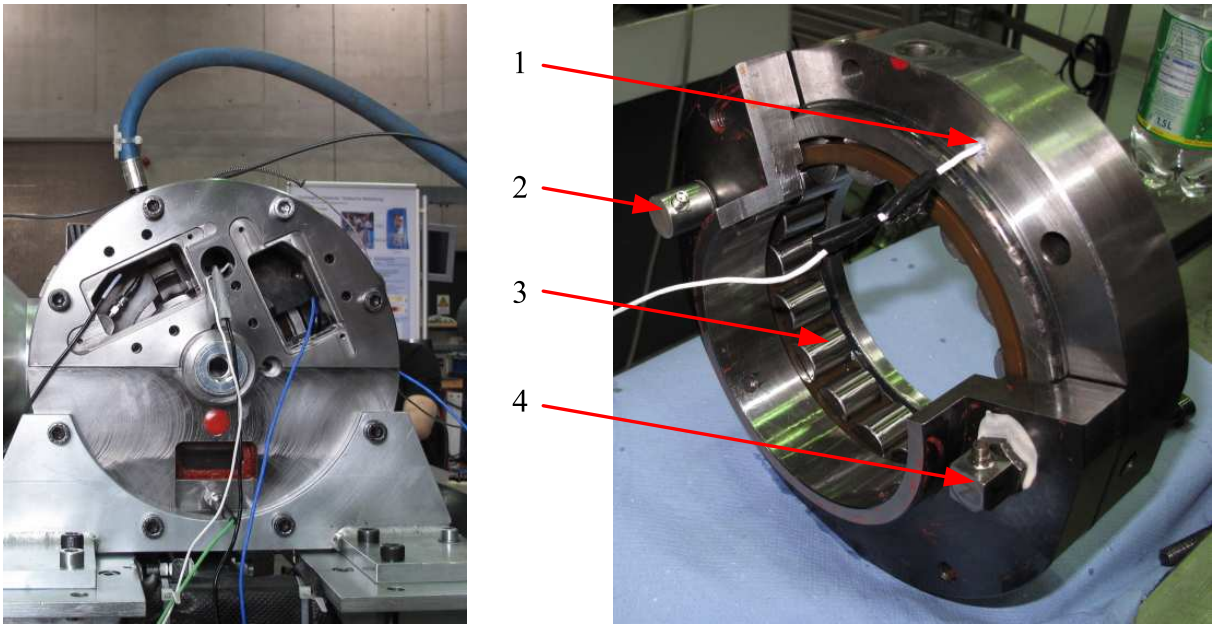
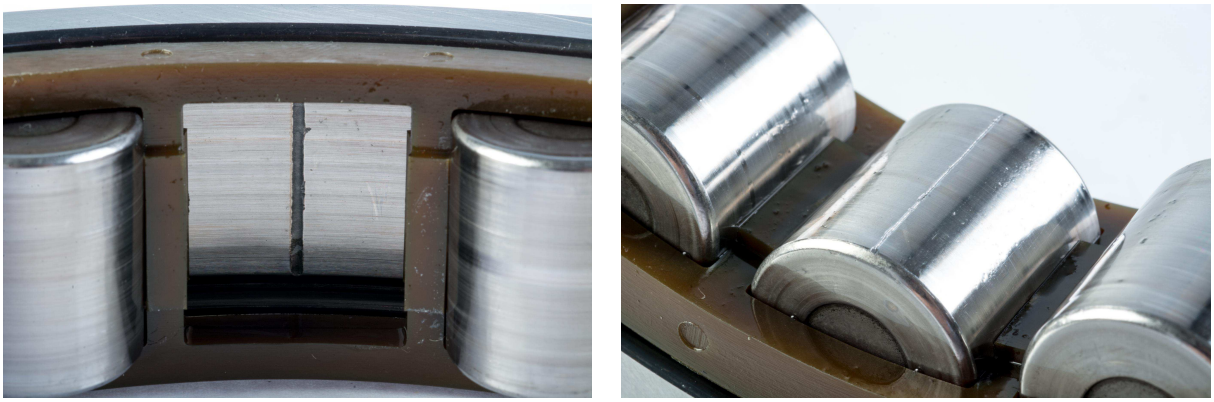


Figure 2.4: Test setup used in experiment 2. **a)** Test unit with bearing mounted. **b)** Test bearing in internal housing. 1: Temperature sensor, 2: AE transducer, 3: NU220 ECP test bearing, 4: Triaxial accelerometer

Two different fault types were tested: A very severe OR fault, and a smaller RE fault.

The faults were seeded using an abrasive tip tool. Both damages are shown in Fig 2.5. Each bearing was run at a range of loads and speeds, as described in Table 2.3.



a) **b)**

Figure 2.5: Faults in NU220 ECP bearing used in experiment 2. **a)** OR fault. **b)** IR fault.

Table 2.3: Dataset 2 test conditions.

Load [kN]	**Damage type**	**speed** [**rpm**]
1, 3	RE	100, 200, 300, 400, 500, 750, 1000
5, 10, 15, 20	OR, RE	100, 200, 300, 400, 500, 750, 1000

2.3 Experiment 3: Axial Bearing Test Rig

Based on the experience from the top drive test described in Section 2.1, it was decided to develop an axial bearing test rig to study wear on roller ends. An overview of the test rig and the bearing test unit is shown in Fig 2.6. The test rig is based on a hydraulic workshop press, where an axial load is applied to the bearings by a manually controlled hydraulic cylinder. An axial load of 50 kN was used in all experiments. A permanent magnet motor drives the shaft, connected with a claw coupling. The location of the AE sensor is also shown. Section 2.3.1 further describes the bearing test unit.

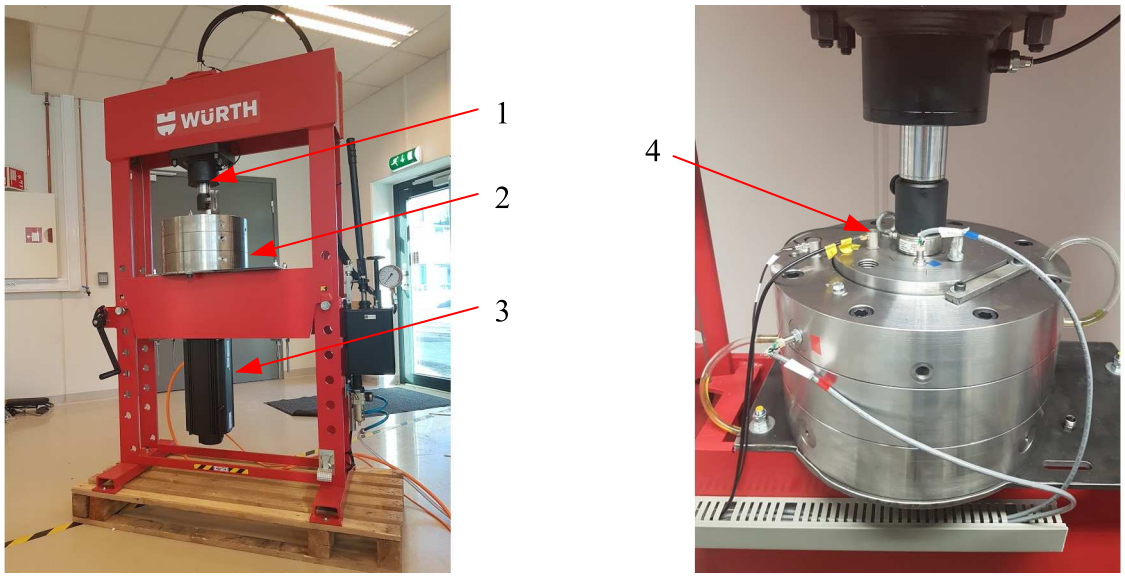


Figure 2.6: **a)** Axial bearing test rig overview. 1: Load cylinder, 2: Bearing test unit, 3: Permanent magnet motor **b)** Bearing test unit. 4: AE sensor

2.3.1 Experiment 3A: Artificial Scratch Test

Inside the test unit shown in Fig 2.7, a test bearing (ISB 29230) and a load bearing (ISB 29336) are mounted on the same shaft. Lip seals are used to separate lubricants for the test and load bearing. The test bearing is mounted with a stationary IR to give the damaging mechanism access to the rib roller contact area. Load is applied to the piston and test bearing, transferred through the shaft and load bearing back to the support structure. As shown in Fig. 2.7b, a needle (yellow) is passed through a hole in the OR to scratch the roller end. Movement of the needle is controlled manually by an adjustment screw (white) and a return spring. The damager needle is stationary while the rollers pass over, emulating a particle. Data from this experiment is used in Paper D.

The damaging mechanism was designed to imitate the formation of roller-end scratches on the load bearing as if a particle was caught between the retaining rib and roller end. A hole in the IR of the test bearing allows a hardened needle to damage the roller end during operation. Due to the brittleness of the needle, the scratch formation tests were conducted at 1 rpm. Experiments were aimed at detecting the formation of single scratches, not damage accumulation. Paper D utilizes AE from a sensor mounted on the load piston, recording raw data at 1 MHz. The resulting scratches on the roller end are shown in Fig 2.8.

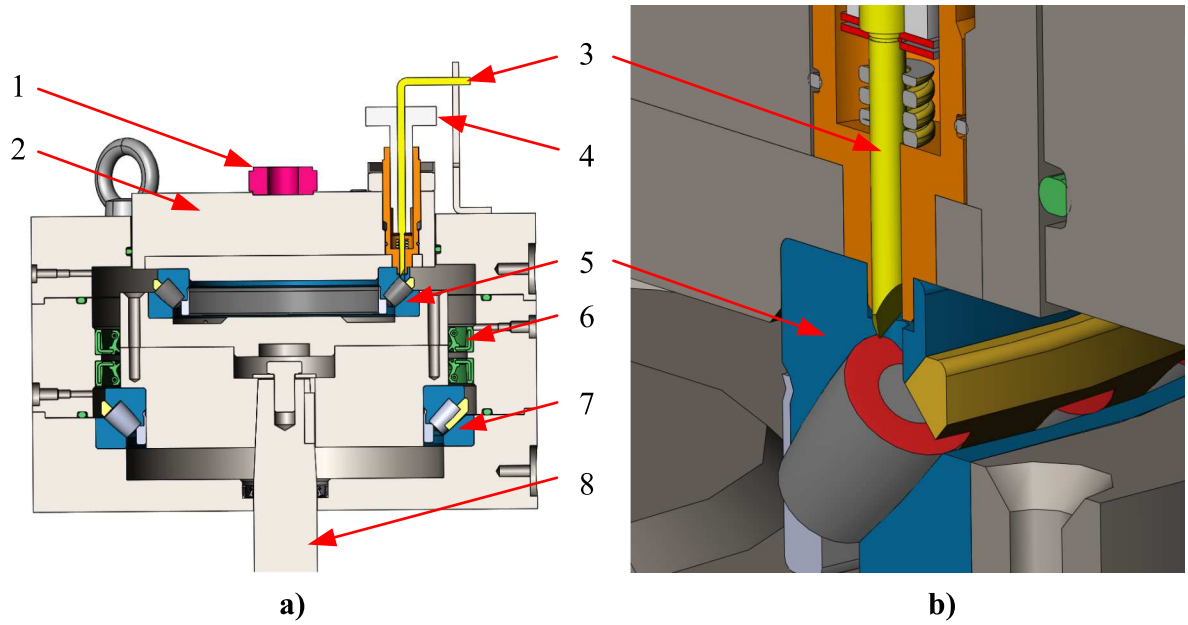


Figure 2.7: Bearing test unit. **a)** Test unit cross section. 1: Load cell, 2: Piston, 3: Damager needle, 4: Damager adjuster, 5: Test bearing, 6: Lip seal, 7: Load bearing, 8: Shaft. **b)** Roller end damager.

Figure 2.8: Roller end scratches created by the damaging mechanism in experiment 3A.

2.3.2 Experiment 3B: Artificial Abrasive Wear Test

In the second test, the roller ends of the test bearing were ground using sandpaper to emulate increasing levels of abrasive wear as the damaging mechanism was not durable enough to accumulate damage. Sandpapers with ISO/FEPA grit grades P400 to P80 were used. The terms *heavy* and *very heavy* refer to increasing severities with the same sandpaper grade. The roller ends before testing are shown in Fig 2.9. AE data was recorded using a measurement system that collected raw waveforms at 1 MHz with the bearing running at 30 and 60 rpm. Data from 30 rpm was recorded before 60 rpm, which is assumed to have an influence on the recorded data as the surface may be smoothened

during operation. This is supported by the results of Paper E.

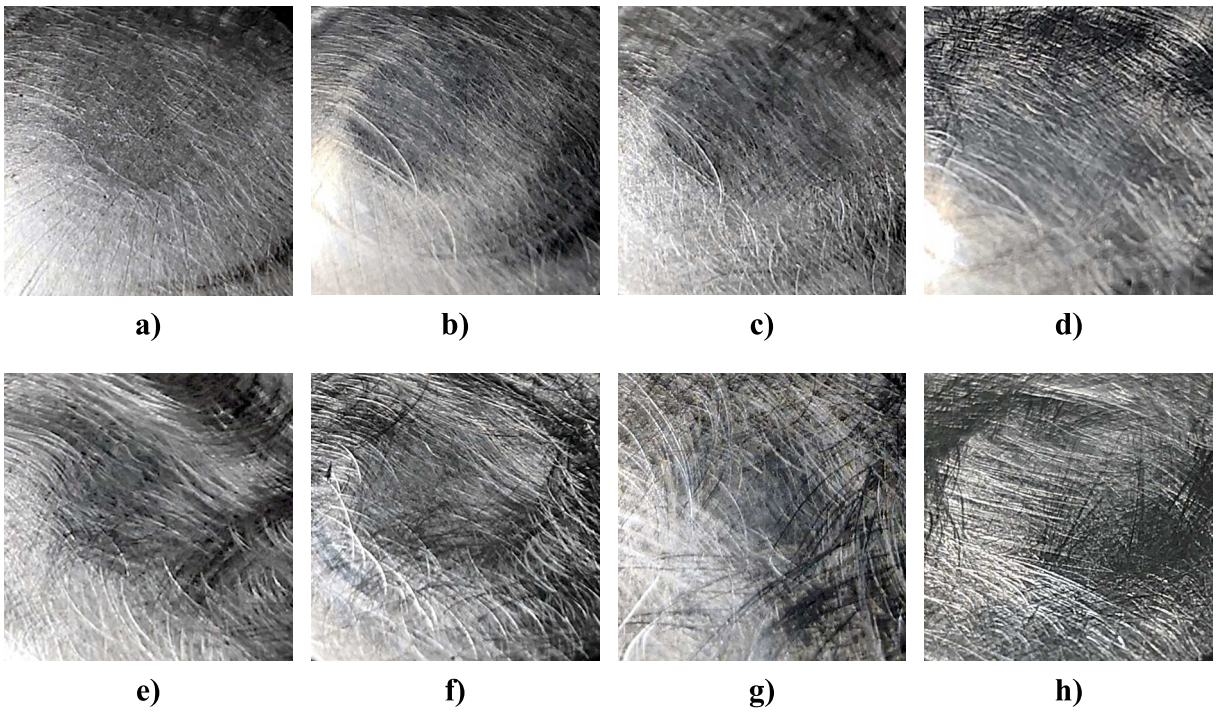


Figure 2.9: Increasing levels of the artificial roller end wear. **a)** DL1: P400, **b)** DL2: P400 (heavy), **c)** DL3: P320, **d)** DL4: P220, **e)** DL5: P220 (heavy), **f)** DL6: P80, **g)** DL7: P80 (heavy), **h)** DL8: P80 (very heavy)

2.4 Experiment 4: Constant Speed Seeded Fault Test

Case Western Reserve University (CWRU) provides a publicly available dataset of bearings with seeded faults [100]. The test setup consists of an induction motor and a dynamometer mounted face-to-face, being connected with a torque transducer and an encoder. The motor contains the test bearings, while the dynamometer acts as an external load. The dataset includes data records of separate seeded OR, IR and RE faults in the drive end, and fan end of a motor, with defect sizes from 0.17 mm to 1.02 mm in diameter. Shaft loads can be varied between zero and four hp. The shaft rate drops from 1797 rpm at no load to 1730 rpm at the maximum load. Data from this experiment is used in Paper F. The data utilized is from the smallest fault size in the drive end of the motor, running at 1797 rpm under a no-load condition with a sample rate of 48 kHz.

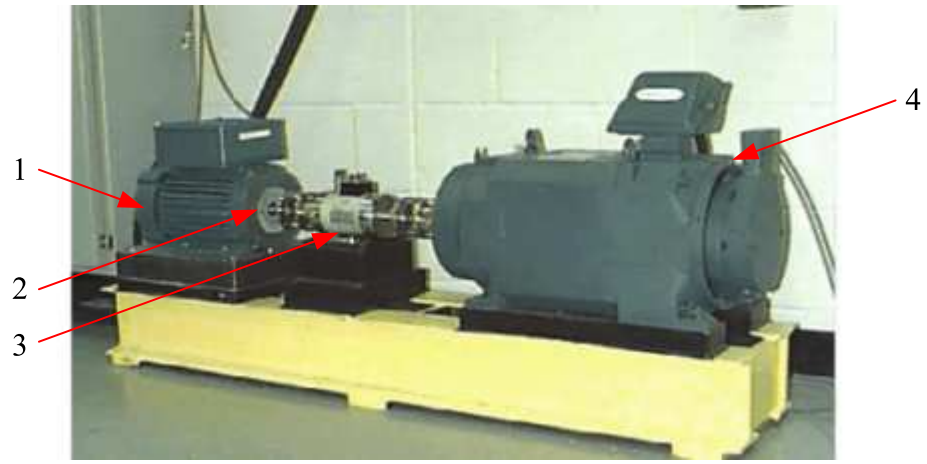


Figure 2.10: CWRU test rig [100]. 1: Fan end bearing, 2: Drive end bearing, 3: Torque sensor, 4: Dynamometer

2.5 Experiment 5: Constant Speed Lifetime Test

Dataset 5 is collected by the NSF I/UCR Center for Intelligent Maintenance Systems (IMS) on behalf of NASA Ames Prognostics Data Repository [101], including three run-to-failure datasets from a four-bearing test setup. A schematic drawing of the test setup is shown in Fig 2.11. The bearings were tested at a constant speed of 2000 rpm and a load of $\simeq$ 26700 N until failure. Further details can be found in a reference paper written by the researchers that performed the measurements [102]. Vibration signals from each bearing were sampled for one second at 20480 Hz with five to ten minute intervals. Data from tests 1 and 2 are used in Paper F.

2.6 Experiment 6: Low-speed Bearing Lifetime Test

An in-house test rig has been developed at the University of Agder for accelerated lifetime tests of a 6008 type bearing under low and variable speed conditions. The shaft speed is controlled by a permanent magnet synchronous motor, while electromechanical actuators apply up to 12 kN radial and axial load separately. An accelerometer is mounted perpendicularly to the load zone. Measurements from a temperature probe and an eddy current proximity sensor are also available, in addition to encoder data. The test setup is shown in Fig. 2.12. Vibration data from a bearing running at 250 rpm with a combined radial and axial load is used in Paper F. A thorough description of the test rig is given in [103].

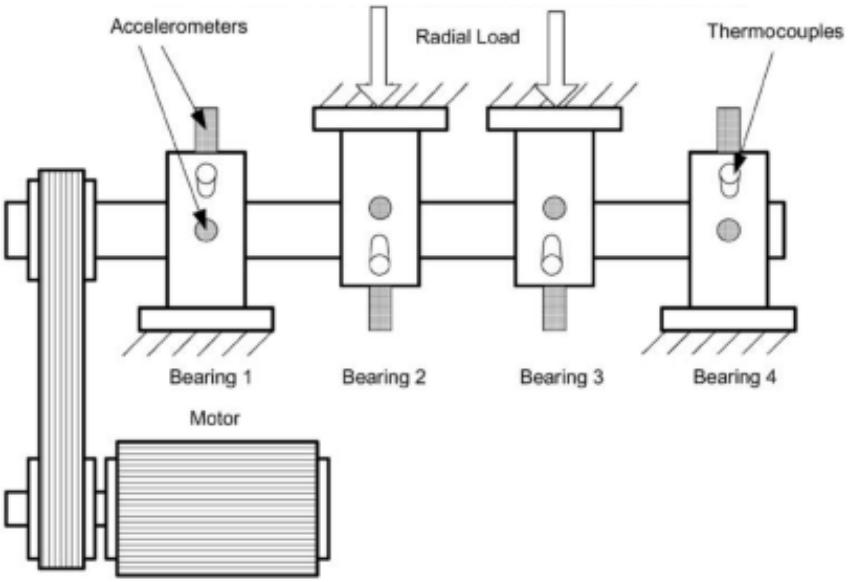


Figure 2.11: Schematic drawing of IMS test rig [101].

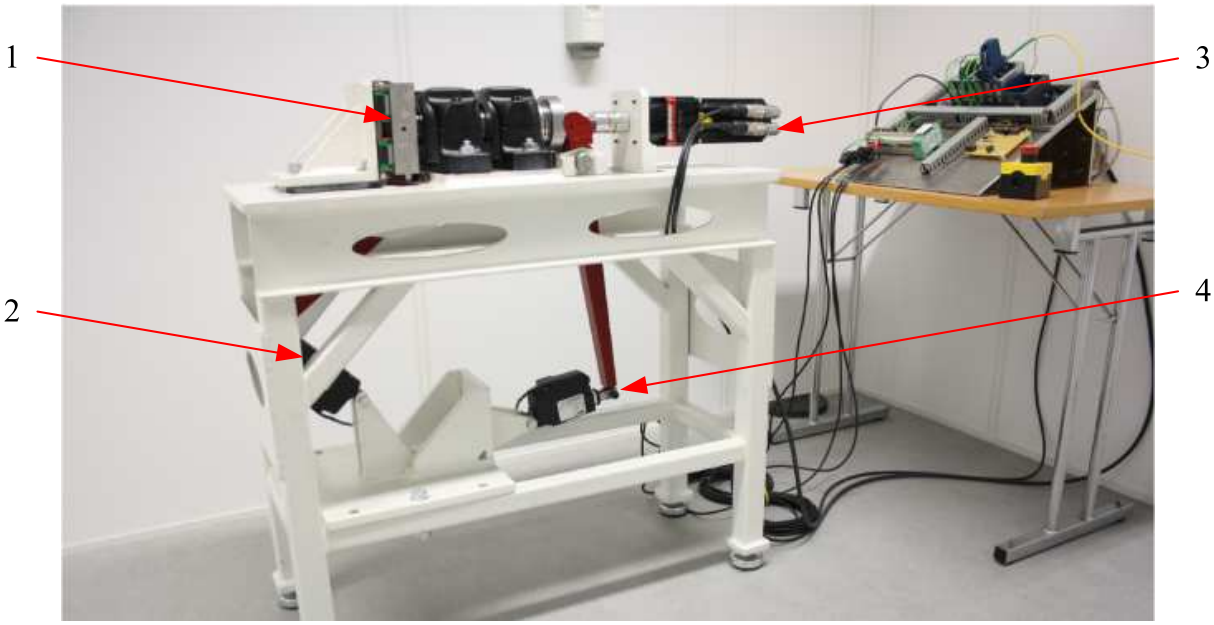


Figure 2.12: In-house test rig for low-speed lifetime testing [103]. 1: Test bearing, 2: Radial load cylinder, 3: Axial load cylinder, 4: Motor

Chapter 3

Methodology and Results

This chapter gives a summary of the results from this study, based on the appended papers. Section 3.1 describes the failure mode that was observed in the top drive test, setting the direction for the remainder of the study. In Section 3.2, TL is applied to detect the formation of roller-end scratches in the axial bearing test rig. A method for unsupervised learning of a HI for anomaly detection and threshold setting is proposed. The method is evaluated on data from the top drive test and an artificially worn axial tapered roller bearing. Section 3.3 focuses on single-point defects, applying TL for fault detection. In addition, a method for simulation-driven fault classification from raw time series data is proposed.

3.1 Roller End Wear in Tapered Roller Bearings

The majority of existing studies on bearing fault detection and diagnosis consider single-point surface defects in either bearing races or REs. In the top drive test described in Paper B, abrasive wear with a characteristic arc-shape was observed, distributed on the roller ends of an axial bearing. This section gives a description of the failure mode and provides an explanation for the characteristic shape of the scratches. An example of such scratches was shown in Fig. 2.2.

As described in Section 2.1 and Paper B, a series of tests were performed with the aim of identifying and comparing potential CIs using AE and acceleration measurements. As the damage was prominent on roller ends, it was expected to observe frequency components at the BSF, possibly with harmonics and shaft frequency sidebands due to amplitude modulation. Several forcing frequencies were more easily identified in the system using AE than acceleration, but none could be tied to the roller end scratches. An additional artificial damage was applied by creating indentations in a roller end, and the amount of lubricant was reduced to amplify this effect. Data was recorded using three different

systems, as described in Table 3.1. System B1 and B2 are identical, but were mounted at two different locations. Plots of the RMS levels at different speeds and DLs are shown in Fig. 3.1. Note that DL2 was not used in Paper B. Refer to Table 2.2 for naming of DLs. Baseline measurements with a HE bearing were performed, however, the acceleration RMS was actually higher than for the HE bearing than the damaged in some acceleration measurements. This can be due to both bearing run-in and differences in assembly, such as shaft alignment. Due to the low axial load, it is unlikely that any new scratches were formed during the experiments and that the main increase is due to insufficient lubrication. In the AE measurements, increasing DL results in an increased RMS level that is more consistent, particularly in the axial direction.

Table 3.1: Dataset 1 instrumentation

Feature	**System A**	**System B1/B2**	**System C**
Transducer	Acceleration	Acceleration	AE
Frequency	2 Hz -10 kHz	2 Hz -10 kHz	50-400 kHz
Sensitivity	10 mV/ms^{-2}	1.0 $\mu A/ms^{-2}$	69 dB (peak)
Sample rate	102.4 kHz	10 kHz	50 kHz
Recording	100 rev	60 s	120 s

This result led to a closer investigation of the failure mode. Tapered roller bearings can accommodate high axial or combined loads. The tapered shape creates a common apex point for roller rotational axes. However, it also requires a seating force to act on the roller end. This is achieved by a retaining rib on the outer race. Under rotation, there is a relative sliding motion between the roller end and the retaining rib. The combination of load and sliding motion makes this area sensitive to wear. Poor lubrication and contamination can create a metal-to-metal contact between the mating surfaces. The force balance of a single roller is shown in Fig. 3.2.

Neglecting friction, the seating force F_{seat} and normal force F_{norm} per roller are a function of taper angle β, the axial force F_{ax} and number of rollers n_r as in Eq. (3.1) and Eq. (3.2).

$$F_{seat} = \frac{F_{ax}}{n_r} sin(\beta) \tag{3.1}$$

$$F_{norm} = \frac{F_{ax}}{n_r} cos(\beta) \tag{3.2}$$

Considering a simplified tapered roller bearing, as shown in Fig. 3.3. Fig. 3.4 shows an "unwrapped", 2-D representation of the rib roller contact area. The roller end can be

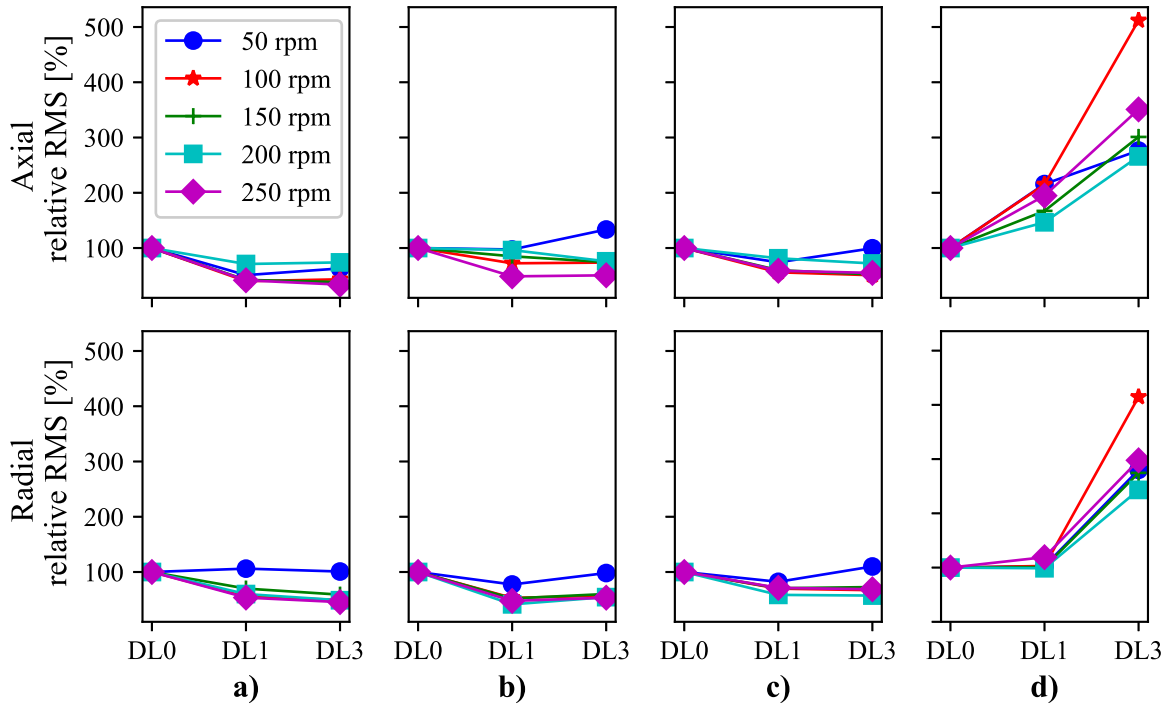


Figure 3.1: Top drive RMS trend in the axial and radial directions, presented as relative change from HE. **a)** vibration system A, **b)** vibration system B1, **c)** vibration system B2, **d)** AE system C.

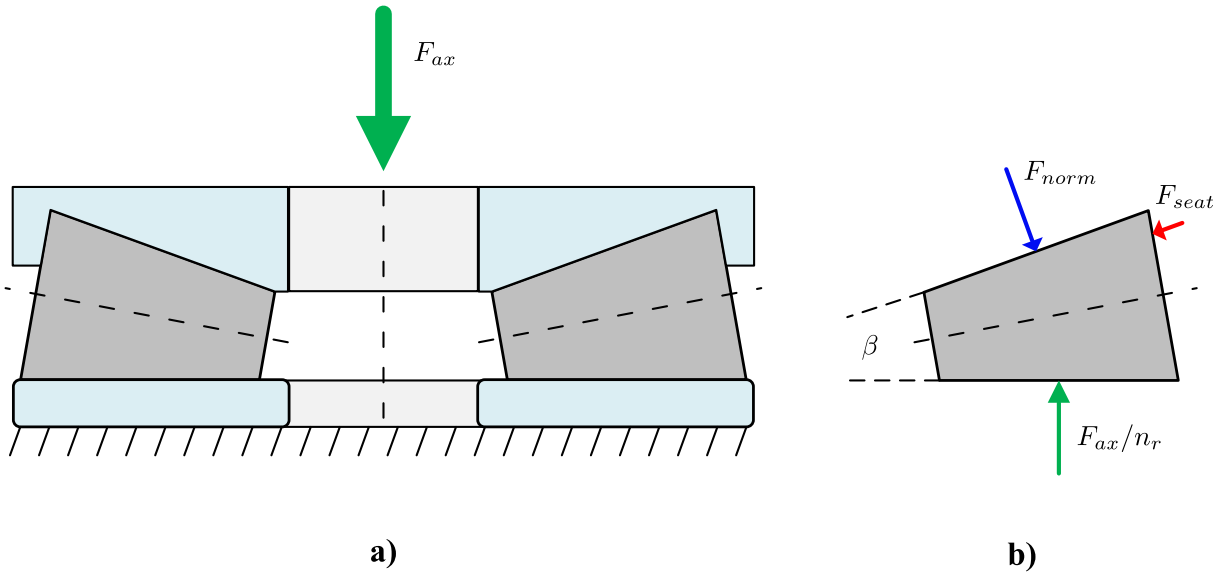


Figure 3.2: Axial tapered roller bearing forces. **a)** Axial tapered roller bearing with an axial load, **b)** Free body diagram of a single roller.

described as a smaller circle A, rolling on a larger circle with center B which represents the outer race. Let the circles have radius a and b, respectively. The OR retaining rib is shown in blue. Now consider a particle P between the roller end and the retaining rib. Given that the scratches are visible on the roller end only, the particle is assumed to be stationary with respect to the OR during scratch formation. The drawing is illustrative, and hence RE and OR diameter are not to scale.

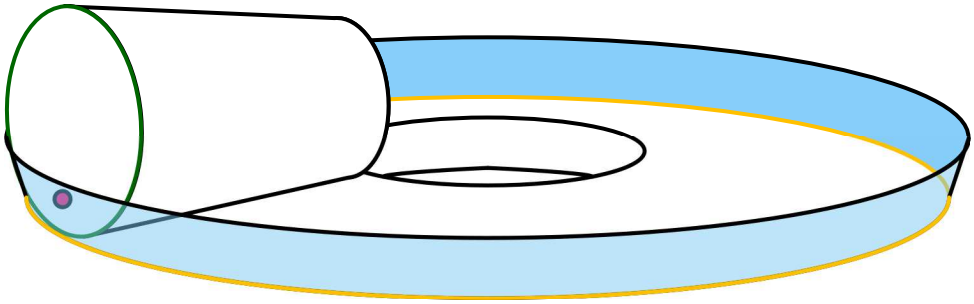


Figure 3.3: 3-D illustration of the model in Fig. 3.4.

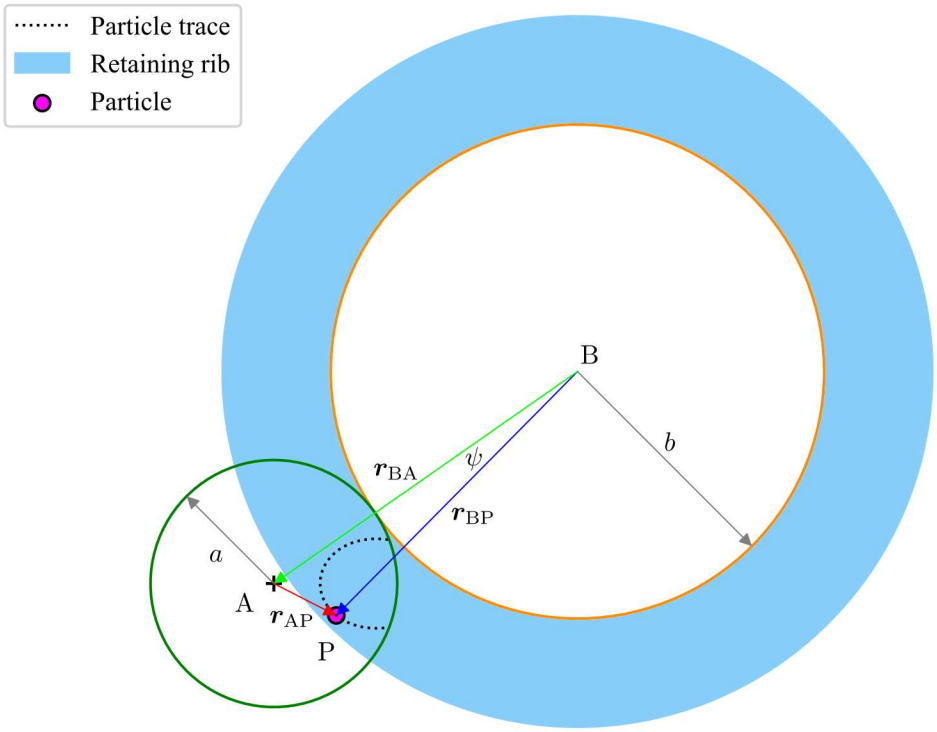


Figure 3.4: Simplified 2-D scratch model. Trace of particle P caught between retaining rib and roller end as A rolls on the perimeter of B.

Angle ψ between $\boldsymbol{r}_{BA}$ and $\boldsymbol{r}_{\mathrm{BP}}$ is a function of roller rotation and particle position vector $\boldsymbol{r}_{\mathrm{BP}}$. $|\boldsymbol{r}_{\mathrm{BP}}|$ is a constant scalar in the roller coordinate frame. For simplification, let γ represent the following relation: $\gamma = (a+b)/b$. The particle position $\boldsymbol{r}_{\mathrm{AP}}$ in the roller coordinate system can modeled as an epitrochoid, being given in Eq. (3.3).

$$\boldsymbol{r}_{\mathrm{AP}} = \begin{bmatrix} \mathrm{P}_x \\ \mathrm{P}_y \end{bmatrix} = \begin{bmatrix} \cos\psi & -\cos\gamma\psi \\ \sin\psi & -\sin\gamma\psi \end{bmatrix} \begin{bmatrix} b\gamma \\ |\boldsymbol{r}_{\mathrm{BP}}| \end{bmatrix} \tag{3.3}$$

The modeled trace resembles the observed failure mode, which supports the assumption of abrasive wear on roller ends originating from particles. As scratch formation was distributed on all rollers, the fault signature can not be expected to exhibit the periodicity of single-point faults. A particle may be trapped between roller and rib at arbitrary time. Thus, analysis methods that seek to identify characteristic fault frequencies were deemed unsuitable for this particular failure mode. Instead, the attention is focused towards ML for feature extraction, fault detection and classification.

3.2 Detection of Roller End Scratches and Wear

As discussed in Section 3.1, the top drive test did not result in identification of good CIs for roller end wear. To gather more knowledge of the failure mode and develop applicable CM techniques, a purpose-built test rig was developed to recreate the roller end scratches during operation. Test rig functionality and design are the topic of Paper C and Section 2.3. This section presents the data processing and analysis results. Two main aspects of roller end wear detection are considered in this section:

1. Detection of scratch formation during operation
2. Detecting a change in machine operational characteristics as a consequence of wear

The axial bearing test rig allows collecting labeled AE time-waveforms with and without scratches. The scratch detection is therefore considered a binary classification problem, determining whether a scratch is present in the time series or not. With access to the labeled data, the problem is reduced to a supervised classification.

3.2.1 Roller End Scratch Detection with Transfer Learning

In the context of ML, TL refers to the concept of taking knowledge acquired from one or more source tasks, and utilizing it to improve performance in another target task [70]. The approach has had some success in bearing fault detection by transferring knowledge of one fault type to another fault severity and operating conditions [72].

Using a CNN to categorize images is probably the most used example of supervised classification. CNNs extract information by convolving the data with a set of kernels, which act as feature extractors. By applying averaging and pooling operations combined with several convolutional layers, the network can learn features on different scales. Based on a large number of labeled images from a set of classes, a CNN learns to extract features and correctly classify unseen images from the same set of classes.

Designing a classifier requires several choices about hyperparameters such as layer type, filter size, number of layers and activation functions. In addition, training a deep neural network is time-consuming, requiring thousands to millions of training examples for multi-class problems. Learning a pre-trained network to classify a new class requires less training data. Hence, utilizing a pre-trained network has several advantages over building a classifier from scratch. In Paper D, it is therefore proposed to use a CNN pre-trained on image to classify spectrograms obtained from time series of HE and faulty bearings. In Paper D, three approaches to utilize the inherent knowledge in a pretrained network were investigated:

1. Replace the final classification layer and fine-tune the network weights to perform a new classification task.

2. Maintain network weights, and extract intermediate layer features that are used to train a separate SVM classifier.

3. Extract intermediate layer features. Train a sparse autoencoder (SAE) for dimensionality reduction of the features, then use the compressed representation to train a separate SVM classifier. This is denoted SAE-SVM.

With approach 1, the learned weights are fine-tuned and associated with a new class. In approaches 2 and 3, the idea is that CNN features become more specialized towards the network output, thus it may be beneficial to extract features from earlier layers when using TL and use them to train a separate classifier, such as SVM.

The AlexNet architecture was chosen for this application [104]. The overall architecture and classifier data flow for the three approaches are shown in Fig. 3.5. This network configuration was the winner of the 2012 *ImageNet Large Scale Visual Recognition Challenge*, which is a competition of classifying images from 1000 different classes. It has five convolutional and three fully-connected (fc) layers, and was considered deep at that time. The network was one of the first to train on GPUs, which is nowadays the de facto standard for training CNNs. Using a network trained on 1000 classes may be an exaggeration for a binary classification problem. However, the network has never seen bearing data during training and relies on utilizing already learned features to the new problem. The

AlexNet was therefore assumed to have a suitable balance between classification power and network complexity. The outputs of fully connected layers fc6 and fc7, as shown in Fig. 3.5, were used in Paper D. An advantage of using SVM is that the objective function is a quadratic optimization problem, unlike the gradient-descent based training of a CNN, which results in a faster training with a guaranteed optimal solution. The idea of training a SAE on the extracted features before a SVM is to obtain a compressed representation that best captures the information in the extracted features.

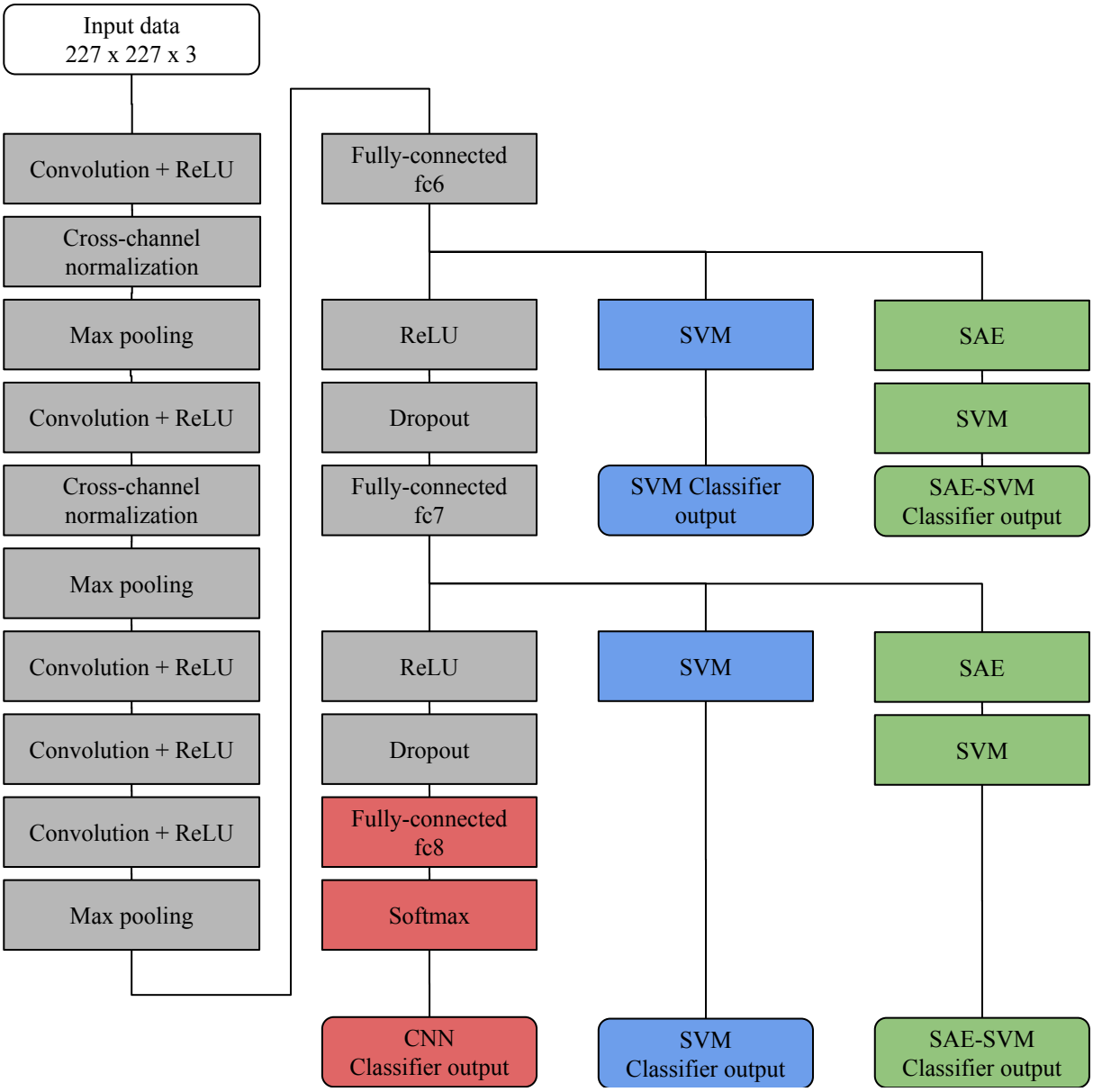


Figure 3.5: AlexNet architecture with branches showing feature extraction after layer fc6 and fc7, followed by the evaluated classifiers. Grey: Pretrained AlexNet architecture. Red: Fine-tuned CNN output layer. Blue: SVM classifier. Green: SAE-SVM classifier.

An important limitation of using pretrained networks is the data input format. Recorded

AE data are time series recorded at 1 MHz, while the AlexNet accepts 227x227x3 RGB images. First, the signal was split in 1-second segments and normalized to unit RMS. The transformation from 1-D time series to a 2-D image was then done by calculating a time-frequency representation of the signal, using the continuous wavelet transform (CWT) function in Matlab with a Morse mother wavelet [105]. The resulting image was resized to the required size.

The network was trained using 75 %, 50 % and 25 % hold out during training to test robustness against little training data. Results are averaged to obtain mean values for accuracy, probability of false alarm (P_{FA}) and probability of detection (P_D) for the different test cases. Results are given in Table 3.2. Overall, a SVM trained on features from a pretrained network was judged the best combination of performance and ease of implementation.

Table 3.2: Validation accuracy: Dataset 3.

	CNN	SVM fc6	SVM fc7	SAE-SVM fc6	SAE-SVM fc7
Accuracy	96.9%	97.8%	97.0%	94.9%	73.2%
P_{FA}	2.0%	1.0%	2.33%	0.67%	33.3%
P_D	96.0%	98.7%	99.3%	88.7%	22.7%

The presented results suggest that TL with a pretrained CNN is a fast way to implement ML-based fault detection. However, representative training data for all classes are required, limiting industrial applicability where historic data is often non-existent. Since detection of scratches is only the first step, a measurement for damage accumulation or severity is also required to make health assessment. The same TL approach is applied to fault detection of single-point faults, described in Section 3.3.

3.2.2 Health Indicator with Variational Autoencoder

This work represents the second approach to wear detection that is listed in Section 3.2, based on the contents of Paper E. Rather than detecting the formation of scratches, the goal is to detect a deviation from normal operating conditions that can be quantified in a single variable - a HI - and evaluated using statistical hypothesis testing. The concept of a HI is appealing, as it enables an asset owner or operator to easily evaluate and compare the current state of a machine. With measurements from HE operation, one can establish a baseline for normal condition and define a threshold based on that. The methodology

is demonstrated on gears and bearings using handpicked CIs and a whitening transform [93, 106, 7], as CIs are often correlated. Given variables with well-defined distributions, this process is efficient to perform hypothesis testing with a predefined significance level α. Let Y be a random variable which is a function of CIs. Further, let Y_0 be a subset of Y, calculated from CIs at HE operation only, with a defined cumulative distribution function ($F_{Y_0}(\cdot)$). The HI is obtained by normalizing a new observation y from Y with the inverse CDF at the desired significance α, as shown in Eq. (3.4)

$$HI = \frac{y}{F_{Y_0}^{-1}(1-\alpha)} \tag{3.4}$$

The probability of observing a HI larger than 1 when the system in nominal condition (HI_0), can then be interpreted as the P_{FA}, as given in Eq. (3.5),

$$P(HI \geq 1) = P_{FA} \tag{3.5}$$

In an industrial application, such as the top drive, a false positive may initiate a very costly maintenance operation, so the P_{FA} is required to be low. Additionally, online monitoring systems enable the equipment owner to perform measurements more often. The family-wise error rate (FWER), as defined in Eq. (3.6), is the probability of having one or more false alarms in q tests with a given threshold α.

$$FWER = 1 - (1-\alpha)^q \tag{3.6}$$

Consider a toy example of a CM system using only two sensors that perform measurements every 10 minutes. In 24 hours, that equals 288 tests. Using a threshold of three standard deviations, i.e. $\alpha \simeq 0.003$, will result in a $FWER = 58\,\%$, meaning that it is more likely than not to have at least one false alarm every day. It is clear that the threshold must be set significantly lower. In the paper, a threshold of 10^{-6} is used. In the toy example, this results in a $FWER = 0.02\,\%$.

Some CIs, such as RMS, are approximately Gaussian for HE bearings. Other CIs may have arbitrary distributions. It is therefore proposed to utilize the framework of VAEs to map arbitrarily distributed CIs to a smaller set of latent variables that follow a Gaussian distribution. Given observable data $\mathbf{x}$, the assumption is that observations are generated from a set of an underlying latent variables $\mathbf{z} \sim p(\mathbf{z})$ and the conditional distribution $p_\theta(\mathbf{x}|\mathbf{z})$ with parameters θ. The distribution of observed data $\mathbf{x} \sim p_d(\mathbf{x})$ is then given as in Eq. (3.7).

$$p_d(\mathbf{x}) = \int_{\mathbf{z}} p(\mathbf{z}) p_\theta(\mathbf{x}|\mathbf{z}) d\mathbf{z} \tag{3.7}$$

However, these distributions are generally not known, thus the integral is intractable. VAEs allow us to approximate $p(\mathbf{z}) \simeq q(\mathbf{z})$ under the assumption that $p(\mathbf{z})$ has a particular form, in this case, being a multivariate Gaussian with identity covariance matrix. The aggregated distribution $q(\mathbf{z})$ given input $\mathbf{x} \sim p_d(\mathbf{x})$ is then defined as in Eq. (3.8) [96].

$$q(\mathbf{z}) = \int_{\mathbf{x}} q_\phi(\mathbf{z}|\mathbf{x}) p_d(\mathbf{x}) d\mathbf{x} \tag{3.8}$$

The approximation is done by an autoencoder. The autoencoder is a neural network structure that consists of an encoder and a decoder network, which learn the parameters of conditional distributions $q_\phi(\mathbf{z}|\mathbf{x})$ and $p_\theta(\mathbf{x}|\mathbf{z})$, respectively. The autoencoder takes data $\mathbf{x}$ as input, and uses the encoder to compress $\mathbf{x}$ into a latent vector, $\mathbf{z}$. The decoder then attempts to reconstruct the input, denoted $\hat{\mathbf{x}}$. The latent space is of lower dimension than the data space, so the encoder is forced to learn features that capture the most information in $\mathbf{z}$. This is done by minimizing reconstruction loss $\mathcal{L}_R$, which is implemented as shown in Eq. (3.9), for a batch $\mathbf{x}$ of size M with N features per datapoint in the batch.

$$\mathcal{L}_R = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{N} (\mathrm{x}_{i,j} - \hat{\mathrm{x}}_{i,j})^2 \tag{3.9}$$

VAEs, proposed in [95], also have an encoder-decoder structure. However, VAEs are generative models that provide an efficient way to approximate variational parameters of a model when closed-form integrals are intractable or sampling-based approaches, e.g. Monte Carlo simulation, is considered too computationally expensive[95]. To control the latent distribution $q(\mathbf{z})$, Kulback–Leibler (KL) divergence (D_{KL}) between the encoder output and the target distribution are added as a loss ($\mathcal{L}_{KL}$) for regularization. In practice, the encoder outputs parameters of a distribution $\boldsymbol{\mu}$ and $\log(\boldsymbol{\sigma}^2)$, each of length J, which is the latent variable dimension. Minimizing the D_{KL} is equivalent to maximizing the evidence lower bound (ELBO) of the marginal probability of a data point [95]. If the target distribution is a Gaussian $p(\mathbf{z}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the $\mathcal{L}_{KL}$ for a batch is as given in Eq. (3.10).

$$\mathcal{L}_{KL} = \frac{1}{M} \sum_{i=1}^{M} D_{KL}\left(q(\mathbf{z})||p(\mathbf{z})\right) = \frac{1}{2M} \sum_{i=1}^{M} \sum_{j=1}^{J} \left(1 + \log\left(\sigma_{i,j}^2\right) - \mu_{i,j}^2 - \sigma_{i,j}^2\right) \tag{3.10}$$

As gradients can not flow through stochastic nodes, the "reparameterization trick" is applied by introducing an auxiliary random variable $\boldsymbol{\epsilon}$ [95]. This allows us to train VAEs using backpropagation. Given an output $\boldsymbol{\mu}$ and $\log(\boldsymbol{\sigma}^2)$ from the encoder, the latent

vector $\mathbf{z}$ is calculated as in Eq. (3.11).

$$\mathbf{z} = \boldsymbol{\mu} + exp\{0.5 \log(\boldsymbol{\sigma}^2)\}\boldsymbol{\epsilon} \tag{3.11}$$

An overview of the network architecture alongside a more detailed network configuration is shown in Fig. 3.6. Only a single hidden layer was used in the experiments, as it was found that more layers did not improve performance, but instead caused convergence to be more unstable.

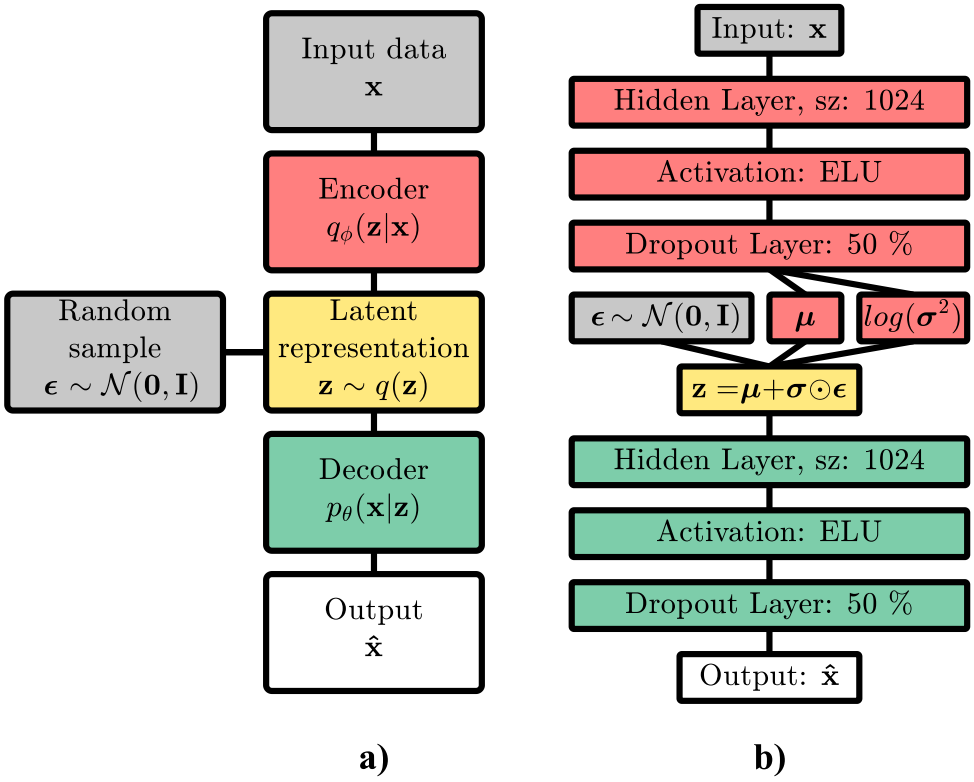


Figure 3.6: **a)** Overview of the VAE architecture. **b)** VAE layer configuration. Colors indicate network architecture affiliation.

As a standard VAE is trained to reconstruct the input while maintaining a Gaussian latent space, one network must be trained for each operating condition. If given multiple classes, the network will cluster similar samples within the Gaussian, and the HI calculation is not valid. The CVAE is an extension to the VAE, where the network is conditioned on an auxiliary variable. In this case, input data and latent vector are concatenated with a one-hot encoded label vector representing different speeds. Experiment 2 contains data from 5 different speeds (50, 100, 150, 200 and 250 rpm), thus 100 and 250 rpm datapoints will be concatenated with a conditioning vector $c_{100} = [0, 1, 0, 0, 0]$ and $c_{250} = [0, 0, 0, 0, 1]$, respectively. This allows a single network to learn representations of all available operating conditions with shared weights, exploiting similarities between two

speeds while maintaining a Gaussian latent space for HE data. The larger dataset will reduce the risk of overfitting at the cost of lower learning capacity, as the same number of weights must learn more than one category of data.

In Paper E, both VAEs and CVAEs were trained on data from the top drive test (experiment 1), and artificial abrasive roller end wear test (experiment 3B) as described in 2.1 and 2.3.2, respectively. In both cases, input data to the model are the coefficients of an AR model of order p, $AR(p)$, fitted to waveform data. Data from experiment 1 showed periodicity in the acceleration autocorrelation function, and was therefore stationarized by differentiation before fitting the AR model. By differentiating, the signal is effectively transformed from acceleration to jerk, enhancing high frequency components. The appropriate model order, i.e. number of coefficients, was determined by calculating the partial autocorrelation function (PACF). The PACF at lag $p+1$ describes the autocorrelation, which is not accounted for by the previous p lags [107]. Model order is considered sufficient when PACF is 0 with 95 % confidence level, calculated as in Eq. (3.12). The number of datapoints in the time series is denoted N.

$$PACF \leq \frac{\pm 1.96}{\sqrt{N}} \tag{3.12}$$

For each record in the dataset, AR model order p was determined based on the PACF. Table 3.3 shows the statistics and selected value for p. In the CVAE, input data was zero-padded to accommodate different values of p in a single network.

Table 3.3: AR model order and statistics.

	DS1-50	DS1-100	DS1-150	DS1-200	DS1-250	DS2-30	DS2-60
Mean (p)	32	12	20	16	15	14	25
STD	9	9	13	9	10	10	11
Median	29	5	23	9	9	7	17

In the VAE loss function, $\mathcal{L}_R$ measures how well the input data is reconstructed. However, $\mathcal{L}_{KL}$ does not show directly how well the aggregated distribution, i.e. sampled values of $\mathbf{z}$, approximates the desired Gaussian prior. To validate this, D_{KL} between a batch of M $\mathbf{z}$'s, $q(\mathbf{z}_M)$ and $p(\mathbf{z})$, as given in Eq. (3.13). Σ_M is the covariance matrix of $\mathbf{z}_M$ and $tr(\cdot)$ is the matrix trace operator.

$$D_{KL}\left(q(\mathbf{z}_M)||p(\mathbf{z})\right)) = \frac{1}{2}\left(tr(\Sigma_M) + \boldsymbol{\mu}_M^T\boldsymbol{\mu}_M - J + \log_e\left(\frac{1}{|\Sigma_M|}\right)\right) \tag{3.13}$$

Table 3.4 compares the loss functions and aggregated KL divergence values at the end of VAE training. Reconstruction loss, $\mathcal{L}_R$, has positive correlation with model order p

Table 3.4: Final loss values after training of the VAEs.

	$\mathcal{L}_R$			$\mathcal{L}_{KL}$			D_{KL}		
	train	val	test	train	val	test	train	val	test
DS1-50	20.668	23.130	0.000	3.484	3.266	2.106	0.033	0.017	0.025
DS1-100	4.432	5.375	0.000	2.765	2.550	1.950	0.033	0.033	0.050
DS1-150	9.814	10.397	0.000	3.205	2.991	2.085	0.038	0.044	0.021
DS1-200	6.987	8.778	5.400	3.001	2.853	2.071	0.028	0.091	0.112
DS1-250	5.328	6.090	6.215	3.178	2.739	1.838	0.020	0.067	0.054
DS2-30	5.255	5.670	6.215	2.844	2.803	2.099	0.050	0.010	0.007
DS2-60	7.418	7.658	7.236	3.586	3.452	1.943	0.041	0.019	0.017

Table 3.5: Final loss values after training of the CVAE.

	$\mathcal{L}_R$			$\mathcal{L}_{KL}$			D_{KL}		
	train	val	test	train	val	test	train	val	test
DS1	14.453	14.080	13.977	2.910	2.825	1.988	0.068	0.007	0.008
DS2	5.302	6.303	8.389	2.702	3.015	2.748	0.038	0.005	0.020

in both datasets. This is expected, as the formulation of $\mathcal{L}_R$ sums square error over the features in a datapoint. $\mathcal{L}_{KL}$, is relatively stable for all datasets and does not appear to be related to $\mathcal{L}_R$. This is reflected in D_{KL}, which is close to 0 for all datasets. This confirms that $\mathbf{z}$ approximates the target Gaussian distribution.

Loss values for the CVAEs, shown in Table 3.5, align well with losses with the VAE. Reconstruction loss is higher than the average of the individual VAE losses, indicating that reconstruction per dataset is not as good. $\mathcal{L}_{KL}$ and D_{KL} are also in the same range. For a given level of $\mathcal{L}_{KL}$, lower reconstruction loss means that the latent variables are able to capture more of the information in the data.

It was selected to calculate the HI based on the norm of the latent vector, $y = \|\mathbf{z}\|$. For $\mathbf{z}$ of dimension J, Y_0 is χ-distributed with J degrees of freedom. The HI is then calculated as shown in Eq. (3.4).

Fig. 3.7 shows the HI for the different test cases in dataset 1, calculated using both separate VAEs for each rpm (left column) and a single CVAE for all rpms. As seen in the box plots, the variance of the predicted HI increases with the value. Whiskers are set to 2.5th and 97.5th percentile. In the following discussion, HI value refers to the median in each DL, marked as the orange horizontal line inside the boxes. The alarm level is set to $HI = 1$, and an alarm level at $HI = 0.75$ There is a consistent HI increase for all DLs at all rpms. DL2 indicates a warning for all rpms except for 150. The alarm threshold of 1 is exceeded at DL 3 in all rpms. HI calculated using CVAE shows the same overall

pattern, with some minor variations. There is no clear relationship between HI and VAE or CVAE at different speeds. All HI values are shown in Table 3.6.

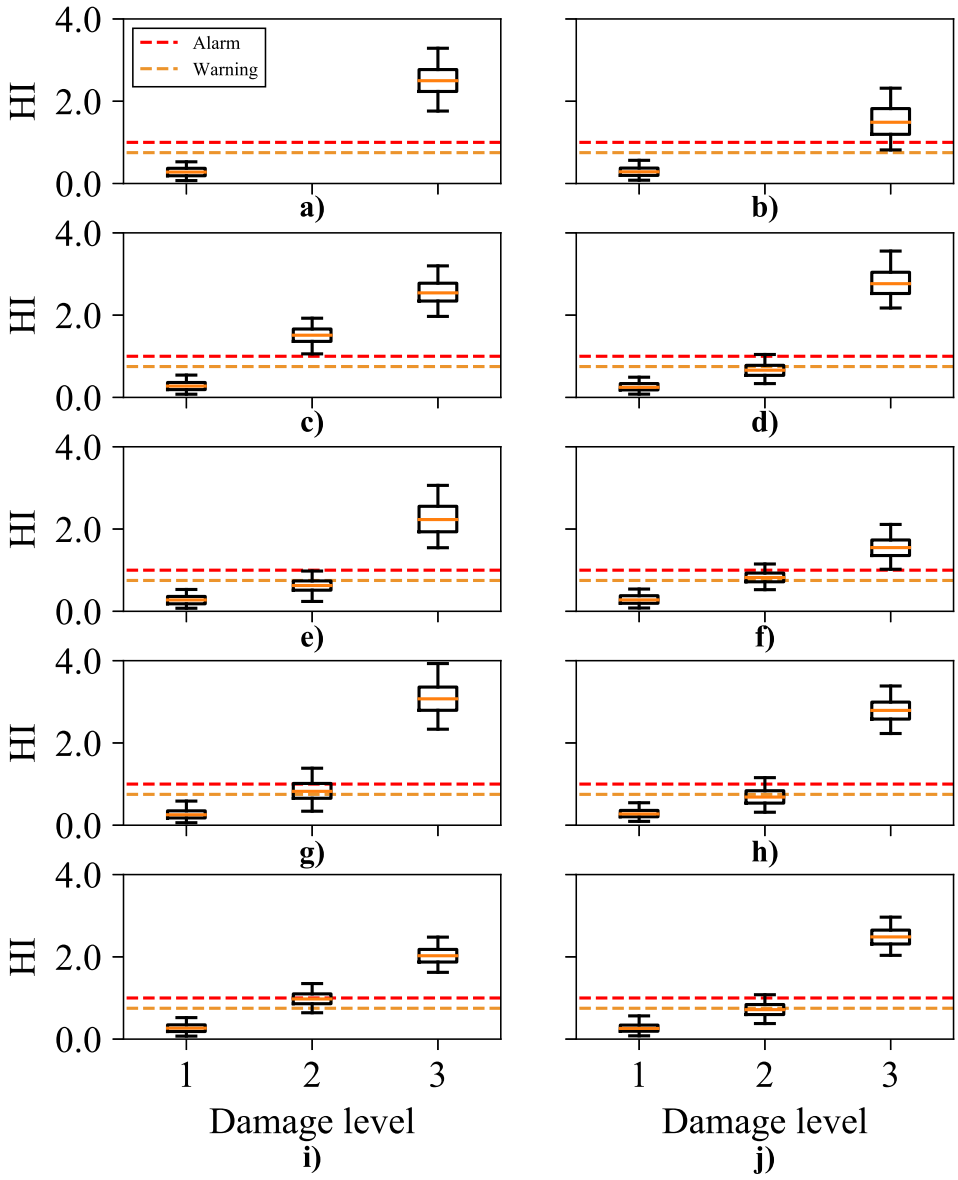


Figure 3.7: HI for top drive data. Left column is calculated using a normal VAE, and right column with a CVAE. **a-b)** 50 rpm, **c-d)** 100 rpm, **e-f)** 150 rpm, **g-h)** 200 rpm, **i-j)** 250 rpm.

Data from experiment 3B shows a less consistent HI trend. At 30 rpm, DL2 and upward has a HI well above the alarm threshold, as shown in Fig. 3.8. Except being consistently above alarm limit, there is little correlation between DL and HI. Results for 60 rpm data shows an overall similar trend, but at a lower HI value. Table 3.7 lists HI values for all rpms and DLs in experiment 3B.

Several factors may have contributed to this behavior. Firstly, the machine and bearing

	DS1-50		DS1-100		DS1-150		DS1-200		DS1-250	
DL	VAE	CVAE	VAE	CVAE	VAE	CVAE	VAE	CVAE	VAE	CVAE
DL1	0.28	0.28	0.28	0.25	0.28	0.28	0.25	0.27	0.27	0.26
DL2	-	-	1.51	0.66	0.63	0.82	0.82	0.68	0.98	0.72
DL3	2.50	1.49	2.54	2.76	2.23	1.55	3.07	2.79	2.03	2.48

Table 3.6: HI values for experiment 1

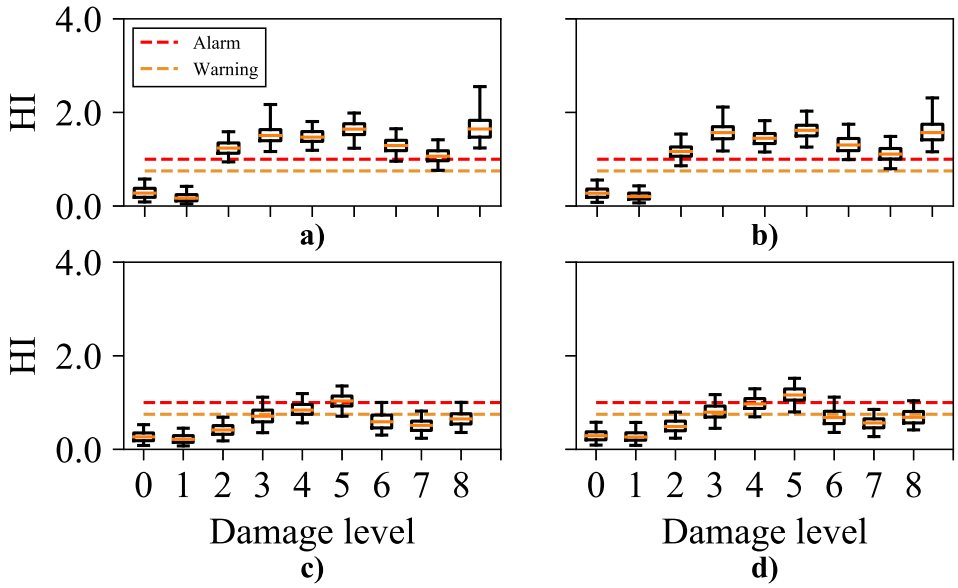


Figure 3.8: HI for axial test rig data. Left column is calculated using a normal VAE, and right column with a CVAE. **a-b)** 30 rpm, **c-d)** 60 rpm.

had to be disassembled to inflict damage on the roller ends. As experienced in Paper A, disassembly can have a large impact on recorded data, even though AE appeared more resistant in that particular case. Secondly, damage is inflicted manually using sandpaper on the roller ends. This may cause inconsistent severity, even if increasingly coarse grit grades were used. Finally, the artificial wear on roller ends may have been smoothened by operating during data collection. Data from 60 rpm was recorded after 30 rpm, which can explain the lower level but similar trend of the HI.

The results show that using VAEs for anomaly detection in large, axial bearings is feasible. Coefficients from AR models using both AE and (differentiated) acceleration data are shown to be capable of capturing differences in operating characteristics, and the inflicted damage resulted in a HI above the alarm level of 1 in both datasets. Further, conditioning on speed generalizes the methodology to more operating conditions in the same network while maintaining a similar HI level.

	DS1-30		DS1-60	
DL	VAE	CVAE	VAE	CVAE
DL0	0.28	0.27	0.27	0.29
DL1	0.17	0.20	0.21	0.26
DL2	1.24	1.17	0.41	0.49
DL3	1.51	1.57	0.71	0.79
DL4	1.47	1.45	0.84	0.97
DL5	1.64	1.62	1.03	1.16
DL6	1.29	1.30	0.59	0.68
DL7	1.06	1.11	0.51	0.56
DL8	1.65	1.57	0.65	0.69

Table 3.7: HI values for experiment 3B

3.3 Detection and Diagnosis of Single-point Defects

3.3.1 Single-point Fault Detection with Transfer Learning

In Paper D, the TL methodology described in Section 3.2.1 was applied to single-point surface defects as well as roller end scratches. Data from experiment 2, described in Section 2.2, was used in testing. The same binary classification setup was used, using both OR and RE fault as the damaged state. In this experiment, the actual faulty signal was corrupted with additive noise, yielding SNR levels of 0 dB, -3 dB and -10 dB relative to the uncorrupted signal. As in Section 3.2.1, training was done using 75 %, 50 % and 25 % of the total dataset as training data. A -10 dB SNR level was the only to show a clear drop in performance, particularly for RE fault. This was expected, as this fault was much less severe. Less training data did not have a consistent effect on performance, presumably due to a homogenous dataset. Table 3.8 and 3.9 gives the classifier accuracy, P_{FA} and P_D for OR and RE damage respectively, averaged for all SNR levels and training data holdout. Training on all three classes simultaneously was not tested, but should be considered for further work.

Table 3.8: Validation accuracy: Dataset 2, OR damage.

	CNN	SVM fc6	SVM fc7	SAE-SVM fc6	SAE-SVM fc7
Accuracy	98,05 %	99,11 %	98,87 %	97,43 %	97,62 %
P_{FA}	1,19 %	0,79 %	0,87 %	2,22 %	2,22 %
P_D	95,71 %	98,81 %	98,10 %	96,19 %	97,14 %

Table 3.9: Validation accuracy: Dataset 2, RE damage.

	CNN	SVM fc6	SVM fc7	SAE-SVM fc6	SAE-SVM fc7
Accuracy	94,11 %	94,35 %	92,86 %	90,60 %	91,31 %
P_{FA}	1,98 %	2,54 %	3,81 %	4,92 %	4,52 %
P_D	82,38 %	85,00 %	82,86 %	77,14 %	78,81 %

TL showed promising results for fast and simple implementation of supervised classification of bearing faults. Despite this, the approach was not pursued further in this project. This decision was based on multiple issues with the approach:

1. Supervised classification requires training data for all relevant fault types, which is often not available in an industrial application.

2. Restrictions on input data dimension result in loss of detail for high-resolution time series. This could limit further performance improvements.

3. Spectrogram calculation is computationally demanding for long time series with high time and frequency resolution.

3.3.2 Simulation-driven Single-point Fault Diagnosis

This section is based on Paper F, which addresses the previously issue of utilizing deep learning when representative fault data is not available. Given the bearing specification, it is possible to generate simulated vibration signals. A simple vibration model for the fault signal was developed. A fault signal $\boldsymbol{x}_F$ is modeled as amplitude modulated bandpass-filtered noise $\boldsymbol{w}_{bp}$. The modulation function was obtained by convolving an exponentially decaying function $\boldsymbol{h}$ of duration t_h with an impulse train $\boldsymbol{p}$. The impulse period corresponded to the characteristic fault frequency. Random jitter in time between impacts and impulse amplitude is also included in $\boldsymbol{p}$, to obtain a pseudo-cyclostationary signal [3]. Amplitude modulation function $\boldsymbol{m}_1$ was added to model the fault transition through the load zone with f_s and FTF for IR and RE faults respectively. Modulating function $\boldsymbol{m}_2$ models the alternating IR and OR impacts for RE damage. The fault signal equation is shown in Eq. (3.14).

$$\boldsymbol{x}_F = \left((\boldsymbol{m}_1 + \boldsymbol{m}_2)\, \boldsymbol{p} * \boldsymbol{h} \right) \boldsymbol{w}_{bp} \tag{3.14}$$

Any parameter that is not known is modeled as a random variable instead of a fixed value. The assumption is that by defining the distributions wide enough, a subset of the simulated signals will approximate the real faults.

3.3.3 Data Preprocessing

Each dataset consists of N records with duration t_r. In paper F, some records are used directly and some are split in multiple parts using a rectangular rolling window with duration $t_w < t_r$ and a stride of t_s. The simulated training data is a combination of a real, HE part and a simulated, faulty part. Because of this combination, it is necessary to know which records are actually HE when simulating the training dataset. This information is readily available in data from experiment 4, where faults are seeded artificially. In run-to-failure datasets 5 and 6, the first N_{HE} records are assumed to be HE. This assessment is done based on results from reference publication utilizing the datasets. In a real application, one would collect baseline records when the machine is new or recently overhauled.

Data preprocessing is slightly different depending on the dataset. First, HE data $\boldsymbol{x}_{HE}$ is augmented with white noise $\boldsymbol{w}$, and normalized to unit RMS. Healthy data to be used in simulation is drawn with replacement from the complete set of HE data. Then each record of simulated faulty data $\boldsymbol{x}_F$ is normalized with a factor $\lambda \operatorname{RMS}\{\boldsymbol{x}_F\}$, where λ is a random variable that controls the power ratio between HE and simulated data.

For dataset 4, the sum is denoted $\boldsymbol{x}'_S$, as shown in Eq. (3.15). $\boldsymbol{x}'_S$ is then normalized again to unit RMS, denoted $\boldsymbol{x}_S$ as in Eq. (3.16).

$$\boldsymbol{x}'_S = \frac{\boldsymbol{x}_{HE} + \boldsymbol{w}}{\operatorname{RMS}\{\boldsymbol{x}_{HE} + \boldsymbol{w}\}} + \frac{\boldsymbol{x}_F}{\lambda \operatorname{RMS}\{\boldsymbol{x}_F\}} \tag{3.15}$$

$$\boldsymbol{x}_S = \frac{\boldsymbol{x}'_S}{\operatorname{RMS}\{\boldsymbol{x}'_S\}} \tag{3.16}$$

The idea behind this normalization is to ensure that training and test data is on a similar scale. As the severity of seeded faults have a stepwise increase, it is hard to predict power of the real faulty signal. Normalizing all data to unit RMS ensures that the signal magnitude at test time is not too different from the simulated training data.

In run-to-failure tests, the situation is different. Here, the increase in RMS is assumed to be more gradual, and may be utilized by the network to perform classification. Again, $\boldsymbol{x}_{HE}$ is augmented with additive white noise w and normalized to the RMS of the original segment $\boldsymbol{x}_{HE}$. The simulated faulty data $\boldsymbol{x}_F$ is also normalized with the same factor $\lambda \operatorname{RMS}\{\boldsymbol{x}_F\}$. This time, the signals are simply added to obtain the simulated signal, as shown in Eq. (3.17).

$$\boldsymbol{x}_S = \frac{(\boldsymbol{x}_{HE} + \boldsymbol{w}) \operatorname{RMS}\{\boldsymbol{x}_{HE}\}}{\operatorname{RMS}\{\boldsymbol{x}_{HE} + \boldsymbol{w}\}} + \frac{\boldsymbol{x}_F}{\lambda \operatorname{RMS}\{\boldsymbol{x}_F\}} \tag{3.17}$$

While this model captures classical characteristics of a faulty bearing vibration signal, there are several unknown parameters. By treating parameters as random variables, sig-

nals with vastly different properties can be generated based on previous experience and less assumptions about system properties. When simulating a time series, distributions are sampled to obtain a set of model parameters. Fig. 3.9 shows the components of a simulated RE fault signal.

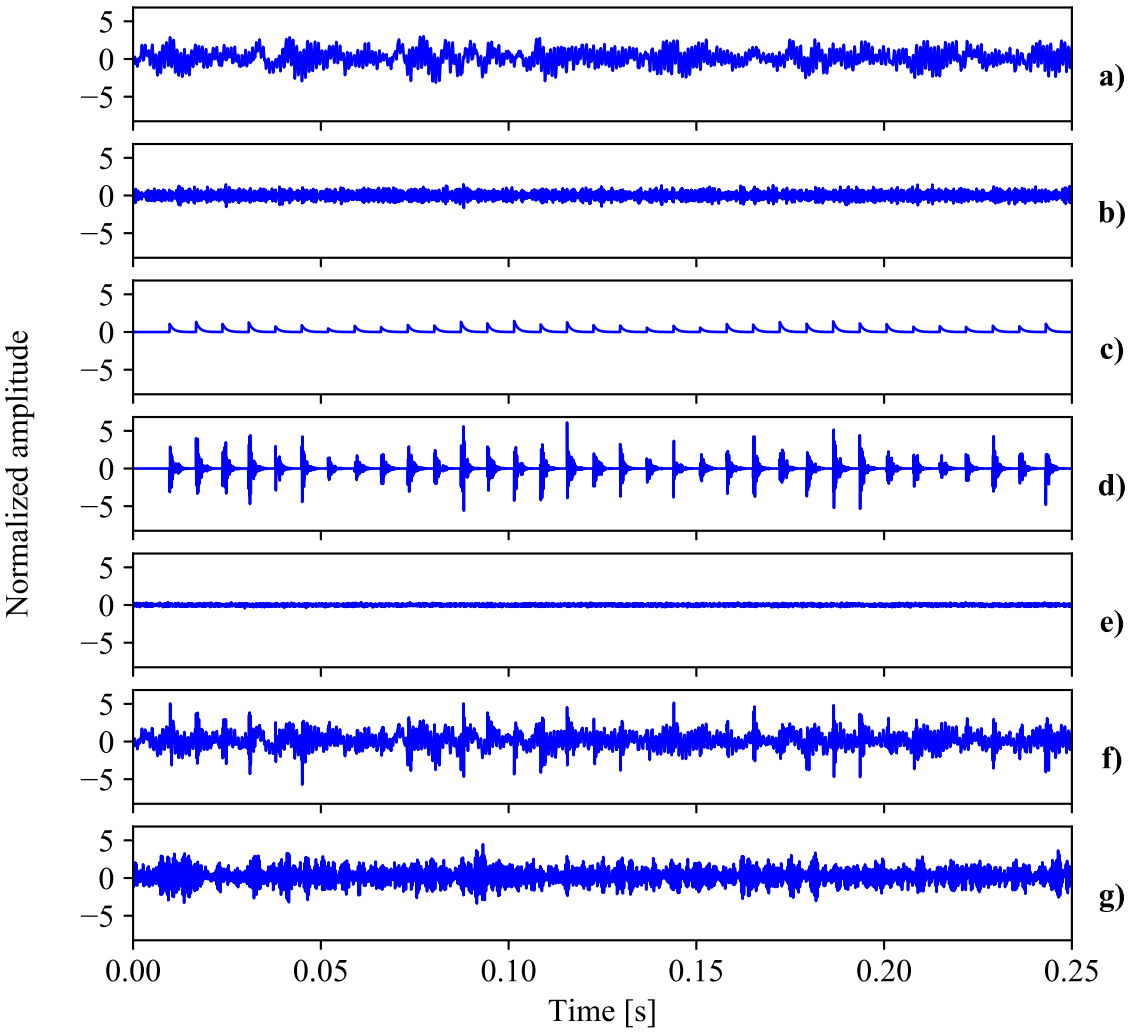


Figure 3.9: Components of generated RE fault compared to actual fault signal. **a)** Healthy signal $\boldsymbol{x}_{HE}$. **b)** Band-pass filtered noise $\boldsymbol{w}_{bp}$. **c)** Modulating envelope $m_1 + m_2$. **d)** Generated fault impacts $\boldsymbol{x}_p$. **e)** Additive noise $\boldsymbol{w}$. **f)** Generated fault signal $\boldsymbol{x}_s$. **g)** Actual fault signal.

The simulated vibration signal were used to train an ensemble of classifiers, each based on stacked blocks of 1D convolutions with ReLU activation and max pooling layers. Two

low-level feature extraction blocks are followed by eight blocks with increasing dilation rates. This approach allows the classifier to maintain a wide receptive field on high-resolution data with relatively few layers. The principle of dilated 1D convolutions is shown in Fig 3.10. Global average pooling is applied in the final layer before classification output. A complete overview of the network structure is given in Table 3.10. Separate classifiers were trained for each channel. The models were fitted to training data over 10 epochs with batch size 32, except for experiment 6 which had to reduce batch size to 16 due to memory constraints. The Adam optimizer with learning rate 10^{-4} was used in all cases.

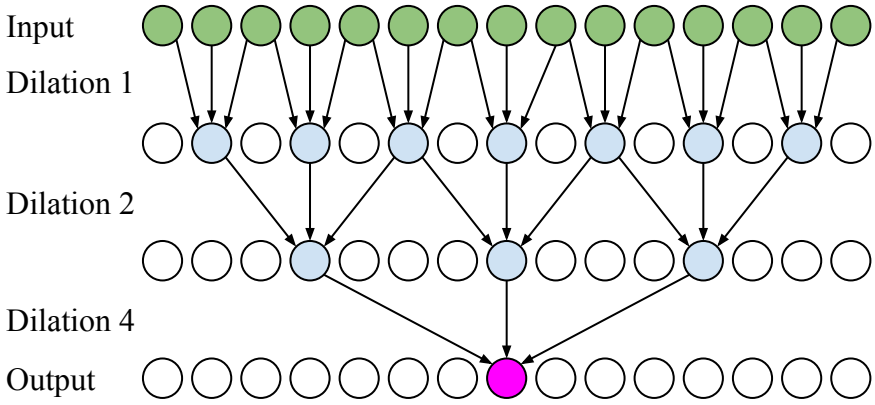


Figure 3.10: Dilated convolutional structure

The ability to classify single point fault types were evaluated by testing on high-resolution accelerometer data, sampled at 48 kHz, from experiment 4 in Section 2.4. As the dataset contains a single record of each fault type, this time series is split into 1 second segments with 0.02 s stride to obtain approximately 200 records per fault type for testing. The confusion matrix is shown in Fig. 3.11. The classifier successfully diagnose 100 % of the IR and OR faults. However, RE faults are misdiagnosed as IR damage as well. This fault type was not diagnosable in the baseline study, and was reported to not show typical characteristics of bearing failure [108]. As the simulated signal is designed to replicate a typical bearing fault, it is likely that none of the training data was representative of the actual RE fault. This illustrates a disadvantage of simulation-driven training of deep learning classifiers.

A similar approach was used on three run-to-failure tests from experiment 5 and 6. Vibration measurements sampled throughout the tests were either used directly (experiment 5) or split in shorter segments (experiment 6). In experiment 6, each classifier in the ensemble have an internal plurality vote over the separate windows in a record. This process is omitted with data from experiment 5, as the complete signal was used

Table 3.10: Network parameters

No.	Type	Filters	Size	Stride	Padding	Dilation	Activation
0	Input						
1	Conv1D	32	11	1	valid	1	ReLU
2	MaxPooling1D		3	2			
3	BatchNormalization						
4	Conv1D	32	5	1	valid	1	ReLU
5	MaxPooling1D		3	2			
6	BatchNormalization						
7	Conv1D	8	5	1	same	1	ReLU
8	Conv1D	8	5	1	same	2	ReLU
9	Conv1D	8	5	1	same	4	ReLU
10	Conv1D	8	5	1	same	8	ReLU
11	Conv1D	8	5	1	same	16	ReLU
12	Conv1D	8	5	1	same	32	ReLU
13	Conv1D	8	5	1	same	64	ReLU
14	Conv1D	8	5	1	valid	128	ReLU
15	GlobalAveragePooling1D						
16	Dense	4					Softmax

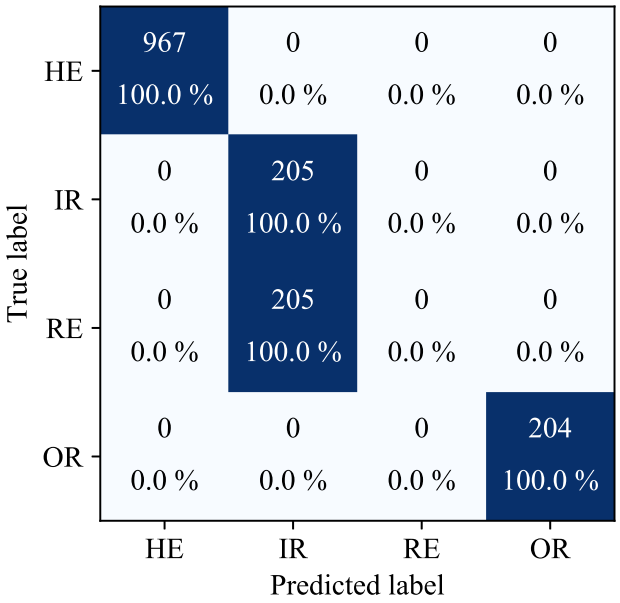


Figure 3.11: Dataset 4 confusion matrix

in training and testing. The ensemble then has another plurality vote to determine final diagnosis output. It is impossible to verify exactly when a fault occurs in the run to

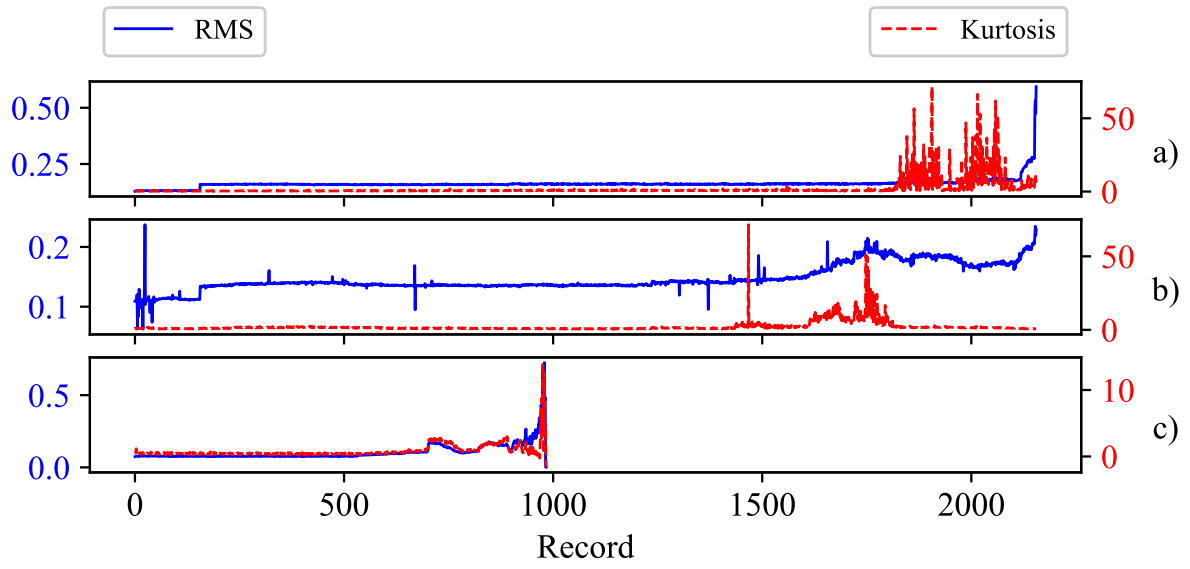


Figure 3.12: Plot of RMS and kurtosis for the faulty bearings in experiment 5. **a)** Test 1, bearing 3. **b)** Test 1, bearing 4. **c)** Test 2, bearing 1.

failure test. For reference, the RMS and kurtosis plots for experiment 5, shown in Fig. 3.12, are examined to estimate the damage initiation. This can not be used for diagnosis alone, but combined with the observed damage at the end of life, it gives an indication of when faults occur, and what fault type is expected. In experiment 5, test 1, kurtosis increase in bearing 4 around record 1435, followed by a larger increase at record 1610. In bearing 3, kurtosis increases from record 1800. In test 2, an increase starting from record 530 is observed in bearing 1.

Classifier output from test 1 until failure is shown in Fig. 3.13. Inspections after failure revealed IR damage in bearing 3 and RE damage in bearing 4. The classifier briefly diagnoses bearing 4 correctly as RE at record 1525, 90 records after an increase in kurtosis. Then, the classifier switches to IR diagnosis output at record 1554. Towards the end of life, the classifier alternates between IR and RE output, with the majority of records being misclassified as IR. The first fault detected in bearing 3 is misclassified as RE damage. Around record 2000, the classifier changes to the correct IR fault diagnosis, before switching to an OR diagnosis towards the end of life. Upon inspection, the IR was severely spalled, which can make the fault signal less impulsive. The simulated training data does not take larger distributed spalls into account, and may not diagnose it correctly. As bearing 4 appears to fail before bearing 3, it is also possible that vibration from the faulty bearings interfere, contributing to misclassification.

Figure 3.14 shows diagnosis output for test 2 in experiment 5. An OR fault in bearing 1 is indicated at record 530, which corresponds very well with the RMS and kurtosis increase shown in Fig. 3.12. The diagnosis is stable towards end of life, before briefly

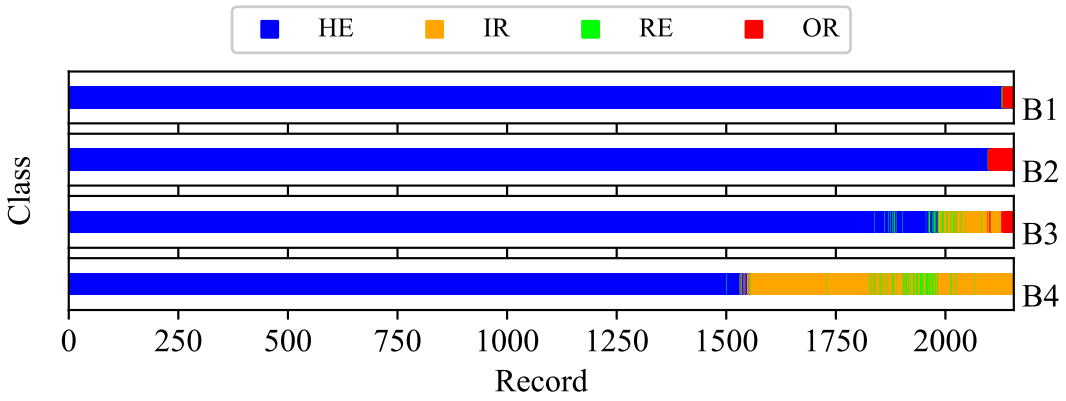


Figure 3.13: Classifier output for experiment 5, test 1. Classification by majority voting. At the end of testing, bearing B3 and B4 were found to have IR and RE damage respectively.

returning to HE output. At this point, the bearing is severely damaged, which may cause less transient impulses masked in high levels of noise. The other bearings are also classified as faulty at the end of testing, but were not actually damaged. Again, it is possible that vibration from the damaged bearing 1 is picked up by the sensors on the other bearings.

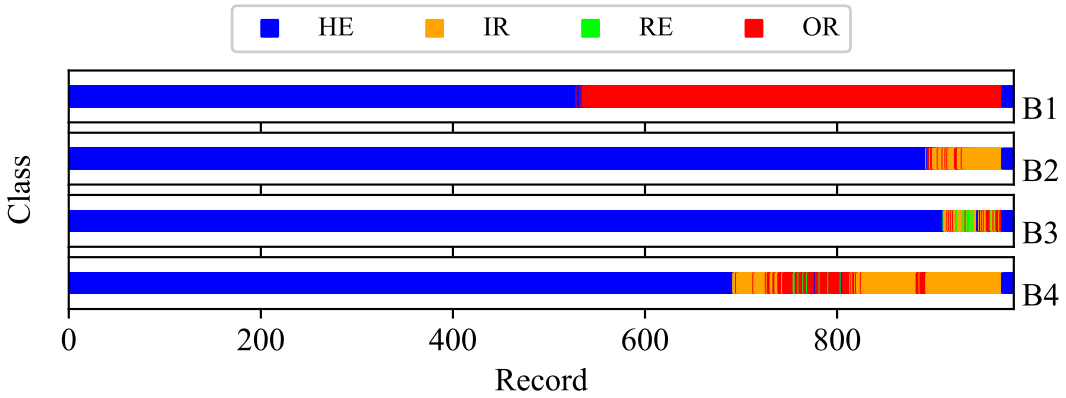


Figure 3.14: Classifier output for experiment 5, test 2. Classification by majority voting. At the end of testing, bearing B1 was found to have OR damage.

Experiment 6 contains a run-to-failure test at 250 rpm. In [103], the bearing was reported to have damage in IR, RE and OR at the end of life. The authors found that indications of RE damage were present from record 171. The proposed classifier output, as shown in Fig. 3.15, also indicate initial RE damage occurring at record 171. The diagnosis changes to HE, before returning to RE damage. As all fault types were present at the end of life, it is not possible to determine a correct initial diagnosis with absolute certainty. The presence of multiple faults is also a likely explanation for the HE classification after

first diagnosing a RE damage.

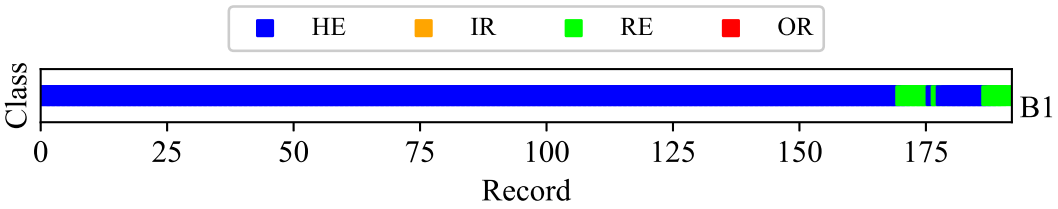


Figure 3.15: Classifier output for dataset 6. Classification by majority voting. At the end of testing, the bearing was found to have RE, OR and IR damage. RE damage is asusmed to occur first.

Simulation-driven classification shows good diagnostics performance on seeded single-point faults, but misclassifies new records if the observed damage differs too much from the simulated training data. The classifier is able to detect faults in more realistic run-to-failure, although the detection is later that the observed kurtosis increase in experiment 5, test 1. In this test, the classifier also tend to misclassify the faults. In the remaining experiments, faults are indicated at the same time or slightly later than in reference papers. It is assumed that multiple fault types and severe damage tend to cause misclassification.

Chapter 4

Concluding Remarks

4.1 Conclusions

This project focuses on developing methods for CM of large rolling element bearings operating at low speed, motivated by unplanned downtimes in offshore drilling rigs caused by top drive bearing failure. A review of CM methods for large, low-speed bearings is presented in Paper A, showing state-of-the-art CM methods under an assumption that faults result in the presence of fundamental frequencies of signals associated with the bearing faults. In Paper B, a used top drive with visible scratches distributed on the roller ends was tested in a workshop. Assuming such roller end wear would also result in harmonic components at the roller frequency, the machine was equipped with acceleration and AE measurement systems that recorded data at different speeds in the operating range. Despite thorough analysis, no fault-related harmonic component was observed in the test data. It was concluded that fault diagnosis methods that rely on the presence of fundamental fault frequency are not effective for diagnosis of such scratches. Other reliable CIs were not discovered, though AE RMS showed some increases with increasing DLs. However, the study provided a better understanding of the failure mode and limitations of state-of-the-art CM methods' ability to detect and diagnose it. In addition to roller end wear, CM methods for single-point surface damage are also considered throughout the study.

To address the lack of reliable CIs for roller end wear, the research focus was shifted towards ML algorithms for feature extraction and classification. A test rig for axial tapered roller bearings was then developed, as described in Paper C. It was specifically designed to imitate roller end scratches by means of a mechanism that allows scratch formation during operation. In Paper D, a pre-trained CNN was used to extract features for supervised binary classification of scratches as well as single-point faults based on spectrograms. This TL approach is proposed as a simple implementation of automated bearing fault

detection in case fault data is available, but manual analysis is still time-consuming and requires skilled personnel. Several variations of feature extraction and classification were tested, and a support vector machine trained on directly extracted features was judged to have the best balance of accuracy, robustness, ease of implementation and computational burden.

Like all supervised classification, TL also requires labeled training data representative for all problem classes. In this study, the test rig was used to obtain training data, but it is acknowledged that historic fault data is often not available in industrial applications. Papers E and F propose two different approaches to tackle this issue. In Paper E, deviation from HE operational characteristics is quantified by using variational autoencoders to infer latent Gaussian variables from the observed data. This approach allows setting statistical thresholds on arbitrarily distributed features. In critical equipment such as top drives, the ability to control a false alarm rate is a valuable tool to avoid unnecessary downtime for maintenance. The method was applied to both data from the axial bearing test rig and top drive test data. In both cases, the HI gave higher readings, including warnings and alarms, at increasing levels of damage. The methodology is general, and can be applied to any kind of data, but it was demonstrated that it is also efficient for bearing health estimation.

Paper F considers single-point defects, being a different scenario than roller end wear. The fault characteristics are relatively well known, and several models exist to simulate a defect bearing using vibration signals. The scope of the proposed method is twofold. The first part considers obtaining sufficiently realistic fault signals from a model with unknown or uncertain parameters. This is done by treating model parameters as distributions that are sampled for signal generation. The second part is the use of 1-D dilated convolutions to extract features from raw, high-resolution data with less layers and smaller filter size. Unlike the spectrograms used in Paper D, the proposed network structure is better suited to capture features on fine and coarse timescales simultaneously. By making the model parameter distributions wide enough, the network was able to learn more general features that successfully classified seeded OR and IR faults in unseen time series, while RE damage was misdiagnosed as IR. In run-to-failure tests at both high and low speed, classification performance was more unstable with regards to fault type. However, if diagnosis is disregarded and instead treating all fault types as a single faulty class, the detection capability was comparable to reference methods from the literature.

Overall, the project has resulted in methods for feature extraction, fault detection and diagnostics of bearings. The proposed methods have been tested on data from high-speed and low-speed bearings. The unsupervised HI was shown effective on actual top drive vibration signals, while the remaining methods were tested on data from in-house test

rigs and publicly available datasets.

4.2 Limitations

The TL approach presented in Section 3.2.1 requires that the new data can be represented in the same format as the training data. In Paper D, the network was initially trained on RGB images sized 227x227x3. As vibration and AE time series are 1D, data was split in segments, transformed to time-frequency representations (spectrograms) and resized to the same input size. This clearly limits the time and frequency resolution of the spectrograms. Only a binary classification was done, so the diagnosis performance is not evaluated. The method also requires that representative training data is available, which may be hard to obtain in a production environment.

The VAE and CVAE approach to HI calculation in Paper E reduces the requirements on feature engineering and selection, but some of the features have to actually carry information that can be used to indicate a machine health. A change can occur over time but still not be an indication of degrading health. The AR model coefficients were capable of capturing changes in time series characteristics, but the HI calculation does not provide a direct physical interpretation of the change. With VAEs, a separate classifier must be trained for each operating condition. This is somewhat remedied by the conditioning in CVAEs, but one-hot encoding of the conditioning vector limits the capability of classifying unseen operating conditions.

The classifier trained on simulated fault signals was in general capable of fault detection. There was no missed detection of seeded single-point defects, i.e. damaged data classified as HE. However, RE faults in the test dataset did not exhibit the typical fault signature as expected, thus they were consistently misclassified as IR faults. This also reflects on classifier performance in more realistic run-to-failure experiments. Tests with multiple faults present simultaneously and with faulty bearings nearby, reduced classifier diagnosis performance. These results emphasize that the simulation-driven classification is limited by the model's ability to mimic the actual fault signal. In all cases, the proposed methods were tested on data from test rigs or in the top drive case, without using an axial load. Thus, the methodologies have not been validated in a production environment.

4.3 Further Work

The effectiveness of applying an image classification network to a bearing fault detection task, shows that knowledge can be utilized across domains despite some limitations. In the experiments, only a binary classification was tested. However, the ability to separate

different faulty classes should also be investigated. Also, instead of transferring knowledge from very separate domains, the TL methodology can be used on a network trained on large amounts of labeled data from different machines. With enough data, the network could be able to separate "style" from "content", i.e. "machine specific signature" from "fault information". In that sense, knowledge of faults in some machines is transferred to classification of new datasets. Developing such a pretrained network on bearing vibration data could have large industrial benefits.

In the VAE, one network was trained for each operating speed. This is time consuming and might reduce generality. The CVAE approach is an improvement, which allowed a single network to be conditioned on a finite set of operating speeds. A further improvement would be to replace one-hot encoding of continuous variables like speed and load with a floating point number scaled from 0 to 1. This can also be seen as a form of TL, where artificial faulty signals can be generated for a specific machine condition based on similar experiences from other machines. This could be used to train a machine-specific classifier as in Paper F. Other formats for input data and encoder/decoder structure should be investigated for use in VAEs, such as the 1-D CNN and recurrent layers for input sequences.

The proposed methods are developed and tested on fault detection and diagnosis problems. Using features from TL for a HI should be considered for RUL estimation. It should also be a priority to validate the proposed methods on actual data from industrial equipment in operation.

References

[1] International Organization for Standardization (ISO). *ISO 281:2007 - Rolling Bearings — Dynamic Load Ratings and Rating Life*. 2007.

[2] International Organization for Standardization (ISO). *ISO 15243:2017 - Rolling Bearings - Damage and Failures - Terms, Characteristics and Causes*. 1 edition, 2004.

[3] R. B. Randall and J. Antoni. Rolling Element Bearing Diagnostics - A Tutorial. *Mechanical Systems and Signal Processing*, 25(2):485–520, 2011. doi:10.1016/j.ymssp.2010.07.017.

[4] E. Y. Kim, A. C. C. Tan, J. Mathew, and B. S. Yang. Condition Monitoring of Low Speed Bearings: A Comparative Study of the Ultrasound Technique Versus Vibration Measurements. In *Proceedings of the 1st World Congress on Engineering Asset Management (WCEAM)*, volume 5, pages 177–189, 2008. doi:10.1007/978-1-84628-814-2_21.

[5] Z. Stamboliska, E. Rusiński, and P. Moczko. *Proactive Condition Monitoring of Low-speed Machines*. Springer International Publishing, 1 edition, 2014. doi:10.1007/978-3-319-10494-2.

[6] B. V. Hecke, J. Yoon, and D. He. Low Speed Bearing Fault Diagnosis Using Acoustic Emission Sensors. *Applied Acoustics*, 105:35–44, 2016. doi:10.1016/j.apacoust.2015.10.028.

[7] E. Bechhoefer, R. Schlanbusch, and T. I. Waag. Techniques for Large, Slow Bearing Fault Detection. *International Journal of Prognostics and Health Management*, 7(1):1–12, 2016.

[8] R. J. Parker. Large-bore Tapered-roller Bearing Performance and Endurance to 2.4 Million Dn Apparatus and Procedure, 1983. Obtained date: 19.02.2018. URL: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19830011863.pdf.

[9] S. M. Muzakkir, H. Hirani, and G. D. Thakre. Lubricant for Heavily Loaded Slow-speed Journal Bearing. *Tribology Transactions*, 56(6):1060–1068, 2013. doi:10.1080/10402004.2013.823530.

[10] A. Rognerud. *Drilling Rig Maintenance: An Analytical Study of the Classification, Treatment, Data Quality of Equipment Failures and Related Downtime*. PhD thesis, University of Stavanger, 2011.

[11] L. Jeffrey. Noble 2012 Analyst & Investor Day Presentation, 2012. Presented: 24.05.2012, Houston, Texas, USA. Date obtained: 31.08.2016.

[12] N. Tandon and A. Choudhury. A Review of Vibration and Acoustic Measurement Methods for the Detection of Defects in Rolling Element Bearings. *Tribology International*, 32(8):469–480, 1999. doi:https://doi.org/10.1016/s0301-679x(99)00077-8.

[13] P. Shakya, A. K. Darpe, and M. S. Kulkarni. Bearing Diagnosis Using Proximity Probe and Accelerometer. *Measurement: Journal of the International Measurement Confederation*, 80:190–200, 2016. doi:10.1016/j.measurement.2015.11.029.

[14] N. Jamaludin and D. Mba. Monitoring Extremely Slow Rolling Element Bearings: Part I. *NDT and E International*, 35(6):349–358, 2002. doi:10.1016/S0963-8695(02)00006-3.

[15] T. Yoshioka. Detection of Rolling Contact Sub-surface Fatigue Cracks using Acoustic Emission Technique. *Lubrication Engineering*, 04(1), 1993.

[16] C. Jiaa and D. Dornfeld. Experimental Studies of Sliding Friction and Wear via Acoustic Emission Signal Analysis. *Wear*, 139(2):403–424, 1990. doi:10.1016/0043-1648(90)90059-J.

[17] T. Yoshioka and T. Fujiwara. A New Acoustic Emission Source Locating System for the Study of Rolling Contact Fatigue. *Wear*, 81(1):183–186, 1982. doi:10.1016/0043-1648(82)90314-3.

[18] T. Yoshioka and T. Fujiwara. Application of Acoustic Emission Technique to Detection of Rolling Bearing Failure. *American society of mechanical engineers*, 14(1):55–76, 1984.

[19] N. Jamaludin and D. Mba. Monitoring Extremely Slow Rolling Element Bearings: Part II. *NDT and E International*, 35(6):359–366, 2002. doi:10.1016/S0963-8695(02)00006-3.

[20] W. Caesarendra, B. Kosasih, A. K. Tieu, H. Zhu, C. A. S. Moodie, and Q. Zhu. Acoustic Emission-based Condition Monitoring Methods: Review and Application for Low Speed Slew Bearing. *Mechanical Systems and Signal Processing*, 72-73:134–159, 2016. doi:10.1016/J.YMSSP.2015.10.020.

[21] S. A. Aye and P. S. Heyns. An Integrated Gaussian Process Regression for Prediction of Remaining Useful Life of Slow Speed Bearings Based on Acoustic Emission. *Mechanical Systems and Signal Processing*, 84:485–498, 2016. doi:10.1016/j.ymssp.2016.07.039.

[22] H. N. Nguyen, J.-M. J. Kim, J.-M. J. Kim, H. N. Nguyen, J.-M. J. Kim, and J.-M. J. Kim. Optimal Sub-band Analysis Based on the Envelope Power Spectrum for Effective Fault Detection in Bearing Under Variable, Low Speeds. *Sensors*, 18(5):1389, 2018. doi:10.3390/s18051389.

[23] Z. Mo, J. Wang, H. Zhang, X. Zeng, H. Liu, and Q. Miao. Vibration and Acoustics Emission Based Methods in Low-speed Bearing Condition Monitoring. In *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pages 871–875, 2018. doi:10.1109/PHM-Chongqing.2018.00156.

[24] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim. Low-speed Bearing Fault Diagnosis Based on ArSSAE Model Using Acoustic Emission and Vibration Signals. *IEEE Access*, 7:46885–46897, 2019. doi:10.1109/ACCESS.2019.2909756.

[25] S. Poddar and N. Tandon. Detection of Particle Contamination in Journal Bearing using Acoustic Emission and Vibration Monitoring Techniques. *Tribology International*, 134:154–164, 2019. doi:10.1016/J.TRIBOINT.2019.01.050.

[26] R. M. Douglas, J. A. Steel, and R. L. Reuben. A Study of the Tribological Behaviour of Piston Ring/cylinder Liner Interaction in Diesel Engines using Acoustic Emission. *Tribology International*, 39(12):1634–1642, 2006. doi:10.1016/j.triboint.2006.01.005.

[27] H. Taura and K. Nakayama. Behavior of Acoustic Emissions at the Onset of Sliding Friction. *Tribiology International*, 123:155–160, 2018. doi:10.1016/j.triboint.2018.01.025.

[28] H. Ocak, K. A. Loparo, and F. M. Discenzo. Online Tracking of Bearing Wear Using Wavelet Packet Decomposition and Probabilistic Modeling: A Method for Bearing Prognostics. *Journal of Sound and Vibration*, 302(4-5):951–961, 2007. doi:10.1016/j.jsv.2007.01.001.

[29] J. Zhu, J. M. Yoon, D. He, and E. Bechhoefer. Online Particle-contaminated Lubrication Oil Condition Monitoring and Remaining Useful Life Prediction for Wind Turbines. *Wind Energy*, 18(6):1131–1149, 2015. doi:10.1002/we.1746.

[30] X. Zhu, C. Zhong, and J. Zhe. Lubricating Oil Conditioning Sensors for Online Machine Health Monitoring – A Review. *Tribology International*, 109:473–484, 2017. doi:10.1016/J.TRIBOINT.2017.01.015.

[31] International Organization for Standardization (ISO). *ISO 10816-3:2009, Mechanical Vibration - Evaluation of Machine Vibration by Measurements on Non-rotating Parts*. 2009.

[32] P. McFadden and J. Smith. Model for the Vibration Produced by a Single Point Defect in a Rolling Element Bearing. *Journal of Sound and Vibration*, 96(1):69–82, 1984. doi:10.1016/0022-460X(84)90595-9.

[33] L. Marple. Computing the Discrete-time "Analytic" Signal via FFT. *IEEE Transactions on Signal Processing*, 47(9):2600–2603, 1999. doi:10.1109/78.782222.

[34] J. Antoni. The Spectral Kurtosis: A Useful Tool for Characterising Non-stationary Signals. *Mechanical Systems and Signal Processing*, 20(2):282–307, 2006. doi:10.1016/j.ymssp.2004.09.001.

[35] J. Antoni. Fast Computation of the Kurtogram for the Detection of Transient Faults. *Mechanical Systems and Signal Processing*, 21(1):108–124, 2007. doi:10.1016/j.ymssp.2005.12.002.

[36] A. Moshrefzadeh and A. Fasana. The Autogram: An Effective Approach for Selecting the Optimal Demodulation Band in Rolling Element Bearings Diagnosis. *Mechanical Systems and Signal Processing*, 105:294–318, 2018. doi:10.1016/J.YMSSP.2017.12.009.

[37] Y. Xu, K. Zhang, C. Ma, L. Cui, and W. Tian. Adaptive Kurtogram and Its Applications in Rolling Bearing Fault Diagnosis. *Mechanical Systems and Signal Processing*, 130:87–107, 2019. doi:10.1016/J.YMSSP.2019.05.003.

[38] M. Žvokelj, S. Zupan, and I. Prebil. Multivariate and Multiscale Monitoring of Large-size Low-speed Bearings Using Ensemble Empirical Mode Decomposition Method Combined With Principal Component Analysis. *Mechanical Systems and Signal Processing*, 24(4):1049–1067, 2010. doi:10.1016/j.ymssp.2009.09.002.

[39] Q. Xiong, Y. Xu, Y. Peng, W. Zhang, Y. Li, and L. Tang. Low-speed Rolling Bearing Fault Diagnosis Based on EMD Denoising and Parameter Estimate With Alpha Stable Distribution. *Journal of Mechanical Science and Technology*, 31(4):1587–1601, 2017. doi:10.1007/s12206-017-0306-y.

[40] T. Han, Q. Liu, L. Zhang, and A. C. Tan. Fault Feature Extraction of Low Speed Roller Bearing Based on Teager Energy Operator and Ceemd. *Measurement*, 138:400–408, 2019. doi:10.1016/J.MEASUREMENT.2019.02.053.

[41] C. Mishra, A. Samantaray, and G. Chakraborty. Rolling Element Bearing Fault Diagnosis Under Slow Speed Operation using Wavelet De-noising. *Measurement*, 103:77–86, 2017. doi:10.1016/J.MEASUREMENT.2017.02.033.

[42] H. Nguyen, J. Kim, J.-M. Kim, H. N. Nguyen, J. Kim, and J.-M. Kim. Optimal Subband Analysis Based on the Envelope Power Spectrum for Effective Fault Detection in Bearing Under Variable, Low Speeds. *Sensors*, 18(5):1389, 2018. doi:10.3390/s18051389.

[43] J. Antoni and R. B. Randall. Differential Diagnosis of Gear and Bearing Faults. *Journal of Vibration and Acoustics*, 124(2):165–171, 2002. doi:10.1115/1.1456906.

[44] D. Abboud, J. Antoni, S. Sieg-Zieba, and M. Eltabach. Envelope Analysis of Rotating Machine Vibrations in Variable Speed Conditions: A Comprehensive Treatment. *Mechanical Systems and Signal Processing*, 84:200–226, 2017. doi:10.1016/j.ymssp.2016.06.033.

[45] S. Kass, A. Raad, and J. Antoni. Self-running Bearing Diagnosis Based on Scalar Indicator Using Fast Order Frequency Spectral Coherence. *Measurement*, 138:467–484, 2019. doi:10.1016/J.MEASUREMENT.2019.02.046.

[46] J. Antoni. Cyclostationarity by Examples. *Mechanical Systems and Signal Processing*, 23(4):987–1036, 2009. doi:10.1016/j.ymssp.2008.10.010.

[47] D. Abboud, M. Elbadaoui, W. Smith, and R. Randall. Advanced Bearing Diagnostics: A Comparative Study of Two Powerful Approaches. *Mechanical Systems and Signal Processing*, 114:604–627, 2019. doi:10.1016/j.ymssp.2018.05.011.

[48] E. Bechhoefer and M. Kingsley. A Review of Time Synchronous Average Algorithms. In *Annual Conference of the Prognostics and Health Management Society*, pages 24–33, 2009.

[49] R. B. Randall and N. Sawalhi. Use of the Cepstrum to Remove Selected Discrete Frequency Components from a Time Signal. In *Proc. Int. Conference on Noise and Vibration Engineering (ISMA)*, pages 451–461. Springer, New York, NY, 2011. doi:10.1007/978-1-4419-9428-8_38.

[50] P. Borghesani, P. Pennacchi, R. B. Randall, N. Sawalhi, and R. Ricci. Application of Cepstrum Pre-whitening for the Diagnosis of Bearing Faults Under Variable Speed Conditions. *Mechanical Systems and Signal Processing*, 36(2):370–384, 2013. doi:10.1016/j.ymssp.2012.11.001.

[51] R. B. Randall, N. Sawalhi, and M. Coats. A Comparison of Methods for Separation of Deterministic and Random Signals. *The International Journal of Condition Monitoring*, 1(1):11–19, 2011. doi:https://doi.org/10.1784/204764211798089048.

[52] K. Fyfe and E. Munck. Analysis of Computed Order Tracking. *Mechanical Systems and Signal Processing*, 11(2):187–205, 1997. doi:10.1006/MSSP.1996.0056.

[53] Y. Pan, R. Hong, J. Chen, Z. Qin, and Y. Feng. Incipient Fault Detection of Wind Turbine Large-size Slewing Bearing Based on Circular Domain. *Measurement*, 137:130–142, 2019. doi:10.1016/J.MEASUREMENT.2019.01.033.

[54] C. Peeters, Q. Leclère, J. Antoni, P. Lindahl, J. Donnal, S. Leeb, and J. Helsen. Review and Comparison of Tacholess Instantaneous Speed Estimation Methods on Experimental Vibration Data. *Mechanical Systems and Signal Processing*, 129:407–436, 2019. doi:10.1016/J.YMSSP.2019.02.031.

[55] D. Jung and C. Sundström. A Combined Data-driven and Model-based Residual Selection Algorithm for Fault Detection and Isolation. *IEEE Transactions on Control Systems Technology*, 27(2):616–630, 2019. doi:https://doi.org/10.1109/TCST.2017.2773514.

[56] X. Dai and Z. Gao. From Model, Signal to Knowledge: A Data-driven Perspective of Fault Detection and Diagnosis. *IEEE Transactions on Industrial Informatics*, 9(4):2226–2238, 2013. doi:https://doi.org/10.1109/tii.2013.2243743.

[57] Z. Gao, C. Cecati, and S. X. Ding. A Survey of Fault Diagnosis and Fault-tolerant Techniques-part I: Fault Diagnosis With Knowledge-based and Hybrid/active Approaches. *Transactions on Industrial Informatics*, 62(6):3757–3767, 2015. doi:https://doi.org/10.1109/tie.2015.2419013.

[58] G. Niu. *Data-driven Technology for Engineering Systems Health Management: Design Approach, Feature Construction, Fault Diagnosis, Prognosis, Fusion and Decision*. Springer, Singapore, 2016. doi:https://doi.org/10.1007/978-981-10-2032-2.

[59] A. Diez-Olivan, J. Del Ser, D. Galar, and B. Sierra. Data Fusion and Machine Learning for Industrial Prognosis: Trends and Perspectives Towards Industry 4.0. *Information Fusion*, 50:92–111, 2019. doi:10.1016/J.INFFUS.2018.10.005.

[60] P. Konar and P. Chattopadhyay. Bearing Fault Detection of Induction Motor using Wavelet and Support Vector Machines (SVMs). *Applied Soft Computing*, 11(6):4203–4211, 2011. doi:https://doi.org/10.1016/j.asoc.2011.03.014.

[61] M. Kang, J. Kim, J.-M. Kim, A. C. C. Tan, E. Y. Kim, and B.-K. Choi. Reliable Fault Diagnosis for Low-speed Bearings Using Individually Trained Support Vector Machines With Kernel Discriminative Feature Analysis. *IEEE Transactions on Power Electronics*, 30(5):2786–2797, 2015. doi:10.1109/TPEL.2014.2358494.

[62] A. Soualhi, K. Medjaher, and N. Zerhouni. Bearing Health Monitoring Based on Hilbert–huang Transform, Support Vector Machine, and Regression. *IEEE Transactions on Instrumentation and Measurement*, 64(1):52–62, 2015. doi:10.1109/TIM.2014.2330494.

[63] J. S. L. Senanayaka, S. T. Kandukuri, H. V. Khang, and K. G. Robbersmyr. Early Detection and Classification of Bearing Faults using Support Vector Machine Algorithm. In *2017 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD)*, pages 250–255, 2017. doi:10.1109/WEMDCD.2017.7947755.

[64] X. Li, Y. Yang, H. Pan, J. Cheng, and J. Cheng. A Novel Deep Stacking Least Squares Support Vector Machine for Rolling Bearing Fault Diagnosis. *Computers in Industry*, 110:36–47, 2019. doi:10.1016/J.COMPIND.2019.05.005.

[65] J. S. L. Senanayaka, H. V. Khang, and K. G. Robbersmyr. Towards Online Bearing Fault Detection using Envelope Analysis of Vibration Signal and Decision Tree Classification Algorithm. In *IEEE International Conference on Electrical Machines and Systems (ICEMS)*, pages 1–6, 2017. doi:https://doi.org/10.1109/icems.2017.8056146.

[66] M. Saimurugan, K. Ramachandran, V. Sugumaran, and N. Sakthivel. Multi Component Fault Diagnosis of Rotational Mechanical System Based on Decision Tree and Support Vector Machine. *Expert Systems with Applications*, 38(4):3819–3826, 2011. doi:10.1016/J.ESWA.2010.09.042.

[67] L. Song, H. Wang, and P. Chen. Vibration-based Intelligent Fault Diagnosis for Roller Bearings in Low-speed Rotating Machinery. *IEEE Transactions on Instrumentation and Measurement*, 67(8):1887–1899, 2018. doi:10.1109/TIM.2018.2806984.

[68] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht. Motor Bearing Fault Detection Using Spectral Kurtosis-based Feature Extraction Coupled With K-nearest Neighbor Distance Analysis. *IEEE Transactions on Industrial Electronics*, 63(3):1793–1803, 2016. doi:10.1109/TIE.2015.2509913.

[69] P. Nguyen, M. Kang, J. Kim, and J.-M. M. Kim. Reliable Fault Diagnosis of Low-speed Bearing Defects Using a Genetic Algorithm. In *PRICAI 2014: Trends in Artificial Intelligence 2014*, pages 248–255. Springer, Cham, 2014. doi:10.1007/978-3-319-13560-1.

[70] S. J. Pan and Q. Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010. doi:10.1109/TKDE.2009.191.

[71] M. Gan, C. Wang, and C. C. Zhu. Multiple-domain Manifold for Feature Extraction in Machinery Fault Diagnosis. *Measurement*, 75:76–91, 2015. doi:10.1016/j.measurement.2015.07.042.

[72] R. Zhang, H. Tao, L. Wu, Y. Guan, L. W. R. Zhang H. Tao, Y. Guan, R. Zhang, H. Tao, L. Wu, and Y. Guan. Transfer Learning With Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions. *IEEE Access*, 5:14347–14357, 2017. doi:10.1109/ACCESS.2017.2720965.

[73] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li. Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines With Unlabeled Data. *IEEE Transactions on Industrial Electronics*, 66(9):7316–7325, 2019. doi:10.1109/TIE.2018.2877090.

[74] X. Li, Z. Liu, Y. Qu, and D. He. Unsupervised Gear Fault Diagnosis Using Raw Vibration Signal Based on Deep Learning. In *2018 Prognostics and System Health Management Conference*, pages 1025–1030. IEEE, 2018. doi:10.1109/PHM-Chongqing.2018.00182.

[75] X. Guo, L. Chen, and C. Shen. Hierarchical Adaptive Deep Convolution Neural Network and Its Application to Bearing Fault Diagnosis. *Measurement*, 93:490–502, 2016. doi:10.1016/J.MEASUREMENT.2016.07.054.

[76] G. Li, C. Deng, J. Wu, X. Xu, X. Shao, Y. Wang, G. Li, C. Deng, J. Wu, X. Xu, X. Shao, and Y. Wang. Sensor Data-driven Bearing Fault Diagnosis Based on Deep Convolutional Neural Networks and S-transform. *Sensors*, 19(12):2750, 2019. doi:10.3390/s19122750.

[77] G. Jiang, H. He, J. Yan, and P. Xie. Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox. *IEEE Transactions on Industrial Electronics*, 66(4):3196–3207, 2019. doi:10.1109/TIE.2018.2844805.

[78] D. Peng, Z. Liu, H. Wang, Y. Qin, and L. Jia. A Novel Deeper One-dimensional CNN With Residual Learning for Fault Diagnosis of Wheelset Bearings in High-speed Trains. *IEEE Access*, 7:10278–10293, 2019. doi:10.1109/ACCESS.2018.2888842.

[79] M. M. Islam and J.-M. Kim. Automated Bearing Fault Diagnosis Scheme Using 2d Representation of Wavelet Packet Transform and Deep Convolutional Neural Network. *Computers in Industry*, 106:142–153, 2019. doi:10.1016/J.COMPIND.2019.01.008.

[80] Z. Chen and W. Li. Multisensor Feature Fusion for Bearing Fault Diagnosis using Sparse Autoencoder and Deep Belief Network. *IEEE Transactions on Instrumentation and Measurement*, 66(7), 2017. doi:10.1109/TIM.2017.2669947.

[81] M. Gan, C. Wang, and an Zhu. Construction of Hierarchical Diagnosis Network Based on Deep Learning and Its Application in the Fault Pattern Recognition of Rolling Element Bearings. *Mechanical Systems and Signal Processing*, 72-73:92–104, 2015. doi:10.1016/j.ymssp.2015.11.014.

[82] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu. Rolling Bearing Fault Feature Learning using Improved Convolutional Deep Belief Network With Compressed Sensing. *Mechanical Systems and Signal Processing*, 100:743–765, 2018. doi:10.1016/J.YMSSP.2017.08.002.

[83] J. Xie, G. Du, C. Shen, N. Chen, L. Chen, and Z. Zhu. An End-to-end Model Based on Improved Adaptive Deep Belief Network and Its Application to Bearing Fault Diagnosis. *IEEE Access*, 6:63584 – 63596, 2018. doi:10.1109/ACCESS.2018.2877447.

[84] M. He and D. He. Deep Learning Based Approach for Bearing Fault Diagnosis. *IEEE Transactions on Industry Applications*, 53(3):3057–3065, 2017. doi:10.1109/TIA.2017.2661250.

[85] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing. A Neural Network Constructed by Deep Learning Technique and Its Application to Intelligent Fault Diagnosis of Machines. *Neurocomputing*, 272:619–628, 2018. doi:10.1016/J.NEUCOM.2017.07.032.

[86] L. Wu, B. Yao, Z. . Peng, and Y. Guan. Fault Diagnosis of Roller Bearings Based on a Wavelet Neural Network and Manifold Learning. *Appl. Sci.*, pages 1–10, 2017.

[87] R. Liu, B. Yang, E. Zio, and X. Chen. Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review. *Mechanical Systems and Signal Processing*, 108:33–47, 2018. doi:10.1016/J.YMSSP.2018.02.016.

[88] M. A. Khan, Y.-H. Kim, and J. Choo. Intelligent Fault Detection via Dilated Convolutional Neural Networks. In *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 729–731. IEEE, 2018. doi:10.1109/BigComp.2018.00137.

[89] H. Sak, A. Senior, and F. Beaufays. Long Short-term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pages 338–342, 2014.

[90] D. Bahdanau, K. Cho, and Y. Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. *arXiv preprint*. arXiv:1409.0473v7.

[91] A. Vaswani, G. Brain, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention Is All You Need. *arXiv preprint*, 2017. arXiv:1706.03762v5.

[92] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A Generative Model for Raw Audio. *arXiv preprint*, 2016. arXiv:1609.03499v2.

[93] E. Bechhoefer and A. P. F. Bernhard. A Generalized Process for Optimal Threshold Setting in Hums. In *IEEE Aerospace Conference Proceedings*, pages 1–9, 2007. doi:10.1109/AERO.2007.352867.

[94] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. arXiv:arXiv:1406.2661v1.

[95] D. P. Kingma and M. Welling. Auto-encoding Variational Bayes. *arXiv preprint*, 2014. arXiv:1312.6114v10.

[96] A. Makhzani, J. Shlens, N. Jaitly, G. Brain, I. G. Openai, and B. Frey. Adversarial Autoencoders. *arXiv preprint*, 2016. arXiv:1511.05644v2.

[97] G. San Martin, E. López Droguett, V. Meruane, and M. das Chagas Moura. Deep Variational Auto-encoders: A Promising Tool for Dimensionality Reduction and Ball Bearing Elements Fault Diagnosis. *Structural Health Monitoring*, 2018. doi:10.1177/1475921718788299.

[98] MHWirth. Derrick Drilling Machine Catalogue, 2019. Date obtained: 03.09.2019. URL: https://www.mhwirth.com/wp-content/uploads/DDM-Series.pdf.

[99] Y. Qu, E. Bechhoefer, D. He, and J. Zhu. A New Acoustic Emission Sensor Based Gear Fault Detection Approach. *International Journal of Prognostics and Health Management*, 4:32–45, 2013.

[100] Case Western Reserve University Bearing Data Center Website. Date obtained: 19.12.2019. URL: http://csegroups.case.edu/bearingdatacenter/home.

[101] Nasa Ames Prognostics Data Repository. Date obtained: 21.08.2018. URL: http://data-acoustics.com/?p=299.

[102] H. Qiu, J. Lee, J. Lin, and G. Yu. Wavelet Filter-based Weak Signature Detection Method and Its Application on Rolling Element Bearing Prognostics. *JOURNAL OF SOUND AND VIBRATION*, 289:1066–1090, 2006. doi:10.1016/j.jsv.2005.03.007.

[103] A. Klausen, R. W. Folgerø, K. G. Robbersmyr, and H. R. Karimi. Accelerated Bearing Life-time Test Rig Development for Low Speed Data Acquisition. *Identification and Control*, 38(3):143–156, 2017. doi:10.4173/mic.2017.3.4.

[104] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet Classification With Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[105] J. M. Lilly and S. C. Olhede. Generalized Morse Wavelets as a Superfamily of Analytic Wavelets. *IEEE Transactions on Signal Processing*, 60(11):6036–6041, 2012. doi:10.1109/TSP.2012.2210890.

[106] E. Bechhoefer, D. He, and P. Dempsey. Gear Health Threshold Setting Based On a Probability of False Alarm. In J. R. Celaya, S. Saha, and A. Saxena, editors, *Annual Conference of the Prognostics and Health Management Society*, pages 275–281, 2011.

[107] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time Series Analysis : Forecasting and Control.* John Wiley & Sons, Inc., 5 edition, 2015.

[108] W. A. Smith and R. B. Randall. Rolling Element Bearing Diagnostics using the Case Western Reserve University Data: A Benchmark Study. *Mechanical Systems and Signal Processing*, 64-65:100–131, 2015. doi:10.1016/j.ymssp.2015.04.021.

Appendices

Paper A

A Review of Methods for Condition Monitoring of Large, Slow-rotating Bearings

Martin Hemmer, Tor I. Waag and Kjell G. Robbersmyr

This paper has been presented as:

Martin Hemmer, Tor I. Waag and Kjell G. Robbersmyr. A Review of Methods for Condition Monitoring of Large, Slow-rotating Bearings. Presented at: *30th Conference on Condition Monitoring and Diagnostic Engineering Management.* University of Central Lancashire, Preston, UK

A Review of Methods for Condition Monitoring of Large, Slow-rotating Bearings

Martin Hemmer*, Tor I. Waag** and Kjell G. Robbersmyr*

*University of Agder
Department of Engineering Sciences
Jon Lilletunsvei 9, 4879 Grimstad, Norway

**Teknova AS
Tordenskjolds gate 9, N-4612 Kristiansand, Norway

Abstract – Rolling element bearings (REBs) are key components in most rotating machinery. Large, slow-rotating REBs found in heavy industrial applications like offshore drilling equipment, steel- and paper mills and wind turbines are the topic of this paper. In such applications, bearings are normally non-redundant components, meaning bearing failure will cause system downtime. Execution of unplanned, on-site maintenance may be costly, time-consuming and difficult or even impossible. Implementation of condition-based maintenance strategies is a means to reduce total lifecycle costs by improving utilization of component lifetime while maintaining system availability. Condition monitoring systems capable of early, reliable detection and diagnosis of incipient faults is necessary for the planning of maintenance actions in due time. In this paper, novel and established condition monitoring methods are surveyed for this purpose. Prominent challenges are speed variations, non-stationary behavior, and low signal-to-noise ratio. Advanced signal processing methods, including order tracking and resampling from time to angular domain, higher order statistics, and cyclic spectral analysis are presented. Methods for data acquisition and maintenance decision making are also discussed. A discussion of the surveyed methods and suggestions for future research concludes the paper.

A.1 Introduction

Roller Element Bearing (REB)s are essential mechanical components, used in virtually all types of rotating machinery. The range of types, variants and sizes match the diversity of applications. This paper aims to provide insight into condition monitoring (CM) methods suitable for large, slow-rotating bearings, typically found in paper and steel mills, offshore drilling equipment, wind turbines and similar heavy industries. It is difficult to define strict limits with regards to size and speed. However, other similarities can be defined. Replacement and maintenance is expensive, time-consuming, and in many cases

not possible on site. Combined with operational non-redundancy, this motivates CM for increased control of machine health. The ability to utilize more of the component lifetime while reducing the risk of unexpected failure potentially reduces lifecycle costs.

A survey presented at the Noble Analyst Day 2012 mapped causes of downtime on drilling rigs in the period 2011-December 2012 [1]. Looking at downtime by equipment on all rig types, top drive failure is the second largest contributor, with a total of 13 %. Further analysis shows that bearing failure is the main overall cause of downtime in top drives, despite variations between manufacturers and types. In top drives, a large tapered roller thrust bearing supports the weight of the drill string. With an outer diameter of up to 750mm and a rotational speed of 240 rpm (4 Hz), top drive main bearings qualify as both large and slow-rotating. The statistics show a potential for improved CM of large, slow-rotating bearings in the offshore industry.

A.2 Condition Based Maintenance Strategies

Most systems require maintenance to a certain extent. The approach to maintenance is influenced by factors such as consequences of failure, maintenance cost, and failure rates. Development in technology has enabled more advanced maintenance strategies beyond corrective and preventive maintenance strategies. Lee et.al [2] reviewed the field of Prognostics and Health Management (PHM) for rotating machinery, presenting a generalized methodology for selection and implementation of a maintenance strategy. A maintenance transformation map is proposed as a guideline for maintenance strategy selection based on system complexity and uncertainty. Typically, condition monitoring of bearings falls in the Condition-Based Maintenance (CBM) category. However, as bearings get larger, maintenance is more complicated. External factors like long spare part lead times, complicated maintenance procedures and limited maintenance opportunities add complexity, justifying strategies like PHM. Common for both CBM and PHM is the need for reliable health assessments for the equipment. Jardine et.al [3] reviews generalized diagnostics and prognostics methods for successful implementation of CBM, identifying data acquisition, data processing and maintenance decision-making as the three main steps. The influence of bearing size and speed on these steps is discussed in the next section.

A.3 Challenges of Large, Slow Bearings

Large and slow-rotating bearings pose challenges for conventional condition monitoring methods. Compared to smaller bearings operating at higher speeds, prominent challenges include low vibration energy, sensitivity to speed fluctuations and a need for accurate

localization in the frequency domain to isolate fault frequencies.

It is commonly accepted that discrete faults in bearings cause impulse-like impacts when the fault interacts with another rolling surface. This impact triggers a transient response at resonance frequencies in the bearing, surrounding structure and transducer. As acceleration is the second derivative of displacement, a reduction in rotational speed leads to significant reduction in acceleration levels. For slow-rotating bearings, fundamental fault frequencies will also be relatively close in the frequency domain, increasing the risk of interference. Bechhoefer recommends a minimum of 10, preferably 30, frequency bins between fault frequencies [4]. Frequency resolution is the inverse of acquisition time, which makes the measurement more prone to capture speed fluctuations, leading to smearing of the frequency spectrum. Larger size also means increased distance between fault and transducer. All these factors contribute to a lower signal to noise ratio (SNR).

A.4 Modeling Bearing Faults

The periodic nature of impacts can be modeled mathematically. Time between impacts is governed by a combination of shaft speed, bearing geometry and localization of the fault. For REBs, faults can be associated with the characteristic fault frequencies of the bearing components, Ball Pass Frequency Inner race (BPFI), Ball Pass Frequency Outer race (BPFO), Cage Pass Frequency (CPF) and Ball Spin Frequency (BSF) [5]. Normalizing by shaft frequency transforms frequencies to shaft order domain for easier comparison across operating speeds.

Characteristic fault frequencies assume ideal operating condition, including perfect rolling motion between rolling elements and races. In reality, rolling elements experience some random slip, causing variation in the time between impacts. Additionally, the impact response amplitude can be periodically modulated with smaller random variations. Antoni includes this randomness in a more realistic model for a bearing vibration signal, given in Eq. (A.1) [6]. The vibration signal is $x(t)$, where $h(t)$ is the response to a single impact, $q(t) = q(t + P)$ is periodic modulation caused by load distribution of period P, and T represents the time between the arrival of two consecutive impacts. The random jitter in arrival time and amplitude is handled by τ_i and A_i respectively.

$$x(t) = \sum_{i=-\inf}^{i=\infty} h(t - iT - \tau_i)q(iT)A_i + n(t) \tag{A.1}$$

This randomness in arrival times causes smearing in the frequency spectrum, but allows separation of the bearing signal from deterministic frequency components from gears and shafts [5].

A.4.1 Cyclostationarity

Processes that shows cyclic behavior is said to exhibit cyclostationarity [6]. Cyclostationary theory provides a generalized framework for describing a wide range of stationary and non-stationary processes [7]. In the context of cyclostationarity, a periodic component at a frequency α is referred to as the cyclic frequency. The period of α is termed cycle. These terms are used to avoid confusion with spectral frequency f and its period T. Figure 1 shows a time signal, highlighting the difference by indicating the cycle of α and period of f.

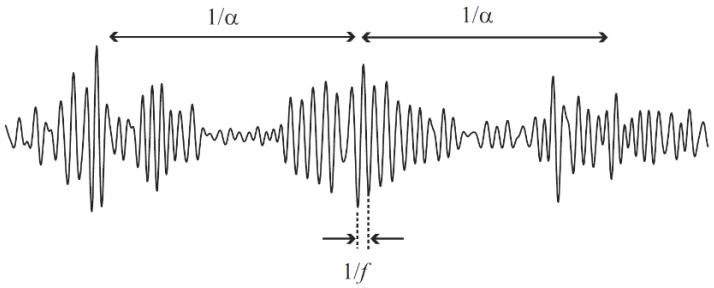


Figure A.1: The difference between cyclic frequency and spectral frequency [7]

A signal can exhibit cyclostationarity at different orders. As an example, a periodic signal masked with additive white noise will have a periodic mean value and thus exhibit first-order cyclostationarity. Consider a signal of amplitude-modulated white noise only. As the mean value is constant, no periodic first-order components exist. Squaring the signal, a second-order transformation, reveals periodic components and consequently second-order cyclostationarity in the signal. A second order transformation is normally enough to reveal bearing diagnostics information. Interested readers can consult the works of Randall [5], [8], [9] and Antoni [6], [7], [10] in particular for further information on the topic and its applications. Cyclic spectral analysis, based on cyclostationary theory, is introduced in Section A.4.1.

A.5 Condition Monitoring Methods

This section is divided in three, discussing methods for data acquisition, data processing and maintenance decision-making; identified by Jardine et.al [11] as the three main steps in CBM. Here, signal enhancement is included as data processing.

A.5.1 Data Acquisition

Choosing a measurement technique capable of observing the symptoms of failure is critical. Tandon and Choudhury [12] identifies four main categories for bearing fault detection

methods; vibration measurements, acoustic measurements, lubrication analysis and temperature measurements.

A.5.1.1 Vibration Measurements

Vibration monitoring using accelerometers is widely used in the industry, and has been researched actively since the 1980s [12]. Traditional vibration analysis faces some challenges when applied to large, slow bearings, as discussed in Section A.2. Displacement measurements can also be used for monitoring vibration. Measuring displacement directly instead of acceleration makes it suitable for slow applications with low acceleration levels. Shakya et.al [13] investigated the use of proximity probe as a standalone method for bearing fault detection and in combination with an accelerometer, and showed improved detectability for inner race defects.

A.5.1.2 Acoustic Measurements

Acoustic measurements refer to vibrations from 20 kHz and upwards, including both ultrasonic and Acoustic Emission (AE) measurements. An increase in AE activity could be an early indication of oil degradation. AE activity in bearings can be related to metal-to-metal contact, indicating a broken oil film. Yoshioka and Fujiwara [14], [15] showed in early research that AE could detect faults before vibration methods. Further research successfully used AE for detection of subsurface cracks [16], which Tan [17] concluded could be useful for detection of pitting. Chacon et. al [18] presented a method for incipient fault detection in REBs using AE measurements for envelope analysis. Jamaludin and Mba review monitoring of extremely slow REBs [19, 20], using AE measurements to detect faults at very low speeds.

A.5.1.3 Lubrication Analysis

Many bearing failure modes can be related to insufficient lubrication; fatigue, wear, corrosion, deformation, and fracture. Examples of condition indicators from lubrication analysis are accumulated particle mass, water content, viscosity, conductivity and debris analysis. Dempsey [21] compared oil debris analysis and vibration based CIs for detection of pitting. All CIs increased when pitting occurred. However the article highlighted the need for improved threshold setting and combination of CIs for improved reliability. Bechhoefer et.al [22] compared lubricant, vibration and temperature data from wind turbine bearings. Lubrication analysis indicated a fault in one damaged bearing but also gave one false alarm. In combination with vibration analysis better accuracy was achieved.

A.5.1.4 Temperature Measurements

The use of temperature as a CI alone is not likely sufficient, especially in offshore equipment where ambient temperature will vary. Load variations will also affect bearing temperature, and lower rotational speed will generate less heat than in high-speed bearings. This combination makes temperature measurements less suitable for condition monitoring. In a comparative study of vibration, lubricant analysis and temperature for condition monitoring [22], temperature failed to indicate failure on a large, slow-rotating wind turbine bearing.

A.5.2 Signal Enhancement

In cases where information carried in the signal is severely masked, pre-processing techniques can be applied to separate, enhance or in other ways improve the signal of interest. The methods do not provide any diagnostics information on their own, but facilitate the use of other CM methods.

A.5.2.1 Correcting for Speed Variation

Variations in shaft rate during sampling will distribute frequency content across more bins in the spectra. As discussed, characteristic fault frequencies are close for large, slow-rotating bearings, which makes low-speed applications sensitive to variations in shaft speed. Order tracking corrects for shaft speed variations by using shaft angle as a reference instead of time. In addition to the original measurement, shaft angle is recorded simultaneously. Measured data is then resampled to angular domain, which effectively manipulates the sampling frequency. Bechhoefer et. al [4] recorded speed variations in the range of 2 % on a wind turbine, which caused smearing of frequency content. Order tracking effectively removed these variations, resulting in a sharper spectrum. Resampling vibration data improved the detection and diagnosis [23, 24, 25].

A.5.2.2 Isolating the Bearing Signal

Shafts and gears can interfere heavily with the bearing signature, particularly in slow-rotating applications where the SNR of fault signatures is low. Linear prediction, adaptive and self-adaptive noise cancellation, Discrete/random separation and Time Synchronous Averaging (TSA) for use in bearing diagnostics are presented by Randall and Antoni in [5]. Borghesani et.al [26] demonstrated a cepstrum based pre-whitening method for extraction of the bearing signature. Common for all methods is that the bearing signature can be isolated from discrete frequency components by exploiting the randomness of bearing vibration as opposed to deterministic gear and shaft signatures.

A.5.3 Established CM methods

In the context of bearing condition monitoring, analysis of time waveform data from vibration, acoustic and ultrasonic is the established industry standard, with notable work by Tandon, Nakra and Choudhury [12, 27], Kim et.al [28], Ho [29] and Randall [8], [30]. Fourier analysis is a fundamental CM method, particularly using the Discrete Fourier Transform (DFT). However, diagnostics information often lies in the periodic modulation of a given carrier frequency. Thus, Fourier analysis of the raw signal alone might not be able to detect or diagnose faults. Envelope analysis is perhaps the best example, widely regarded as a benchmark for bearing fault detection [5]. Here, Fourier analysis is preceded by bandpass filtering around an assumed carrier frequency before the signal envelope is calculated. A challenge is to choose the correct bandpass filter. In digital signal processing, envelope extraction of a signal is often done by taking the absolute value of its Hilbert transform [5]. Then, Fourier analysis is performed on the envelope to reveal frequency component corresponding to the fault frequencies from Section A.4. Envelope analysis and applications have been thoroughly examined in [8], [12], [27], [29, 30, 31].

Other methods can be applied directly in time domain, such as Root-Mean-Square (RMS) and Crest Factor (CF). Kim et. Al [28] compared vibration and ultrasonic measurements for bearing fault detection across a range of low speeds. RMS was shown to decrease almost linearly with shaft speed. However, healthy bearings have shown big variance in RMS values, indicating RMS change is a better CI than predefined threshold values. CF, the ratio of peak amplitude to RMS value, will increase immediately when a fault first appears. Williams et.al [32] recorded the CF in bearing run-to-failure experiments, and reported an increase followed by a decrease as the fault developed. This indicates fault detection capabilities, but limitations as a trendable parameter for diagnostics purposes.

A.5.4 Advanced CM methods

Traditional CM methods are often insufficient for reliable fault detection in large, slow-rotating bearings. This section presents a selection of advanced and novel CM methods available to overcome the challenges.

A.5.4.1 Higher Statistical Moments

For nominal bearings, the acceleration Probability Density Function (PDF) can be assumed to have a Gaussian distribution. Thus, any changes in the shape of the PDF can indicate failure [12]. Statistical moments of first and second order, mean (μ) and standard deviation (σ) respectively, are well known. For CM purposes, moments of a higher order

k, calculated as in Eq. (A.2). Mainly skewness ($k = 3$) and kurtosis ($k = 4$), are used as CIs.

$$\frac{1}{N}\sum_{n=1}^{N}\left(\frac{x[n]-\mu}{\sigma}\right) \tag{A.2}$$

Skewness describes the asymmetry of a distribution, i.e. the relative energy above and below the mean. Nguyen et.al [33] identified skewness as one of three optimal features for reliable fault detection in low-speed bearings, but skewness is not consistently reported as a reliable CI. Kurtosis is as a measure of tailedness, i.e. the presence of tail extremities in a dataset [34]. A Gaussian distribution always has a kurtosis of 3. High amplitude accelerations from impacts yield a heavy-tailed distribution and high kurtosis. This makes kurtosis suitable as a standalone CI, requiring no prior knowledge to quantify the condition.

A.5.4.2 Spectral Kurtosis and the Kurtogram

Spectral Kurtosis (SK) identifies non-Gaussian components in signals along with their location in the frequency spectrum. The method was proposed in 1983 by Dwyer [35]. Wang, Y et. al [36] and Wang, P. et. Al [37] published a review on the use of SK for fault detection, diagnostics, and prognostics for bearings. SK has also been shown to aid optimal selection of frequency band for envelope analysis [38]. The kurtogram was proposed by Antoni in [39], mapping SK as a function of center frequency and filter bandwidth. Wang, P. et. al [37] similarly utilized SK for frequency band selection, but proposed an enhanced kurtogram based on kurtosis of the power spectrum.

A.5.4.3 Wavelets

A wavelet is a waveform with a limited duration that integrates to zero and can be scaled and shifted in time. The Wavelet transform (WT) provides a time-scale representation of the signal, where scale is qualitatively comparable to frequency. An important advantage is the good time resolution at high frequencies and high frequency resolution at low frequencies. Klepka presented a wavelet-based demodulation technique [40], which combined the use of the continuous and discrete-time WT for filtering, envelope estimation and fault detection on synthetic bearing data. Gelman et. al [41] proposed an improved method, using SK for optimal selection of frequency band while maintaining the advantages of wavelet demodulation compared to Fourier analysis.

A.5.4.4 Empirical Mode Decomposition

Empirical Mode Decomposition (EMD), also known as the Hilbert-Huang Transform, obtains instantaneous frequency information of an oscillatory signal by separating it into several Intrinsic Mode Functions (IMFs) which can be amplitude- and frequency-modulated non-linearly. EMD is the data-driven and adaptive, as IMFs are based on the sampled signal only. Lei et.al [42] reviews the application of EMD to fault diagnosis of rotating machinery. Žvokelj et. al [43] demonstrated fault detection on large, slow rotating bearings using EMD and Principal Component Analysis (PCA).

A.5.4.5 Cepstrum Analysis

The (real) cepstrum, defined in Eq. (A.3), identifies repeating "echoes" of a signal, which can be used for detection of periodic signatures. Bechhoefer et.al [4] tested cepstrum analysis for fault detection on wind turbine main bearings, and observed indications of an outer race fault. Cepstrum RMS and kurtosis were tested as possible CIs, but were not able to give actionable results alone. Further study was recommended.

$$cepstrum = IFT\{\ln|FT\{x\}|\} \tag{A.3}$$

A.5.5 Cyclic Spectral Analysis

Cyclic spectral analysis relies on the concept of cyclostationarity from Section A.4.1. Here, two approaches for detecting cyclostationarity in signals will be introduced. The first extracts periodic components of the instantaneous power, while the second is based on the autocorrelation function. An operator $P\{\cdot\}$ is presented in [7]. The operator is implemented as an estimator, shown in Eq. (A.4), where n is the sample number and T_s is the sample period. The estimator extracts Fourier coefficients at cyclic frequencies α in set A, from a given discrete data sequence $\{\cdot\}$.

$$\hat{P}\{\cdot\}\sum_{\alpha\in A} DFT_\alpha\{\cdot\}e^{j2\pi\alpha nT} \tag{A.4}$$

A.5.5.1 Power Decomposition and Instantaneous Autocorrelation

Cyclostationary behavior can be detected by decomposition of signal power to periodic components. Consider a signal $x[n]$ with power P_x. Estimation of mean instantaneous power $P_x[n]$ is done by applying $\hat{P}\{\cdot\}$ to the signal power $|x[n]|^2$. A Fourier series expansion of $P_x[n]$ then gives the cyclic powers P_x^α. The quantities mean instantaneous power

spectrum and cyclic modulation spectrum are obtained by a time frequency decomposition and Fourier series expansion respectively, further elaborated in [7]. The presence of periodicity in a signal creates a correlation of spectral components at the cyclic frequency. Antoni describes in [6], [7] how the instantaneous autocorrelation function can be utilized to detect cyclostationary behavior. Given a signal $x(t)$, the instantaneous autocorrelation $R_x(t,\tau)$ is defined as applying $P\{\cdot\}$ to the symmetric autocorrelation function, shown in Eq. (A.5).

$$R_x(t,\tau) = P\{x(t-\tau/2)x(t-\tau/2)\} \tag{A.5}$$

The Fourier series expansion of the instantaneous autocorrelation function expressed is called the cyclic autocorrelation function $R_x^\alpha(\tau)$. Applying a Fourier transform to $R_x^\alpha(\tau)$ yields the spectral correlation density $SC_x^\alpha(f)$, a frequency-frequency representation of $x(t)$. Note that cyclic frequency α is the frequency counterpart of time, and spectral frequency f is the dual of shift τ. The spectral correlation (SC) is non-zero if a frequency component f is periodic with cyclic frequency α. These connections and the relationship to the Wigner-Ville spectrum, classical autocorrelation function, and the PSD are examined in [7].

A.5.6 Maintenance Decision-making

A condition monitoring system should be able to make or aid in maintenance decisions. However, a single CIs may be insufficient to provide reliable decisions. Fusion of data of different types, both on sensor and feature level can be utilized for improved diagnostics and prognostics of bearings [2]. Dempsey and Loutas [21], [44] investigates a combination of on-line oil analysis, AE and vibration as a way of improving CI performance. Bechhoefer et. al [45] presents a method for optimal threshold setting, by fusing several CIs in a Health Index (HI), which quantifies bearing damage without the need for user interpretation. The HI is constructed from the norm of n Gaussian CIs, and can be shown to form a Nakagami-distributed PDF. This method allows for setting a desired Probability of False Alarm (PFA) and normalizing the HI to be 1 when this probability is reached. The method is successfully demonstrated on data from three large, slow wind turbine bearings [4], where the faulty bearing was shown to have a HI well above one.

A.6 Conclusions

This paper presents an overview of relevant CM methods for large, slow-rotating bearings. The combined requirement of cost reduction and uptime facilitates the emergence of more

advanced condition monitoring systems. Main challenges of condition monitoring of large, slow-rotating bearings can be summarized by a low energy impacts, large distance from fault to transducer, comprehensive background noise and speed variations, resulting in a low SNR. Detection capabilities of traditional CM methods, especially envelope analysis, can be improved by longer acquisition times, order tracking and separation of random and discrete components. SK aided bandpass filtering before envelope extraction further improves performance. Other data acquisition methods can also be used. AE signals carries similar diagnostics information as vibration, but in a frequency band less subjected to noise.

Another development in bearing condition monitoring is the transition from a stationarity assumption implicated by the Fourier transform to a more realistic, non-stationary or cyclostationary approach. Time-frequency and cyclostationary analysis tools takes this into account. In cases where cyclic behavior is heavily masked in non-stationary signals, cyclostationary analysis appears to be a powerful tool.

It seems unlikely to find a single CI, data acquisition or signal processing method that solves all challenges for CM of large, slow-rotating REBs. Hence, combining CM data from different sources seems more reasonable. The concept of a PFA-controlled HI is attractive from an operator point of view, and can preferably be utilized in systems for automated fault detection and diagnostics. Finding good CIs and methods for fusing them should be a priority in future work.

A.7 Acknowledgement

The research presented in this paper has received funding from the Norwegian Research Council, SFI Offshore Mechatronics, project number 237896.

References

[1] L. Jeffrey. Noble 2012 Analyst & Investor Day Presentation, 2012. Presented: 24.05.2012, Houston, Texas, USA. Date obtained: 31.08.2016.

[2] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel. Prognostics and Health Management Design for Rotary Machinery Systems—reviews, Methodology and Applications. *Mechanical Systems and Signal Processing*, 42(1-2):314–334, 2014. doi:10.1016/j.ymssp.2013.06.004.

[3] A. K. S. Jardine, D. Lin, and D. Banjevic. A Review on Machinery Diagnostics and Prognostics Implementing Condition-based Maintenance. *Mechanical Systems and Signal Processing*, 20(7):1483–1510, 2006. arXiv:0208024, doi:10.1016/j.ymssp.2005.09.012.

[4] E. Bechhoefer, R. Schlanbusch, and T. I. Waag. Techniques for Large, Slow Bearing Fault Detection. *International Journal of Prognostics and Health Management*, 7(1):1–12, 2016.

[5] R. B. Randall and J. Antoni. Rolling Element Bearing Diagnostics - A Tutorial. *Mechanical Systems and Signal Processing*, 25(2):485–520, 2011. doi:10.1016/j.ymssp.2010.07.017.

[6] J. Antoni. Cyclic Spectral Analysis of Rolling-element Bearing Signals: Facts and Fictions. *Journal of Sound and Vibration*, 304(3-5):497–529, 2007. doi:10.1016/j.jsv.2007.02.029.

[7] J. Antoni. Cyclostationarity by Examples. *Mechanical Systems and Signal Processing*, 23(4):987–1036, 2009. doi:10.1016/j.ymssp.2008.10.010.

[8] R. B. Randall. State of the Art in Monitoring Rotating Machinery – Part 1. *Journal of Sound and Vibration*, 38(3):14–21, 2004.

[9] R. B. Randall, J. Antoni, and S. Chobsaard. The Relationship Between Spectral Correlation and Envelope Analysis in the Diagnostics of Bearing Faults and

Other Cyclostationary Machine Signals. *Mechanical Systems and Signal Processing*, 15(5):945–962, 2001. doi:doi:10.1006/mssp.2001.1415.

[10] J. Antoni. Cyclic Spectral Analysis in Practice. *Mechanical Systems and Signal Processing*, 21:597–630, 2007. doi:10.1016/j.ymssp.2006.08.007.

[11] A. H. C. Tsang, W. K. Yeung, A. K. S. Jardine, and B. P. K. Leung. Data Management for CBM Optimization. *Journal of Quality in Maintenance Engineering*, 12(1):37–51, 2006.

[12] N. Tandon and A. Choudhury. A Review of Vibration and Acoustic Measurement Methods for the Detection of Defects in Rolling Element Bearings. *Tribology International*, 32(8):469–480, 1999. doi:https://doi.org/10.1016/s0301-679x(99)00077-8.

[13] P. Shakya, A. K. Darpe, and M. S. Kulkarni. Bearing Diagnosis Using Proximity Probe and Accelerometer. *Measurement: Journal of the International Measurement Confederation*, 80:190–200, 2016. doi:10.1016/j.measurement.2015.11.029.

[14] T. Yoshioka and T. Fujiwara. Application of Acoustic Emission Technique to Detection of Rolling Bearing Failure. *American society of mechanical engineers*, 14(1):55–76, 1984.

[15] T. Yoshioka and T. Fujiwara. A New Acoustic Emission Source Locating System for the Study of Rolling Contact Fatigue. *Wear*, 81(1):183–186, 1982. doi:10.1016/0043-1648(82)90314-3.

[16] T. Yoshioka. Detection of Rolling Contact Sub-surface Fatigue Cracks using Acoustic Emission Technique. *Lubrication Engineering*, 04(1), 1993.

[17] C. K. Tan and D. Mba. Identification of the Acoustic Emission Source During a Comparative Study on Diagnosis of a Spur Gearbox. *Tribology International*, 38(5):469–480, 2005. doi:10.1016/j.triboint.2004.10.007.

[18] J. L. Ferrando Chacon, V. Kappatos, W. Balachandran, and T.-H. Gan. A Novel Approach for Incipient Defect Detection in Rolling Bearings Using Acoustic Emission Technique. *Applied Acoustics*, 89:88–100, 2015. doi:10.1016/j.apacoust.2014.09.002.

[19] N. Jamaludin and D. Mba. Monitoring Extremely Slow Rolling Element Bearings: Part I. *NDT and E International*, 35(6):349–358, 2002. doi:10.1016/S0963-8695(02)00006-3.

[20] N. Jamaludin and D. Mba. Monitoring Extremely Slow Rolling Element Bearings: Part II. *NDT and E International*, 35(6):359–366, 2002. doi:10.1016/S0963-8695(02)00006-3.

[21] P. J. Dempsey. A Comparison of Vibration and Oil Debris Gear Damage Detection Methods Applied to Pitting Damage. Technical Report September, 2000.

[22] E. Bechhoefer, R. Schlanbusch, and T. I. Waag. Fault Detection on Large Slow Bearings. In *European Conference of the Prognostics and Health Management Society 2016*, volume 7, pages 1–8, 2016.

[23] S. Priya, M. R. Ramesh, and V. Naidu. Bearing Health Condition Monitoring: Frequency Domain Analysis Multi-sensor Data Fusion. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (An ISO Certified Organization)*, 3(5):260–268, 2014.

[24] 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*, volume 24, pages 1998–2011, 2010. doi:10.1016/j.ymssp.2010.05.005.

[25] E. Bechhoefer and M. Kingsley. A Review of Time Synchronous Average Algorithms. In *Annual Conference of the Prognostics and Health Management Society*, pages 24–33, 2009.

[26] P. Borghesani, P. Pennacchi, R. Ricci, and S. Chatterton. Testing Second Order Cyclostationarity in the Squared Envelope Spectrum of Non-white Vibration Signals. *Mechanical Systems and Signal Processing*, 40(1):38–55, 2013. doi:10.1016/j.ymssp.2013.05.012.

[27] N. Tandon and B. C. Nakra. Comparison of Vibration and Acoustic Measurement Techniques for the Condition Monitoring of Rolling Element Bearings. *Tribology International*, 25(3):205–212, 1992. doi:10.1016/0301-679X(92)90050-W.

[28] E. Y. Kim, A. C. C. Tan, J. Mathew, and B. S. Yang. Condition Monitoring of Low Speed Bearings: A Comparative Study of the Ultrasound Technique Versus Vibration Measurements. In *Proceedings of the 1st World Congress on Engineering Asset Management (WCEAM)*, volume 5, pages 177–189, 2008. doi:10.1007/978-1-84628-814-2_21.

[29] D. Ho and R. B. Randall. Optimisation of Bearing Diagnostic Techniques Using Simulated and Actual Bearing Fault Signals. *Mechanical Systems and Signal Processing*, 14(5):763–788, 2000. doi:10.1006/mssp.2000.1304.

[30] R. B. Randall. State of the Art in Monitoring Rotating Machinery – Part 2. *Journal of Sound and Vibration*, 38(5):10–16, 2004.

[31] D. Hochmann and E. Bechhoefer. Envelope Bearing Analysis: Theory and Practice. In *IEEE Aerospace Conference Proceedings*, pages 3658–3666, 2005. doi:10.1109/AERO.2005.1559671.

[32] T. Williams, X. Ribadeneira, S. Billington, and T. Kurfess. Rolling Element Bearing Diagnostics In Run-to-failure Lifetime Testing. *Mechanical Systems and Signal Processing*, 15(5):979–993, 2001. doi:10.1006/mssp.2001.1418.

[33] P. Nguyen, M. Kang, J. Kim, and J.-M. M. Kim. Reliable Fault Diagnosis of Low-speed Bearing Defects Using a Genetic Algorithm. In *PRICAI 2014: Trends in Artificial Intelligence 2014*, pages 248–255. Springer, Cham, 2014. doi:10.1007/978-3-319-13560-1.

[34] P. H. Westfall. Kurtosis as Peakedness, 1905–2014. R.i.p. *The American Statistician*, 68(December):191–195, 2014. doi:10.1080/00031305.2014.917055.

[35] R. Dwyer. Detection of Non-gaussian Signals by Frequency Domain Kurtosis Estimation. *ICASSP '83. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 8:607–610, 1983. doi:10.1109/ICASSP.1983.1172264.

[36] Y. Wang, J. Xiang, R. Markert, and M. Liang. Spectral Kurtosis for Fault Detection, Diagnosis and Prognostics of Rotating Machines: A Review With Applications. *Mechanical Systems and Signal Processing*, 66-67:679–698, 2016. doi:10.1016/j.ymssp.2015.04.039.

[37] D. Wang, P. W. Tse, and K. L. Tsui. An Enhanced Kurtogram Method for Fault Diagnosis of Rolling Element Bearings. *Mechanical Systems and Signal Processing*, 35(1-2):176–199, 2013. doi:10.1016/j.ymssp.2012.10.003.

[38] E. Bechhoefer, M. Kingsley, and P. Menon. Bearing Envelope Analysis Window Selection using Spectral Kurtosis Techniques. In *2011 IEEE International Conference on Prognostics and Health Management, PHM 2011 - Conference Proceedings*, pages 1–6, 2011. arXiv:ICPHM.2011.6024338, doi:10.1109/ICPHM.2011.6024338.

[39] J. Antoni and R. B. Randall. The Spectral Kurtosis: Application to the Vibratory Surveillance and Diagnostics of Rotating Machines. *Mechanical Systems and Signal Processing*, 20(2):308–331, 2006. doi:10.1016/j.ymssp.2004.09.002.

[40] A. Klepka. Wavelet Based Signal Demodulation Technique for Bearing Fault Detection. *Mechanics and Mechanical Engineering*, 15(4):63–71, 2011.

[41] L. Gelman, T. H. Patel, G. Persin, B. Murray, and A. Thomson. Novel Technology Based on the Spectral Kurtosis and Wavelet Transform for Rolling Bearing Diagnosis. *International Journal of Prognostics and Health Management*, 4(2):1–7, 2013.

[42] Y. Lei, J. Lin, Z. He, and M. J. Zuo. A Review on Empirical Mode Decomposition in Fault Diagnosis of Rotating Machinery. *Mechanical Systems and Signal Processing*, 35(1-2):108–126, 2013. doi:10.1016/j.ymssp.2012.09.015.

[43] M. Žvokelj, S. Zupan, and I. Prebil. Multivariate and Multiscale Monitoring of Large-size Low-speed Bearings Using Ensemble Empirical Mode Decomposition Method Combined With Principal Component Analysis. *Mechanical Systems and Signal Processing*, 24(4):1049–1067, 2010. doi:10.1016/j.ymssp.2009.09.002.

[44] T. H. Loutas, D. Roulias, E. Pauly, and V. Kostopoulos. The Combined Use of Vibration, Acoustic Emission and Oil Debris On-line Monitoring Towards a More Effective Condition Monitoring of Rotating Machinery. *Mechanical Systems and Signal Processing*, 25(4):1339–1352, 2011. doi:10.1016/j.ymssp.2010.11.007.

[45] E. Bechhoefer and A. P. F. Bernhard. A Generalized Process for Optimal Threshold Setting in Hums. In *IEEE Aerospace Conference Proceedings*, pages 1–9, 2007. doi:10.1109/AERO.2007.352867.

Paper B

A Comparison of Acoustic Emission and Vibration Measurements for Condition Monitoring of an Offshore Drilling Machine

Martin Hemmer and Tor I. Waag

This paper has been published as:

Martin Hemmer and Tor I. Waag. A Comparison of Acoustic Emission and Vibration Measurements for Condition Monitoring of an Offshore Drilling Machine. *Proceedings of the Annual Conference of the Prognostics and Health Management Society 2017*, 278-285, 2017. ISBN 978-1-936263-26-4.

A Comparison of Acoustic Emission and Vibration Measurements for Condition Monitoring of an Offshore Drilling Machine

Martin Hemmer* and Tor I. Waag**

*University of Agder
Department of Engineering Sciences
Jon Lilletunsvei 9, 4879 Grimstad, Norway

**Teknova AS
Tordenskjolds gate 9, N-4612 Kristiansand, Norway

Abstract – This paper investigates the application of heterodyned Acoustic Emission (AE) compared to more conventional vibration measurements for Condition Monitoring (CM) of an offshore drilling machine, with a particular focus on the large, axial tapered roller bearing supporting the drill string weight in a top drive. The focus on cost reduction and operational uptime in the oil and gas industry motivates research on improved CM methods for fault detection, identification and ultimately prediction. However, bearing failure on this type of machines are currently responsible for a significant share of operational downtime on drilling rigs. In the experiment, a previously used and replaced bearing is compared to a new, healthy bearing with the purpose of identifying possible condition indicators (CI) from the vibration and AE measurements. AE root-mean-square values (RMS) was identified as a CI, being more consistent with the expected bearing health than vibration measurements and also less affected by operating speed. The AE measurements also show complementary forced frequency identification capabilities compared to the vibration measurements. The particular failure mode with bearing roller end damage is described and seen in conjunction with the results.

B.1 Introduction

The Rolling Element Bearing (REB) is a component found in basically all rotating machinery. It is also a common cause of premature machine failure. As bearings get larger, the consequence of failure typically increases in terms of unplanned downtime cost and potential safety hazards. Simultaneously, maintenance actions are time-consuming, expensive and sometimes impossible to do on-site. The oil and gas industry is moving towards condition based maintenance strategies which require reliable CM methods. For the case of drilling machines, CM of the main bearing has proven to be difficult. A study

of downtime causes on drilling rigs, the drilling machine was shown to be responsible for 13 % of downtime, with the bearing as the largest cause of failure [1]. The current industry standard is mainly a combination of visual inspection, periodic offline vibration measurements, and lubricant analysis. The statistics show an obvious potential for improvement, motivating the development of methods suitable for online CM of the bearing. This experiment is a rare opportunity to do measurements on a real drilling machine with a known damage to the bearing. The goal is to identify condition indicators for the main bearing and compare vibration measurements with AE measurements with regards to fault detection and identification.

Vibration analysis using accelerometers is the current industry standard for bearing CM, and has been researched for decades. Piezoelectric transducers are common for bearing CM, but MEMS-type accelerometers are also in use [2]. Likewise, the application of AE for bearing fault detection is not new in the field of CM of REBs, with early work done by [3, 4]. Heterodyning of the AE signal before sampling can be done to reduce the amount of stored and analyzed data for AE measurements, but maintain diagnostics information [5]. This reduction in data volume makes the technology more accessible for the industry, and is utilized in the experiment. AE measurements has also been combined with self-learning neural networks [6], which can further automate the analysis process.

The majority of research on bearing CM focus on signal processing methods for analysis of the time-waveform output by transducers. An extensive, categorized review of methods for bearing fault detection was presented by [7]. A recent advance is the development of cyclic spectral analysis, with notable work by [8, 9]. Cyclic spectral analysis can be applied to AE as well as vibration signals [10].

Detection and diagnostics of discrete bearing faults are typically done by detecting the presence of one or more fundamental fault frequency; Ball Pass Frequency Inner Race (BPFI), Ball Pass Frequency Outer Race (BPFO), Cage Pass Frequency (CPF) and Ball Spin Frequency (BSF). However, to the authors' knowledge, CM methods specific for the failure mode described in this paper has not been thoroughly researched, opening new possibilities for future work in the field.

B.2 Methods

This section describes the experimental setup, data acquisition systems, and data processing methods used to obtain the presented results.

B.2.1 Experimental Setup

Tests were performed on a large drilling machine taken out of operation for onshore maintenance. The main point of interest is the axial bearing, which normally supports the weight of the drillstring. The idea was to apply different systems to a relevant industrial application to compare performance. The machine already has measurement points for the radial and axial direction of the shaft, as shown in /ure B.1. In this experiment, the only axial load on the bearing is caused by the shaft self-weight. The rotating motion is powered by a single hydraulic motor, connected to the shaft by a spur gear with a 8:1 ratio. Normally the machine uses up to four motors, but as the torsional load is low, only one motor was installed for this experiment. A pump for lubricant circulation is mounted to the side of the machine.

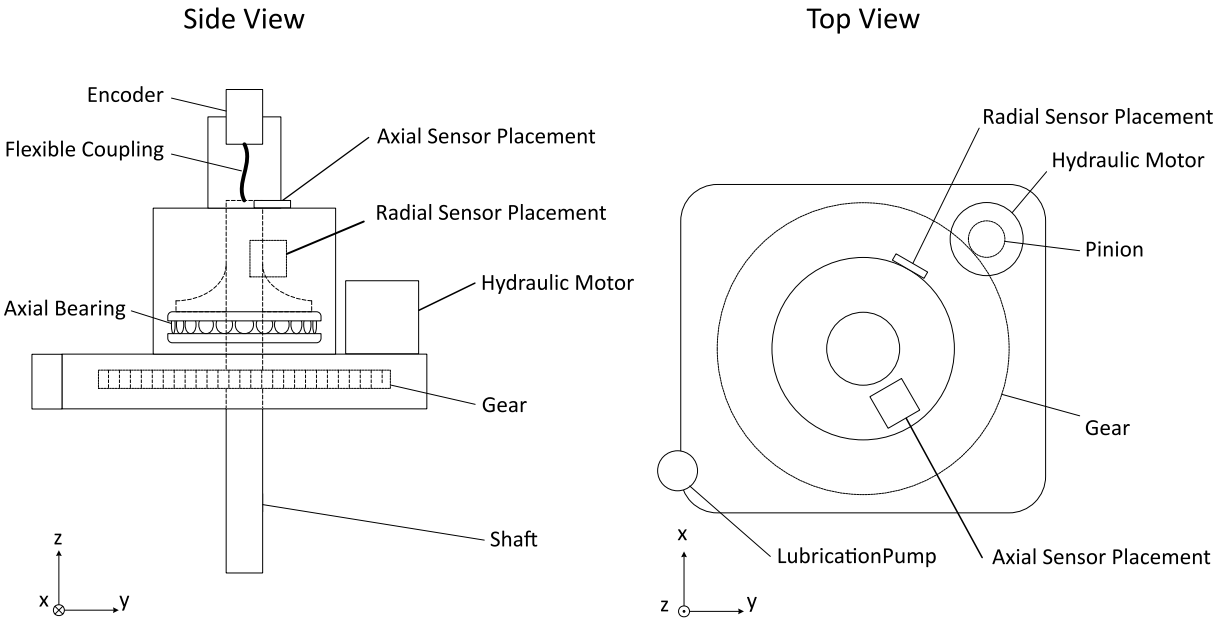


Figure B.1: Experimental setup

B.2.2 Test Parameters

The experiment had two variable parameters; rotational speed and three stages of declining bearing condition. The three stages of declining bearing condition are termed Health Level (HL), with an index from 0 to 2. First, the machine was tested using a new, healthy bearing. This is referred to as HL0. HL1 refers to a previously used bearing, which was replaced due to observed roller end damage, described in Section B.3. The same bearing was subject to artificially created indentations on a roller end and run under poorer lubrication conditions to produce HL2. Basic test parameters including the main bearing fault frequencies is shown in Table B.1. The fault frequencies are given in orders, with

the main shaft as 1X. For each health level, the machine was tested at 5 rotational speeds ranging from 50 rpm (0.83 Hz) to 250 rpm (4.17 Hz). The main bearing was not subject to external axial load except shaft self-weight. The bearing load of 9.83 kN is significantly less than typical load under operation, further complicating fault detection.

Table B.1: Bearing operating information

Test Information	
Bearing diameter	650 mm
Bearing load	9.83 kN
Test speeds, [rpm]	$\{50 - 100 - 150 - 200 - 250\}$
Test speeds, [Hz]	$\{0.83 - 1.67 - 2.5 - 3.34 - 4.17\}$
Main shaft speed	1X
Motor shaft speed	8X
BPFI	8.68X
BPFO	8.32X
BSF	6.75X
CPF	0.49X

B.2.3 Sensor Placement

The drilling machine is equipped with measurement points for routine CM. However, to accommodate several transducers logging simultaneously, an adapter plate was made from a 20mm steel plate. The sensors were placed on a circle centered on the mounting flange to make the signal transmission path as equal as possible.

B.2.4 Instrumentation

Data acquisition was done using 3 different systems, two of which utilizes acceleration transducers while the third one uses an AE transducer. The benchmark for acquisition time was a frequency resolution of 10-30 bins between fault frequencies, as recommended in [2]. System A recorded for 100 revolutions regardless of rotational speed while the other systems sampled fixed length time series. Key specifications are given in Table B.2.

B.2.4.1 System A

System A consists of a piezoelectric accelerometer with a sensitivity of 100 mV/g and linear range of 2 Hz to 10 kHz. However, with a sampling rate of 102.4 kHz, the signal is

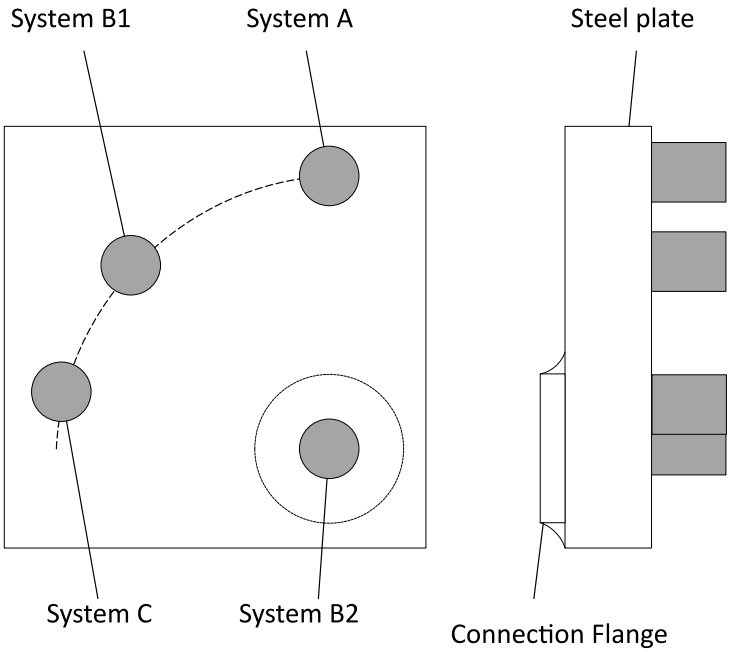


Figure B.2: Sensor placement on adapter plate

Feature	**System A**	**System B**	**System C**
Transducer	Acceleration	Acceleration	AE
Frequency	2 Hz -10 kHz	2 Hz -10 kHz	50-400 kHz
Sensitivity	10 mV/ms^{-2}	1.0 $\mu A/ms^{-2}$	69 dB (peak)
Sample rate	102.4 kHz	10 kHz	50 kHz
Recording	100 rev	60 s	120 s

Table B.2: Instrumentation

oversampled compared to the linear range. This was done to capture encoder data, which was logged simultaneously on the same system to allow rpm calculation and order tracking.

B.2.4.2 System B

System B is part of a commercially available CM system, using an Integrated Electronics Piezoelectric Accelerometer with a current output sensitivity of 1.0 $\mu A/ms^{-2}$. For the purpose of this paper, the time-waveform is extracted and analyzed to ensure control of signal processing methods used. Sampling is limited to 10 kHz for 60 seconds. The system is installed at two locations, named B1 and B2 for distinction, shown in Figure B.2. System B2 is mounted directly over the connection flange, where the accelerometer is normally mounted, to serve as a reference measurement.

B.2.4.3 System C

System C uses an AE transducer glued to the measurement surface. The frequency range is 50 to 400 kHz. However, the signal is heterodyned and demodulated by an analog circuit before sampling, which allows the sampling frequency of 50 kHz. The preprocessing method applied to AE signals was presented in [11]. The high frequency heterodyne product is filtered out, and the low frequency is phase-shifted by $\frac{\pi}{2}$, which effectively is the complex part of the Hilbert transform of the signal. Sampling both the original signal and the phase-shifted version allows demodulation simply by calculating the absolute value of the two.

B.3 Failure Mode

The bearing is a pure axial tapered roller bearing, supporting the drillstring. The rollers are tapered to maintain rolling line contact between roller and races during rotation. A retaining flange is necessary to keep the rollers in place due to the wedge effect of the tapered roller. Simplified, the forces acting on the roller is shown in Figure B.3. The axial load F_A is decomposed in two components; F_N normal to the tapered bearing raceway, and F_C acting on the roller end towards the roller apex point. The relative magnitude of F_N and F_C depends on the cone angle β. Due to the inevitable sliding contact at the roller end, this area is exposed to surface wear. This corresponds well with the observed damage on the worn bearing in the drilling machine, which initiated the maintenance action.

Arc-shaped scratches with varying radius are observed across the roller end surface, as shown in Figure B.4. Generation of such damage can be explained by observing the trace of a particle stuck on the retaining flange, which is passed by a roller. Figure B.5 shows this trace for 3 particles at different distances from the rolling surface.

The distribution of scratches appears to be relatively even between rollers. As it is not a clear, discrete fault, an assumption of periodic behavior may be invalid, which makes traditional frequency analysis tools less useful. Generation of scratches implies permanent changes in the metallic structure, which will generate an AE transient. Also, as the number of scratches increases, the accumulated amount of particles in the lubricant should grow at an increasing rate, leading to a similar growth in scratch formation and AE activity.

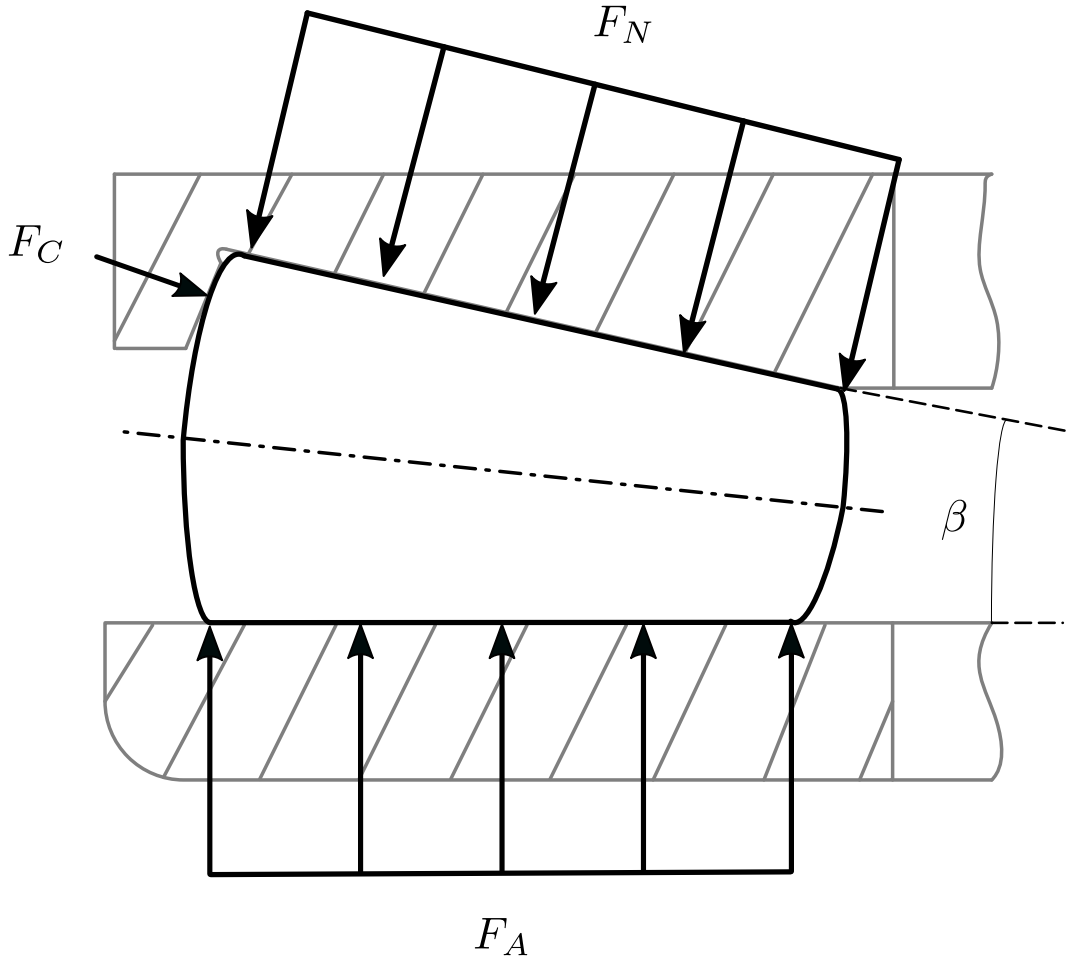


Figure B.3: Internal bearing forces

B.4 Results

The presented results highlight observed differences between the vibration and AE measurement system for overall value trending and feature detection in the frequency domain.

B.4.1 RMS Trending

A change in signal RMS values can be used to indicate a change in condition. In this experiment, there was a known change component health, which was expected to cause an increase in RMS. Measurements from system B2 show an increase in RMS with rpm for all HLs, shown in Figure B.6. However, from HL0 to HL1, the RMS decrease unexpectedly, and remains low at HL2. The same trend occurs in data from systems A and B1, shown in Table B.3. The levels for system B2 are generally higher than for B1 in the axial direction and lower in radial direction. Sensor placement is the only difference between systems B1 and B2, indicating that the adapter plate transfer function has an effect on the measurements. System B2 has the shortest signal transmission path and the stiffest

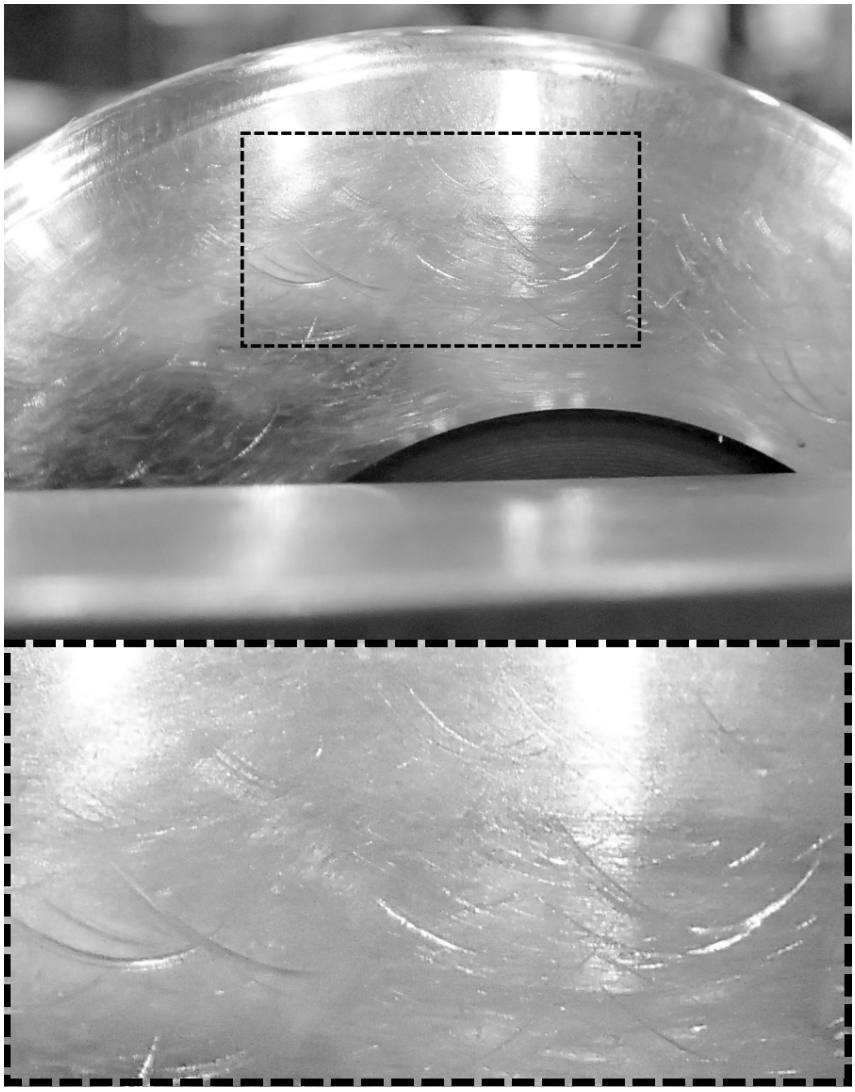


Figure B.4: Characteristic roller end damage

connection, which should produce the most accurate results.

Interestingly, results from the AE measurements (system C) show a different trend. As shown in Figure B.7 and Table B.3, there is an increase in AE RMS at HL1 and HL2. At HL1, radial RMS values increase with a factor of 1.5-2.2, whereas axial RMS increase by a maximum factor of 1.2 at 250 rpm. At HL2 the increase is distinct in both directions, with a relative increase from HL0 of 5.1 and 4.2 for radial and axial RMS respectively. The radial measurement point, shown in Figure B.1, is located closer to the bearing. The longer signal transmission path can explain the higher AE RMS levels, assuming that the bearing is a source of AE activity. The AE RMS increases with rpm, but less than the corresponding vibration measurements. For an increase in rpm from 50 to 250 rpm, System A and B RMS increase by a factor in the range from 2.6 to 12.9, while for system C the range is 1.6 to 2.3.

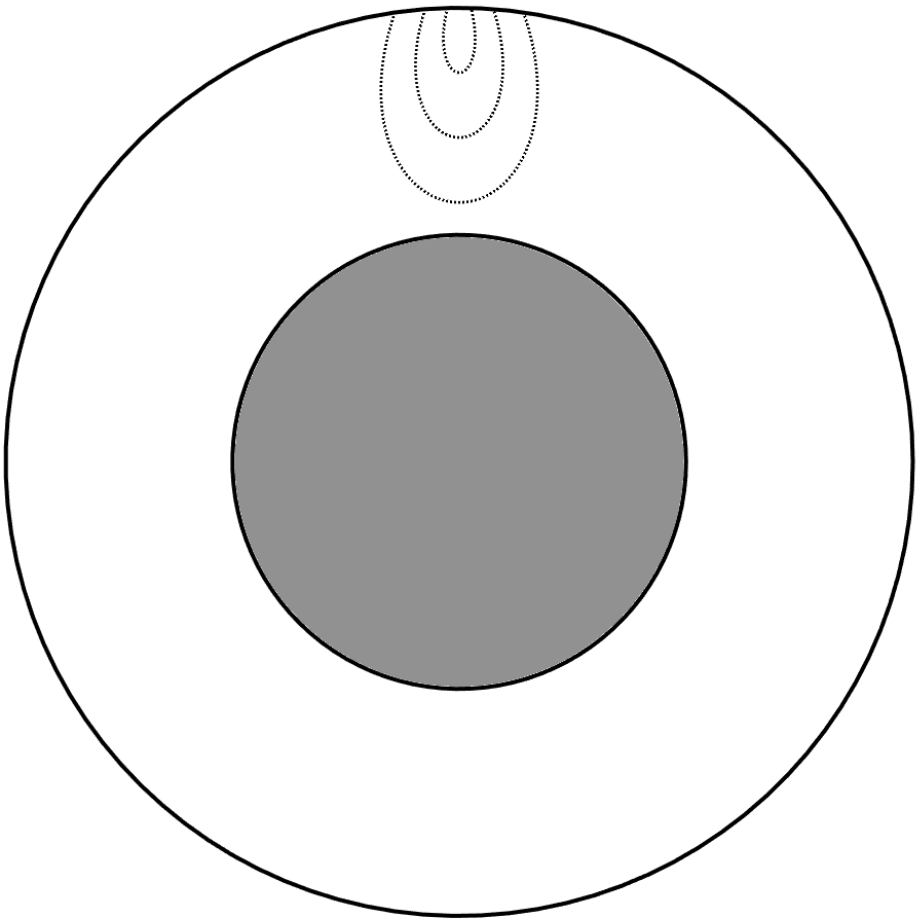


Figure B.5: Trace of particles on a roller end

The power spectrum of the AE signal is shown for 150 rpm in Figure B.8. The reveals an overall, broadband increase in activity. The power spectrum $P(f)$ is calculated as shown in Equation B.1, simply by applying the Fast Fourier Transform, denoted $FFT\{\cdot\}$ to the squared signal. Peaks in the spectrum can not be related to the bearing fault frequencies, but is addressed in Section B.4.2.

$$P(f) = FFT\{|x(t)|^2)\} \tag{B.1}$$

Acceleration RMS-value [10^{-3} g]											
		Axial					Radial				
	rpm	50	100	150	200	250	50	100	150	200	250
System A	HL 0	11.5	30.9	40.0	59.9	85.7	15.7	32.5	46.8	44.6	99.8
	HL 1	12.2	17.1	28.0	36.0	46.0	8.0	13.0	19.8	31.7	41.5
	HL 2	11.6	15.3	23.6	29.3	39.0	9.9	14.1	17.9	33.0	33.7
System B1	HL 0	96.6	256.7	352.8	764.1	819.3	83.2	216.9	278.7	479.5	1069.3
	HL 1	74.8	136.4	183.9	316.7	390.7	81.0	156.9	237.0	461.7	521.0
	HL 2	94.8	146.6	212.4	416.6	438.2	111.1	159.8	207.4	364.3	541.0
System B2	HL 0	266.9	466.0	584.8	973.0	1094.5	45.6	106.5	142.1	195.1	311.7
	HL 1	219.5	325.7	413.2	570.1	780.6	33.8	59.7	85.1	158.9	183.5
	HL 2	293.6	312.5	425.4	558.1	752.6	45.4	54.1	74.1	140.6	171.3
AE RMS-value [10^{-3} V]											
System C	HL 0	59,9	64,2	71,9	87,0	113,3	94,5	93,6	104,2	134,0	168,1
	HL 1	61,4	66,1	72,6	85,9	135,0	203,9	201,2	174,2	196,4	327,4
	HL 2	167,8	266,2	197,4	211,6	337,4	261,2	479,2	313,7	356,2	589,8

Table B.3: RMS Values

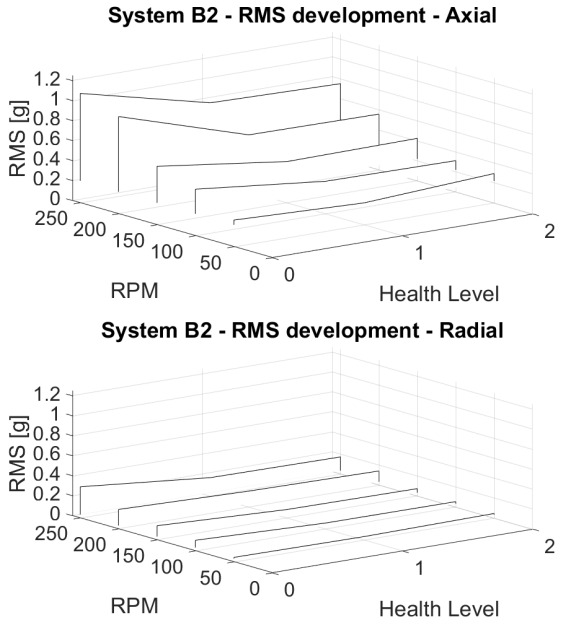


Figure B.6: RMS trend for system B2

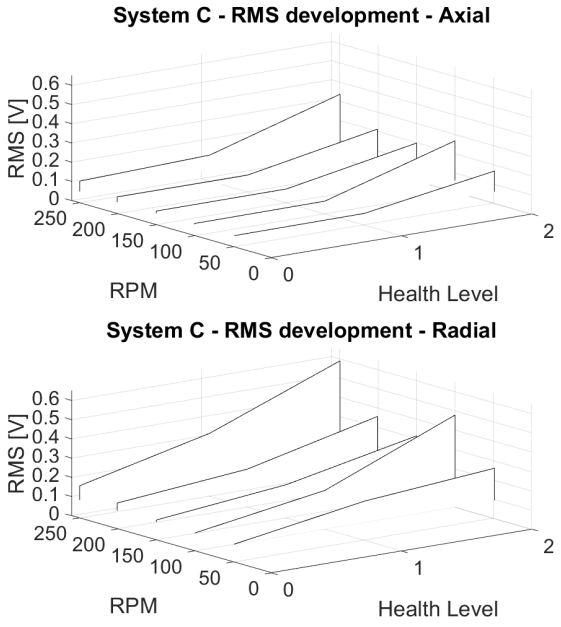


Figure B.7: RMS trend for system C

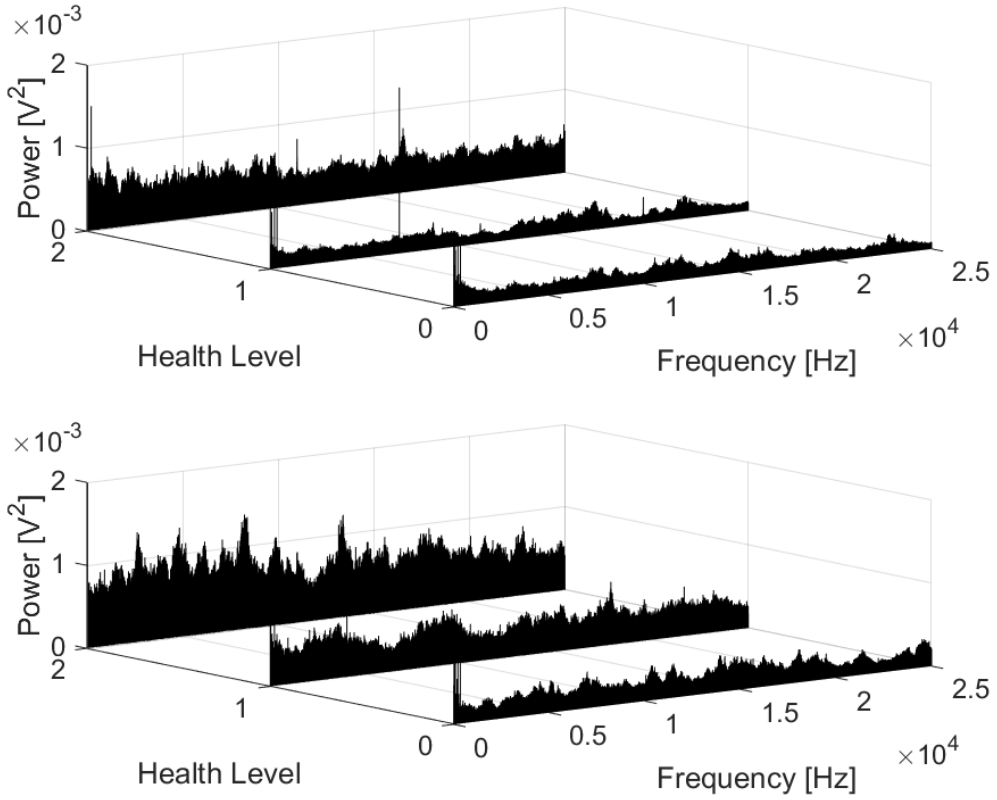


Figure B.8: System C power spectrums at 150 rpm

B.4.2 Forcing Frequency Identification

The processing of recorded data did not reveal any clear indications of faults on component level. Still, the systems show differences in the ability to identify forcing frequencies. Figure B.9 and B.10 show the envelope power spectrum from HL2 at 150 rpm, using data from system A and C, respectively. 1X and 8X harmonics with 1X sidebands dominates the spectrum from vibration measurements, shown in Figure B.9. However, one particular, non-synchronous feature was detected in 14 of 15 AE measurements. Table B.4 shows the peak frequency at the different operating speeds and health levels.

Health Level	0	1	2
rpm	Observed Frequency [Hz]		
50	17.24	16.71	18.91
100	17.30	16.78	16.9
150	17.34	17.04	18.87
200	17.38	17.19	18.76
250	17.47	-	17.1

Table B.4: Identified pump frequencies

The peak appeared at similar frequencies regardless of operating speed. The source of this frequency is assumed to be a small lubrication pump located on the side of the machine, shown in Figure B.1, approximately 1 meter from the measurement point. Completely uncoupled from the main shaft, it was identified as the only component rotating in the detected frequency range.

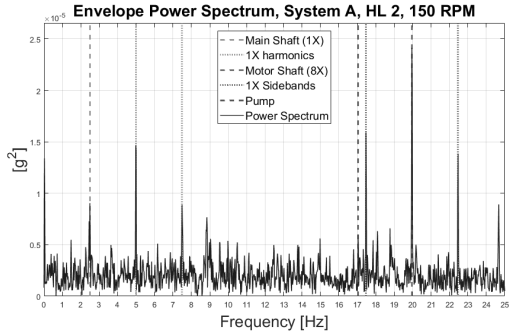


Figure B.9: System A envelope power spectrum

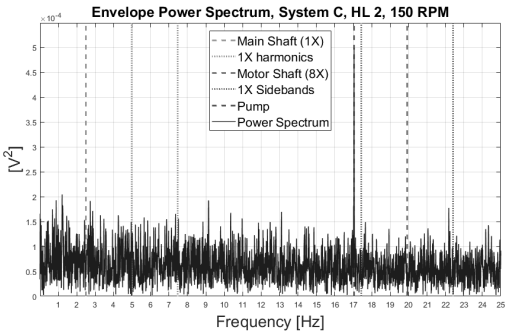


Figure B.10: System C envelope power spectrum

Computation of the fast kurtogram, as described in [12], was used to identify frequency bands with elevated kurtosis. However, no clear improvement in forcing frequency identification was observed. The results shown in table B.4 are calculated for the top level of the kurtogram, which corresponds to low-pass filtering up to the Nyquist frequency before calculating the envelope. For the vibration measurements, the envelope power spectrum is calculated from the square of time-waveform x. The envelope is then the absolute value of the analytic signal, computed using the Hilbert transform, denoted $H\{\cdot\}$, as shown in Equation B.2.

$$Envelope = |x(t)^2 + jH\{x(t)^2\}| \tag{B.2}$$

For the AE measurements, an approximation of the analytic signal is created by analog pre-processing as described in Section B.2.4.

B.5 Discussion and Conclusion

In this paper, 3 different CM systems utilizing accelerometer and AE transducers have been applied to an offshore drilling machine. A healthy bearing was used as reference for a worn bearing, which then was tested at two stages of declining health.

Despite low utilization of the axial load capacity, it was possible to detect an increase in RMS of the AE signal that corresponds to the change in bearing health. The fact that

vibration measurement systems gave higher RMS values for the healthy bearing simply highlights the need for comparable operating conditions when trending. The process of replacing the healthy bearings with the used one, implied complete disassembly of the machine, effectively changing the basis of comparison. The consistent increase in AE RMS makes this CI more promising. AE RMS measurements were also less affected by rotational speed, which is an advantage for machines under varying operating conditions. AE measurements also showed an ability to detect some modulation frequencies not visible in the acceleration spectrum. In particular, a frequency which is assumed to be a lubrication pump was detected, in spite of a long signal transmission path. In the authors' opinion, this illustrates that measurement systems using AE transducers can complement vibration based systems.

The observed failure mode lacks a dominant localized fault, which results in a lack of periodic impacts. Hence, methods based on detection of bearing fundamental frequencies were ineffective. Artificially induced indentations were applied to a roller end, but a roller fault could not be identified from the measurements. Due to low utilization of bearing load capacity, the bearing is less prone to breakage of the oil film on the roller ends, which is a requirement for detection of roller end damage. While the overall increase in AE activity still corresponded to declining bearing health, it is unlikely that the increase is due to formation of new scratches. The results support AE as a CM technology for axial tapered roller bearings in drilling machines and other rotating machinery. The shown fault frequency identification capabilities combined with possible detection of scratch formation on roller ends motivates further research on the topic, in particular to capture and identify the failure mode propagation.

Acknowledgment

The research presented in this paper has received funding from the Norwegian Research Council, SFI Offshore Mechatronics, project number 237896.

References

[1] L. Jeffrey. Noble 2012 Analyst & Investor Day Presentation, 2012. Presented: 24.05.2012, Houston, Texas, USA. Date obtained: 31.08.2016.

[2] E. Bechhoefer, R. Schlanbusch, and T. I. Waag. Techniques for Large, Slow Bearing Fault Detection. *International Journal of Prognostics and Health Management*, 7(1):1–12, 2016.

[3] T. Yoshioka and T. Fujiwara. A New Acoustic Emission Source Locating System for the Study of Rolling Contact Fatigue. *Wear*, 81(1):183–186, 1982. doi:10.1016/0043-1648(82)90314-3.

[4] T. Yoshioka and T. Fujiwara. Application of Acoustic Emission Technique to Detection of Rolling Bearing Failure. *American society of mechanical engineers*, 14(1):55–76, 1984.

[5] B. V. Hecke, J. Yoon, and D. He. Low Speed Bearing Fault Diagnosis Using Acoustic Emission Sensors. *Applied Acoustics*, 105:35–44, 2016. doi:10.1016/j.apacoust.2015.10.028.

[6] M. He, D. He, and E. Bechhoefer. Using Deep Learning Based Approaches for Bearing Fault Diagnosis With AE Sensors. In *Annual Conference of the Prognostics and Health Management Society*, pages 1–10, 2016.

[7] R. B. Randall and J. Antoni. Rolling Element Bearing Diagnostics - A Tutorial. *Mechanical Systems and Signal Processing*, 25(2):485–520, 2011. doi:10.1016/j.ymssp.2010.07.017.

[8] J. Antoni. Cyclic Spectral Analysis of Rolling-element Bearing Signals: Facts and Fictions. *Journal of Sound and Vibration*, 304(3-5):497–529, 2007. doi:10.1016/j.jsv.2007.02.029.

[9] J. Antoni. Cyclostationarity by Examples. *Mechanical Systems and Signal Processing*, 23(4):987–1036, 2009. doi:10.1016/j.ymssp.2008.10.010.

[10] B. Kilundu, X. Chiementin, J. Duez, and D. Mba. Cyclostationarity of Acoustic Emissions (ae) for Monitoring Bearing Defects. *Mechanical Systems and Signal Processing*, pages 2061–2072, 2011. doi:10.1016/j.ymssp.2011.01.020.

[11] Y. Qu, E. Bechhoefer, D. He, and J. Zhu. A New Acoustic Emission Sensor Based Gear Fault Detection Approach. *International Journal of Prognostics and Health Management*, 4:32–45, 2013.

[12] J. Antoni. Fast Computation of the Kurtogram for the Detection of Transient Faults. *Mechanical Systems and Signal Processing*, 21(1):108–124, 2007. doi:10.1016/j.ymssp.2005.12.002.

Paper C

Rib-Roller Wear in Tapered Rolling Element Bearings: Analysis and Development of Test Rig for Condition Monitoring

Martin Hemmer, Kjell G. Robbersmyr, Tor I. Waag, Rolf Albrigtsen, Torfinn Pedersen, Thomas J. J. Meyer and Chloë Vercammen

This paper has been presented and accepted for publication as:

Martin Hemmer, Kjell G. Robbersmyr, Tor I. Waag, Rolf Albrigtsen, Torfinn Pedersen, Thomas J. J. Meyer and Chloë Vercammen. Rib-Roller Wear in Tapered Rolling Element Bearings: Analysis and Development of Test Rig for Condition Monitoring. *Engineering Assets and Public Infrastructures in the Age of Digitalization.* Springer, in press. Presented at: *13th World Congress on Engineering Asset Management.* Stavanger, Norway

Rib-Roller Wear in Tapered Rolling Element Bearings: Analysis and Development of Test Rig for Condition Monitoring

Martin Hemmer*,**, Kjell G. Robbersmyr*, Tor I. Waag***, Rolf Albrigtsen****, Torfinn Pedersen****, Thomas J.J. Meyer*** and Chloë Vercammen*****

*University of Agder
Department of Engineering Sciences
Jon Lilletunsvei 9, N-4879 Grimstad, Norway

**MHWirth AS
Butangen 21, N-4639 Kristiansand, Norway

***Teknova AS
Tordenskjolds gate 9, N-4612 Kristiansand, Norway

****Macgregor Norway AS
gate 9, N-4612 Kristiansand, Norway

*****ArianeGroup
Rue de Touban, 33185 Le Haillan, France

Abstract – Rolling Element Bearings (REBs) are present in virtually all machines with moving or rotating parts, and are vital for proper performance and safe operation. Condition Monitoring (CM) of bearings often receive particular interest, as this component group rarely reach design lifetime and hence is responsible for unplanned machine downtime. Unplanned maintenance can represent a large cost which motivates development of improved CM methods for implementation of advanced maintenance regimes. Based on observations of a used bearing from an offshore drilling machine, wear on roller ends in the ribroller contact area was identified as an area of interest for future research. A test rig for creating and observing accelerated roller end damage is developed, intended for use with vibration and Acoustic Emission (AE) sensors. In addition to normal continuous rotation of the bearing, the test rig is also designed with performing oscillation motion tests in mind. This mode of operation is of interest to manufacturers and end users of cranes and winches with heave compensation. Plans and challenges for future work are also discussed, in conjunction with the experimental setup.

C.1 Introduction

Dynamic mechanical systems of all types depend on bearings; a basic machine element that both allows and constrains movement by transferring loads to support structure while minimizing friction for the desired rotation or translation. As the diversity of applications is great, so is the number of bearing types and sizes. Given the widespread use and importance, bearing failure is also a large contributor to reduced asset availability. Under nominal operating conditions, fatigue will be the main failure mode of bearings. However, degradation and contamination of the lubricant can create wear that changes the internal geometry of the bearing [1], leading to premature failure and unplanned maintenance. This paper focuses on roller end wear in tapered REBs under axial load, rooted in a previous case study of an offshore drilling machine. The machine was taken out of operation for maintenance, and available for testing using both new and worn bearings. A schematic drawing of an axial spherical tapered roller bearing is seen in Figure . Because of the tapered roller design, a rib is necessary to keep the rollers in place, applying a seating force to the roller end. The rib-roller contact area is separated by an oil film, and the mating surfaces are sliding relative to each other. A thorough dynamic analysis of tapered REBs is given by [2].

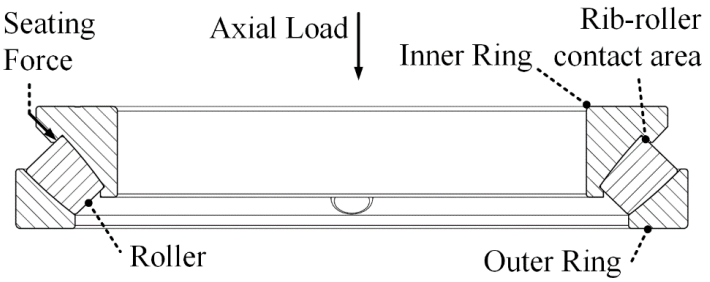


Figure C.1: Schematic drawing of an axial spherical tapered roller bearing

Tapered REBs are sensitive to proper lubrication of the rib-roller contact area, and lubrication starvation is a root cause of failure for bearing wear. In extreme cases, wear can lead to bearing seizure and catastrophic failure [3]. Ideally, online lubricant monitoring systems could detect wear particles at an early stage and prevent breakdowns. While such systems exist [4, 5] this approach has not replaced periodic sampling and analysis of the lubricant. Due to the inherent lag in this process, damage can occur and progress unnoticed. Alternative methods for online detection of wear will reduce the interval between detectable wear and maintenance action. Research of such methods is the main motivation for the development of this test rig.

Vibration monitoring using accelerometers is the current industry practice, and a range of well-established methods exists for fault detection [6] and health assessment

[7]. However, these methods are more suitable for detection of repetitive transients from localized damage. Roller end wear differs from localized faults in the load path, as scratch formation is not assumed to be related to the shaft frequency. This indicates that CM methods based on detection of repetitive transients at bearing defect frequencies will be ineffective.

It is known that AE has potential for monitoring of wear and sliding friction [8]. Investigation of AE activity in sliding friction has shown a proportional relation between sliding speed and amplitude, and frictional work and energy. Additionally, the frequency components associated with sliding was not present in the noise, which could simplify wear detection [9]. In the early stages of degradation, formation of scratches should therefore result in acoustic activity. AE has also been used for detection of wear in metal cutting tools [10]. As wear progresses, local changes in internal geometry and clearance can lead to increased mechanical vibration detectable by accelerometers.

C.2 Roller End Wear

In a previous study by the authors [11], one new and one previously used main bearing from an offshore drilling machine was run in a controlled environment. Scoring, a type of abrasive wear, was observed on all roller ends during visual inspection of the bearing, as seen in Figure C.2. However, any formation of new scratches could be not detected in the collected data. It is assumed that each scratch originates from a particle in the lubricant which enters the rib-roller contact area, then sliding across the roller surface, stationary with respect to the rib. If the particle size is greater than the oil film thickness, mechanical contact occurs, and a scratch is formed. The abrasion then contributes to further contamination of the oil. As this process is not expected to be periodic, any detection method based on detecting the presence of bearing fault frequencies will not be effective. Instead, an event-based approach is assumed to be more effective. If one can successfully detect the formation of new scratches, both the total number of events and changes in mean time between events are potential measures of bearing health and fault propagation. Tapered REBs are also sensitive to scuffing, a form of adhesive wear. As scuffing can be caused by excessive surface roughness, it is a possibility that the roller end scoring develops to scuffing.

C.2.1 Roller End Scoring Model

The scoring occurs as distinctly shaped arcs with varying radius. A simplified 2D visualization is given in Figure C.3. Let A be a circle of radius a, representing a roller end with a reference coordinate system centred on the roller end, while circle B of radius b

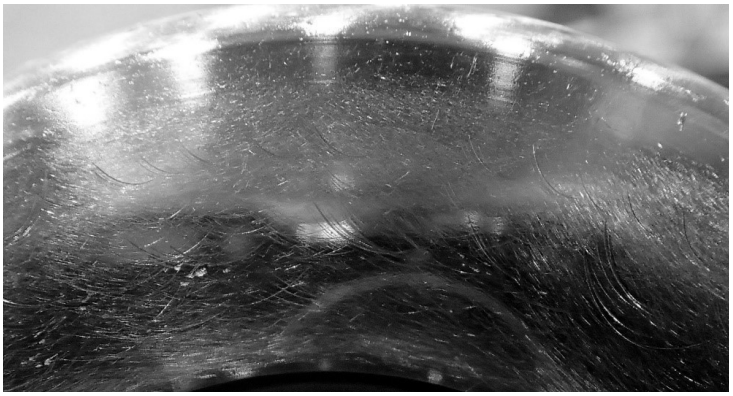


Figure C.2: Scoring on roller end of tapered roller bearing

represents a path of rolling contact with the bearing race. Then, consider B rolling on the circumference of A with a particle in point P, fixed at a distance h from the centre of B. The parameter θ is defined as the angle between the x-axis and the line between circle centres.

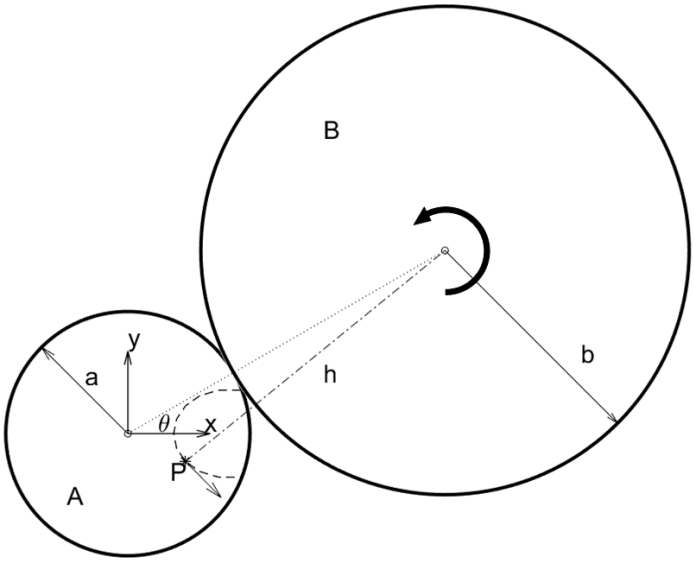


Figure C.3: Trajectory of P on A forms an epitrochoid.

Let γ represent the following relation:

$$\gamma = \frac{a+b}{b} \tag{C.1}$$

The position vector of P as a function of θ is given in Equation C.3:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\cos\gamma\theta \\ \sin\theta & -\sin\gamma\theta \end{bmatrix} \begin{bmatrix} b\gamma \\ h \end{bmatrix} \tag{C.2}$$

If θ is a function of time, the instantaneous velocity vector of P, indicated by an arrow, is given in Equation 3:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -\sin\theta & \sin\gamma\theta \\ \cos\theta & -\cos\gamma\theta \end{bmatrix} \begin{bmatrix} b\gamma\dot{\theta} \\ h\gamma\dot{\theta} \end{bmatrix} \tag{C.3}$$

Note that the sliding velocity is a harmonic function of θ, resulting in variable sliding velocity for a constant rpm. In a similar application, monitoring piston seals in a combustion engine, AE activity showed a linear relationship to sliding velocity [12]. As seen by this simplified model, the same behaviour could be utilized in establishing a fault signature for roller end wear detection in bearings. With a known fault signature, it is possible to determine the presence of a fault in recorded data through hypothesis testing, providing decision support for scheduling of maintenance.

C.3 Test Rig

To the authors' knowledge, no test equipment for emulating roller end damage in a controlled environment during operation exists, motivating the development. Figure C.4 shows a sketch of the test rig (left) and the bearing test unit (right).

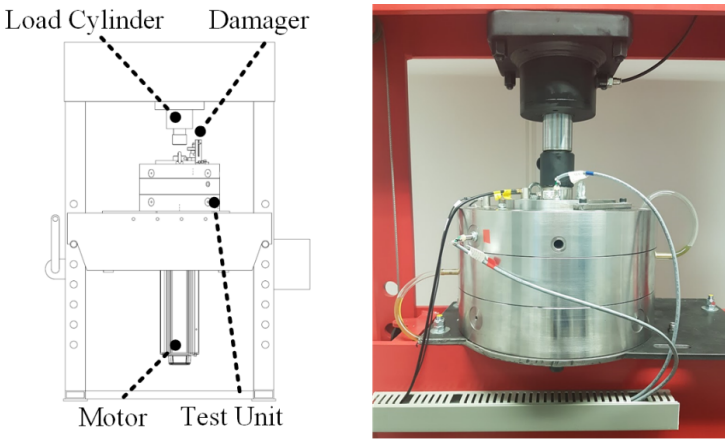


Figure C.4: Bearing test rig sketch (left), Bearing test unit (right)

The test bearing (designation 29230) is contained inside the bearing test unit, attached to the motor through a flexible coupling. Axial load is applied by a hydraulic load cylinder. A second, higher capacity bearing (designation 29336) transfers the load to the support structure, removing any axial load from the motor shaft. The test bearing is attached using an adapter, allowing for easy replacement and testing of other bearing types and sizes, up to 240 mm outer diameter. The load bearing and test bearing runs in two separate chambers, allowing different lubricants and preventing cross-contamination. The lubricant can also be circulated for monitoring purposes. The test rig is developed with two main test types in mind: artificial roller end damage and oscillating motion testing, described more in detail below.

C.3.1 Artificial Roller End Damage Test

The goal of artificial roller end damage end testing is to identify any characteristics of the failure mode. What separates this test rig from most other machines, is the capability to create damage while the bearing is running in the machine, emulating a particle in the lubricant. This allows control of when the scratch is made, as opposed to having loose particles in the lubricant. Figure C.5 shows the mechanism for applying the damage. A hardened needle is inserted through a predrilled hole to scratch the roller. By varying the hole location and shape of the tip, analogous to changing the distance h in Figure C.5, the scratch location on the roller can be modified.

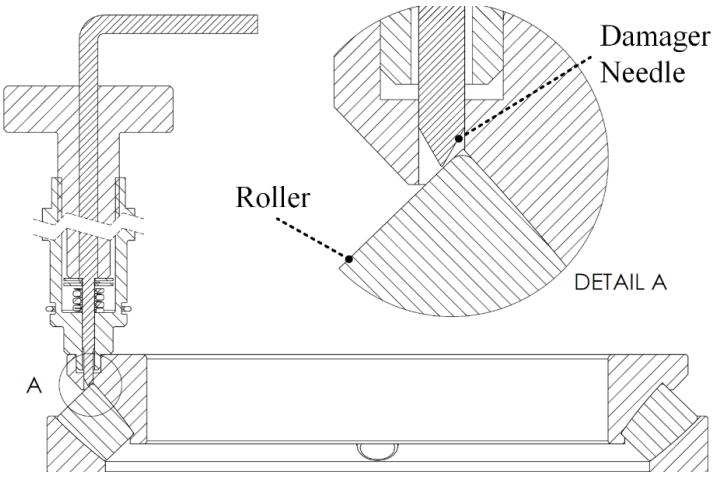


Figure C.5: Cross-section of artificial damage mechanism

C.3.2 Oscillating Motion Test

Heave-compensating winches is an example of equipment experiencing oscillating motion in parts of the operating cycle. Oscillating motion results in uneven distribution of stress on the bearing components, which is known to reduce fatigue lifetime [13]. Oscillating motion is also known to cause fretting corrosion. Degradation of high loaded oscillating bearings was investigated by [14]. For the context of this paper, oscillation implies that the rolling elements perform incomplete revolution before changing direction. The motor is a synchronous permanent magnet AC (PMAC) motor, capable of precise control of shaft torque and rotation at low speed.

C.3.3 Instrumentation

The test rig is prepared for accommodating a range of sensors. It is currently equipped with an axial load transducer, accelerometers and AE transducers. Additionally, the motor has an integrated encoder, making position, velocity, acceleration available from

Table C.1: System overview

Drive System	Loading System	Acquisition System
PMAC Motor	Hydraulic	NI 9232 - 2 x 3 ch., 24 bit AI
Direct drive	50 ton load capacity	NI 9205 - 16 ch, 16 bit AI
48 Nm rated torque	Manual control	NI 9401 - 8 ch, DI/O
3000 rpm limit	Force measurement	VFD Ethernet communication
		MISTRAS PCI-2 (AE specific)

Table C.2: Sensor specification

Sensor Type	Model
Accelerometer	TE Connectivity 805M1-0020-01
AE sensor	Physical Acoustics R15a
Force transducer	TECSIS F6210 500 kN

the Variable Frequency Drive (VFD) over Ethernet. Data is recorded using a National Instruments compactDAQ equipped with Analog Input (AI) and Digital Input/Output (DI/O) cards. An overview of the drive system, loading system and acquisition system is give in Table 1.

Table 2 gives an overview of the sensors currently in use on the machine.

Table 2 gives an overview of the sensors currently in use on the machine.

C.4 Discussion

This paper discusses roller end wear, in particular scoring on roller ends. The characteristic scoring shape is described mathematically, and a test rig is designed for replicating this failure mode through artificial roller end damage testing and oscillating motion testing. Through these tests, it may be possible to identify fault signatures for use in online condition monitoring systems. A data-driven approach allows the use of statistical tools for detection algorithms and bearing health assessment, providing decision-making assistance for maintenance actions. Compared to offline lubricant analysis, this gives more time to plan and execute maintenance.

Further, experiments with different needle shapes, speeds, lubricants and loads are possible. With minor modifications, the test rig can also accommodate different test bearings. While the test rig is prepared for vibration and AE measurements, other transducers can be retrofitted, including but not limited to motor current sensors, online lubrication

monitoring and electrical methods for metal-to-metal contact detection.

Acknowledgement

The research presented in this paper has received funding from the Norwegian Research Council, SFI Offshore Mechatronics, project number 237896.

References

[1] B. Fitzsimmons and H. D. Clevenger. Contaminated Lubricants and Tapered Roller Bearing Wear. *ASLE Transactions*, 20(2):97–107, 1977. doi:10.1080/05698197708982822.

[2] S. Crep, I. Bercea, and N. Mitu. A Dynamic Analysis of Tapered Roller Bearing Under Fully Flooded Conditions Part 1: Theoretical Formulation. *WEAR*, 188, 1995.

[3] R. J. Parker. Large-bore Tapered-roller Bearing Performance and Endurance to 2.4 Million Dn Apparatus and Procedure, 1983. Obtained date: 19.02.2018. URL: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19830011863.pdf.

[4] J. Zhu, J. M. Yoon, D. He, and E. Bechhoefer. Online Particle-contaminated Lubrication Oil Condition Monitoring and Remaining Useful Life Prediction for Wind Turbines. *Wind Energy*, 18(6):1131–1149, 2015. doi:10.1002/we.1746.

[5] X. Zhu, C. Zhong, and J. Zhe. Lubricating Oil Conditioning Sensors for Online Machine Health Monitoring – A Review. *Tribology International*, 109:473–484, 2017. doi:10.1016/J.TRIBOINT.2017.01.015.

[6] R. B. Randall and J. Antoni. Rolling Element Bearing Diagnostics - A Tutorial. *Mechanical Systems and Signal Processing*, 25(2):485–520, 2011. doi:10.1016/j.ymssp.2010.07.017.

[7] D. Wang, K. L. Tsui, and Q. Miao. Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators. *IEEE Access*, 6:665–676, 2018. doi:10.1109/ACCESS.2017.2774261.

[8] C. Jiaa and D. Dornfeld. Experimental Studies of Sliding Friction and Wear via Acoustic Emission Signal Analysis. *Wear*, 139(2):403–424, 1990. doi:10.1016/0043-1648(90)90059-J.

[9] H. Taura and K. Nakayama. Behavior of Acoustic Emissions at the Onset of Sliding Friction. *Tribiology International*, 123:155–160, 2018. doi:10.1016/j.triboint.2018.01.025.

[10] M. S. H. Bhuiyan, I. A. Choudhury, and M. Dahari. Monitoring the Tool Wear, Surface Roughness and Chip Formation Occurrences Using Multiple Sensors in Turning. *Journal of Manufacturing Systems*, 33(4):476–487, 2014. doi:10.1016/j.jmsy.2014.04.005.

[11] M. Hemmer and T. I. Waag. A Comparison of Acoustic Emission and Vibration Measurements for Condition Monitoring of an Offshore Drilling Machine. In *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, pages 278–285, 2017.

[12] R. M. Douglas, J. A. Steel, and R. L. Reuben. A Study of the Tribological Behaviour of Piston Ring/cylinder Liner Interaction in Diesel Engines using Acoustic Emission. *Tribology International*, 39(12):1634–1642, 2006. doi:10.1016/j.triboint.2006.01.005.

[13] L. Houpert. Bearing Life Calculation in Oscillatory Applications. *Tribology Transactions*, 42(1):136–143, 1999. doi:10.1080/10402009908982200.

[14] F. Massi, N. Bouscharain, S. Milana, G. Le Jeune, Y. Maheo, and Y. Berthier. Degradation of High Loaded Oscillating Bearings: Numerical Analysis and Comparison With Experimental Observations. *Wear*, 317(1-2):141–152, 2014. doi:10.1016/J.WEAR.2014.06.004.

Paper D

Fault Classification of Axial and Radial Roller Barings Using Transfer Learning through Pretrained Convolutional Neural Network

Martin Hemmer, Huynh Van Khang, Kjell G. Robbersmyr and Thomas J. J. Meyer

This paper has been published as:

Martin Hemmer, Huynh V. Khang, Kjell G. Robbersmyr, Tor I . Waag and Thomas J. J. Meyer. Fault Classification of Axial and Radial Roller Bearings Using Transfer Learning through a Pretrained Convolutional Neural Network. *Designs*, 2(4), 56, 2018. doi: 10.3390/designs2040056.

Fault Classification of Axial and Radial Roller Bearings Using Transfer Learning through a Pretrained Convolutional Neural Network

Martin Hemmer*,**, Huynh Van Khang*, Kjell G. Robbersmyr*, Tor I. Waag*** and Thomas J.J. Meyer***

*University of Agder
Department of Engineering Sciences
Jon Lilletunsvei 9, N-4879 Grimstad, Norway

**MHWirth AS AS
Butangen 21, N-4639 Kristiansand, Norway

***NORCE Norwegian Research Centre AS
Jon Lilletuns vei 9 H, N-4877 Grimstad, Norway

Abstract – Detecting bearing faults is very important in preventing non-scheduled shutdowns, catastrophic failures, and production losses. Localized faults on bearings are normally detected based on characteristic frequencies associated with faults in time and frequency spectra. However, missing such characteristic frequency harmonics in a spectrum does not guarantee that a bearing is healthy, or noise might produce harmonics at characteristic frequencies in the healthy case. Further, some defects on roller bearings could not produce characteristic frequencies. To avoid misclassification, bearing defects can be detected via machine learning algorithms, namely convolutional neural network (CNN), support vector machine (SVM), and sparse autoencoder-based SVM (SAE-SVM). Within this framework, three fault classifiers based on CNN, SVM, and SAE-SVM utilizing transfer learning are proposed. Transfer of knowledge is achieved by extracting features from a CNN pretrained on data from the imageNet database to classify faults in roller bearings. The effectiveness of the proposed method is investigated based on vibration and acoustic emission signal datasets from roller bearings with artificial damage. Finally, the accuracy and robustness of the fault classifiers are evaluated at different amounts of noise and training data.

D.1 Introduction

Failure on rolling bearings is one of the most frequent system failures, resulting in huge losses of productivity in drivetrains installed in remote and harsh environment areas. Defects on bearings contribute to over 40% of faults in rotating machinery [1]. If a bearing fault is well predicted, the risk of long-term system breakdown can be prevented, and a replacement of the faulty bearing will be done at the right time. Faulty bearings can be detected by analyzing current, vibration, or acoustic emission signals. Current signature analysis can be useful to detect limited faults on bearings, which need to be connected to a shaft driven by an electric motor. Vibration analysis is preferred to monitor conditions of bearings in most mechanical systems, where accelerometers are usually installed in place. In critical applications, acoustic emission signals can be used to detect bearing faults at an early stage due to its higher sensitivity and convenient installation without being involved in the system.

Processing data and understanding faulty features in vibration and acoustic emission analysis need skilled manpower with advanced knowledge of bearing faults [2]. Vibration signals associated with faults typically originates from high-frequency resonance in the housing structure excited by low-frequency impacts related to the contact between a fault and other bearing components. The accelerometers installed on the bearing housing are very sensitive to any forces generated in a system. This makes collected signals from the accelerometers very complicated due to the interference of noise. The complexity of the output signals from the collected acoustic emission sensors can be even worse due to its higher sensitivity and is worse in highly disturbed environments.

To predict bearing faults based on the mentioned signals, common processing, i.e., fast Fourier transform (FFT), short-time Fourier transform (STFT), and continuous Wavelet transform (CWT) in [3], and wavelet transform (WT) with kurtosis [4], could be used to detect signals associated with the faults. Such a signal processing technique is useful to observe characteristic frequencies in time and frequency representations. However, missing a harmonic in a spectrum or the appearance of harmonics at characteristic frequencies due to noise might cause misclassification. Further, the effectiveness of this conventional approach strongly depends on manpower skill, training, and relevant experience.

Unlike conventional bearing faults such as spalling on races or rolling elements, roller-end wear in axial roller bearings might not produce periodic harmonic components associated with faults. In a previous study by the authors [5], scratches on an axial bearing were observed on roller ends of a tapered axial roller bearing in an offshore drilling machine, but no particular bearing frequency was steadily detected in the spectrum. However, acoustic emission data showed an energy increase with higher damage severities. Detecting bearing defects without a characteristic frequency or predefined knowledge of the fault signature

is a big challenge in fault diagnosis.

To address the mentioned challenges, an automatic system for fault detection and classification applicable to both vibration and acoustic emission signals can reduce the manpower dependence and time consumption for condition monitoring of the roller bearing in industry. As argued in [6], increasing the performance of the detection system might be more important than looking for highly reliable features. Model-based, data-driven, or hybrid algorithms are common in automatic fault diagnosis [7, 8, 9, 10]. The model-based diagnosis needs both a detailed physical model of the system and its accurate parameters, which are very hard to obtain in reality. Without a physical model [11], the data-driven approach using statistical or machine learning algorithms is attractive for an automatic diagnosis system. To enhance the accuracy of fault detection, statistics methods should be based on the frequency spectrum to reduce false and missing alarms [12]. Alternatively, machine learning methods, namely support vector machine (SVM) [13, 14], decision tree (DT) [15], and various neural network architectures [16, 17] combined with advanced signal processing can be used to find the complex relations on the feature space by using predefined time-frequency features, being based on fault characteristic frequencies. However, without the characteristic frequencies, the mentioned methods have great difficulty in classifying bearing faults [18].

This work focuses on developing a simple automatic fault diagnosis method for roller bearings, requiring less human intervention or domain knowledge of features. Using transfer learning (TL) allows us to reduce the time and complexity of generating features for fault classification. Further, TL is very helpful for a bearing fault diagnosis if the available data for training and validation are limited in industry [16]. Within this work, a pretrained version of the well-known AlexNet convolutional neural network (CNN) architecture [19] is applied to CWT spectrograms of vibration and acoustic emission signals. Then, the CNN is either fine-tuned to perform classification directly or to extract features used to train and validate two classifiers using SVM and sparse autoencoder-based SVM (SAE-SVM). The robustness of the proposed method is tested at different signal-to-noise ratio (SNR) levels.

The remainder of the article is organized as follows. In Section D.2, the proposed methods are detailed. In Section D.3, the experimental setup and preprocessing are presented. In Section D.4, the results of the fault detection and classification are presented. Further, the discussion of the presented results is detailed in Section D.5. Finally, the paper is concluded in Section D.6.

D.2 The Proposed Method

A diagram of the proposed fault classification is shown in Figure D.1. The vibration signals need to be converted to images as pixels or a matrix ($227 \times 227 \times 3$, height by width by depth) before feeding them to the AlexNet architecture. Within this process, images or spectrograms of vibration signals are formed by three channels (red, green, and blue (RGB)), resulting in a depth of three. The spectrograms go though several convolutional layers, acting as learnable filters to detect the presence of specific features from the input, and produce matrices ($M \times M \times L$) with $M < N$ and L-size filters. This work utilizes a pretrained version of the AlexNet architecture [19], obtained from the Berkeley Vision and Learning Center caffe repository on GitHub [20], through the MATLAB Deep Learning Toolbox. AlexNet consists of five convolutional (conv1–conv5) and three fully-connected (fc6–fc8) layers, as illustrated in Figure D.2, in which numbers outside the boxes illustrate the dimension in each layer and numbers inside the boxes indicate the filter sizes of the convolutions. The architecture uses rectified linear units (ReLU) as activation functions and dropout layers to prevent overfitting. Three classifiers will be described in this section: CNN, SVM, and SAE-SVM. For the two latter, features from the pretrained network will be extracted at both layer fc6 and fc7 and used to train two instances of the classifiers.

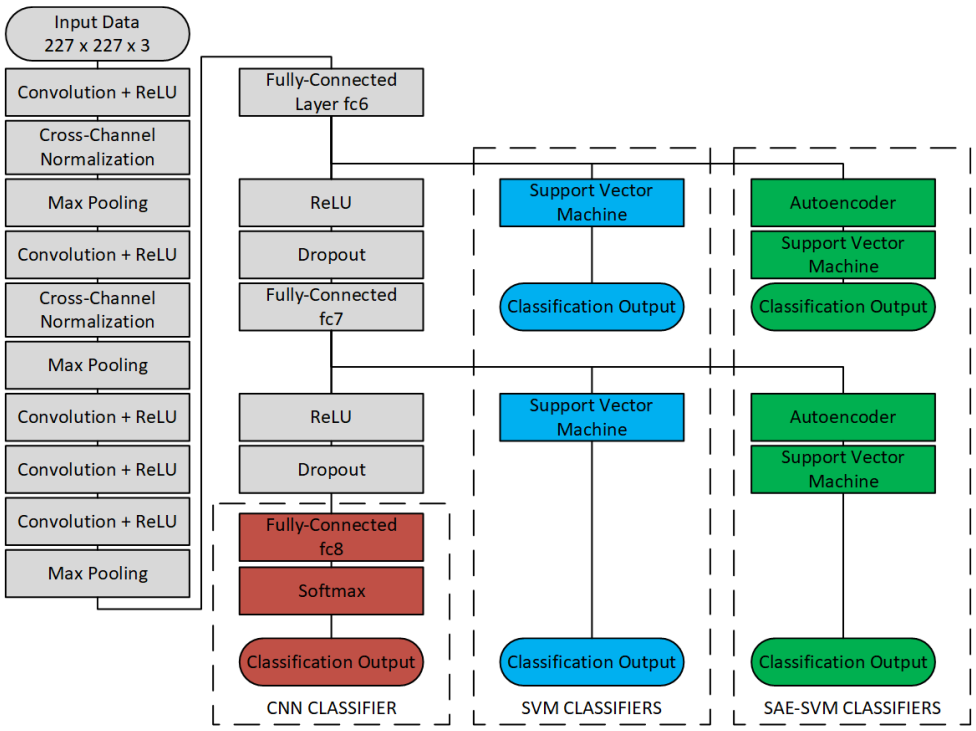


Figure D.1: Methodology visualization. SAE, sparse autoencoder.

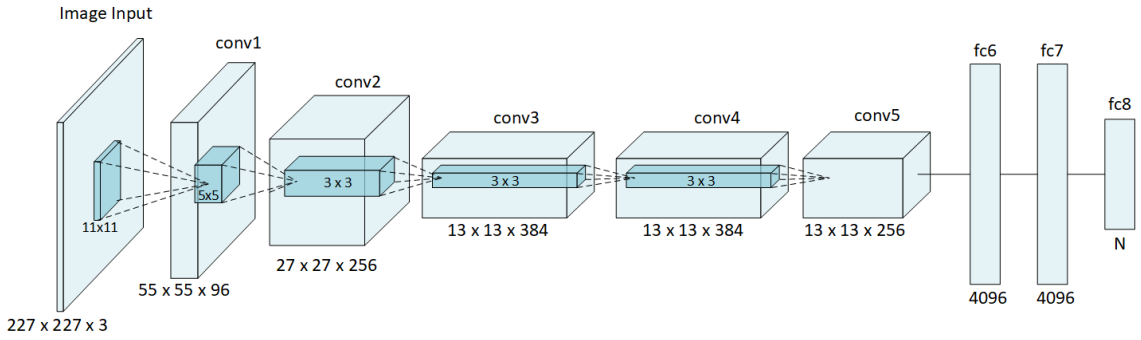


Figure D.2: Simplified illustration of the AlexNet architecture.

D.2.1 Convolutional Neural Network-based Fault Classifiers or Retrained CNN

Given the performance of CNNs in image classification, they can be fully trained to classify patterns or spectrograms generated by CWT correctly [21]. However, training a network from scratch is very time consuming, requiring GPU programming, tuning of hyperparameters, etc. The first classifier uses transfer learning through fine-tuning of a pretrained CNN, e.g., the AlexNet architecture, to reduce the complexity of the training process. Instead of retraining the complete network from scratch, only the final classification layer is replaced, which maintains most of the already gathered information from the training on the imageNet database. To make sure most of the pretrained weights are maintained, the learning rate for the classification layer fc8 is increased to 20-times the overall learning rate. After replacing fc8 with a fully-connected layer of size 2 (equal to the number of classes), the network has approximately 60 million trainable parameters. Because of pre-training, fine-tuning the CNN to classify new data can then be done by using a smaller dataset of CWT spectrograms. This adaptation of a pretrained network is time-saving and very helpful for inexperienced users. Table D.1 describes the parameter setting for the CNN-classifier.

Table D.1: CNN settings.

Solver	Stochastic Gradient Descent with Momentum (SGDM)
Training epochs	50
Initial learning Rate, overall	10^{-3}
Initial learning Rate, fc8	0.02
Learning Rate Drop Factor	0.2
Learning Rate Drop Period	10

D.2.2 Support Vector Machine-based Fault Classifier

Support vector machines are supervised learning models for data classification. Given a set of training data of dimension K, the algorithm finds the hyperplane, a subset of the feature space of dimension $K-1$, which provides the best separation between classes in the training data [14]. This is a quadratic optimization problem, which also removes the local minima being present in neural networks [22]. Each input image generates a set of features at each layer throughout the network. Instead of retraining the final classification layers like in Section D.2.1, this classifier extracts the features directly at a higher level. This method is built on on the assumption that filters in the convolutional layers are trained to detect features that are also suited to discriminate features associated with the bearing faults.

The generated feature space from AlexNet has dimension $N = 4096$ for both layer fc6 and fc7. The objective of using data at fc6 or fc7 is to study whether the extra ReLU, dropout, and fully-connected layer from TL affects the accuracy of SVM classifications or not. By using the pretrained network to generate features, it is not necessary to design any features tailor-made to the application. Instead, the SVM is trained as set in Table D.2.2.

Table D.2: SVM settings.

Feature Layers	fc6 and fc7
Kernel	Linear
Standardization	Unit mean and variance
Solver	Iterative Single Data Algorithm (ISDA)
Outlier fraction	0.0

D.2.3 Sparse Autoencoder Combined with SVM Classifier

The autoencoder is designed to replicate its input at its output in an unsupervised fashion, which can be used for both unsupervised feature extraction and image denoising [23]. An autoencoder is basically a single fully-connected layer of size P, referred to as the hidden layer, that is trained to reconstruct its input by minimizing error over the training dataset. Labels are not considered during training. One could assume that not all these features are equally important or necessary in order to perform classification. However, computational burden is a product of feature space dimension, N, and hidden layer size, P. By using the features from fc6 or fc7 instead of the input image, N is reduced from 154,587 to 4096, dramatically reducing the computational burden. Denoising AEs and

sparse AEs are commonly used in literature. The denoising AEs are to partially corrupt input data and capture the original data removing noise, while the sparse AEs are to learn the features and structures within the input data. With a hidden layer size of $P = 100$, the SAE is used in this paper with a sparsity proportion of 0.05 to identify features from fc6 or fc7. The identified individual features are classified by SVM as described in Section D.2.2. Detailed settings of the SAE parameters are given in Table D.3.

Table D.3: SAE settings.

Hidden layer size	100
Training algorithm	Scaled conjugate gradient descent
Training epochs	2500
Encoder transfer function	Logistic sigmoid function
Decoder transfer function	Logistic sigmoid function
L_2 Weight Regularization	0.001
Loss function	Mean squared error with L_2 and sparsity regularization
Sparsity proportion	0.05

D.3 Experimental Setups and Datasets

The proposed method is evaluated using two separate datasets, consisting of time-waveform signals from two different bearing test rigs:

- Dataset 1: Vibration data from the NU220 ECP radial bearing with seeded faults.
- Dataset 2: Acoustic emission signals from an artificial roller end scratch generation from an in-house test setup for axial tapered roller bearings.

The two datasets represent the two different situations discussed in Section D.1. Dataset 1 contains periodic transients at bearing fault frequencies, while Dataset 2 contains aperiodic bursts of acoustic energy from scratch formation on the roller end.

D.3.1 Dataset 1: Radial Roller Bearing Test Rig

The radial bearing tests were conducted using a radial bearing test rig at the Institute for Machine Elements and Systems Engineering at RWTH Aachen University. A triaxial accelerometer was mounted on the bearing housing, giving two radial channels and one axial channel. The location is shown in Figure D.3. For this paper, the two radial channels were used in the analysis.

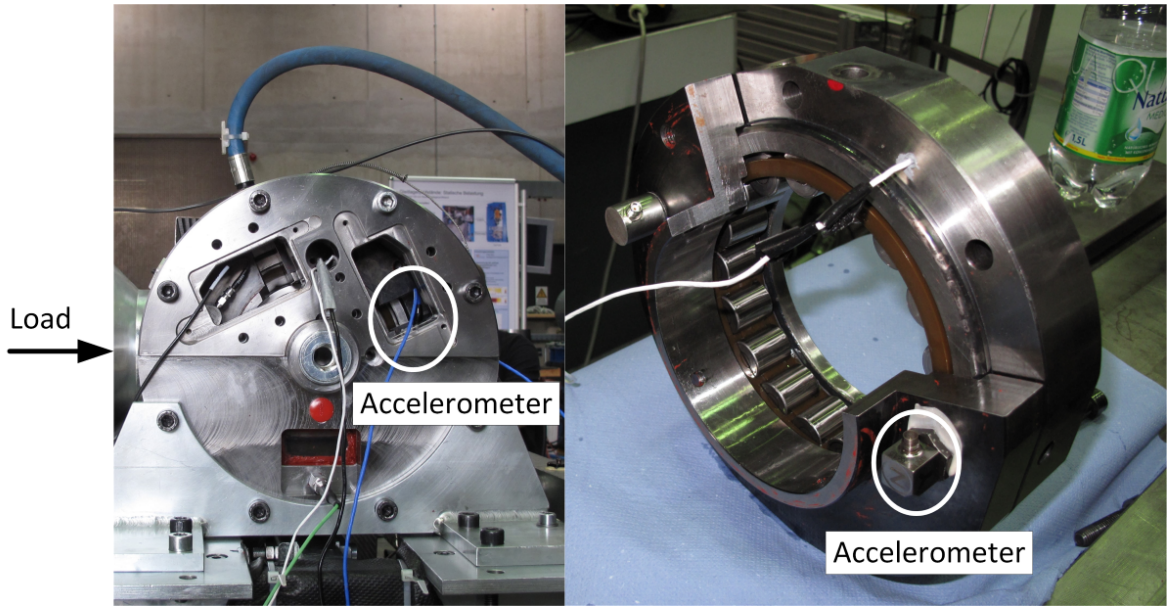


Figure D.3: Radial bearing test rig.

The test bearing, a radial bearing of type NU 220 ECP, was tested with two different fault types. Fault Type 1 (FT1) is a severe outer race (OR) fault, while Fault Type 2 (FT2) is a smaller rolling element (RE) fault. Data were classified as either healthy (HE) or damaged with subcategories of OR and RE faults. However, each bearing only had one type of damage or a single fault under test; thus, detecting faults in each bearing is a binary classification problem. The seeded OR and RE damages in the bearing can be seen in Figure D.4.

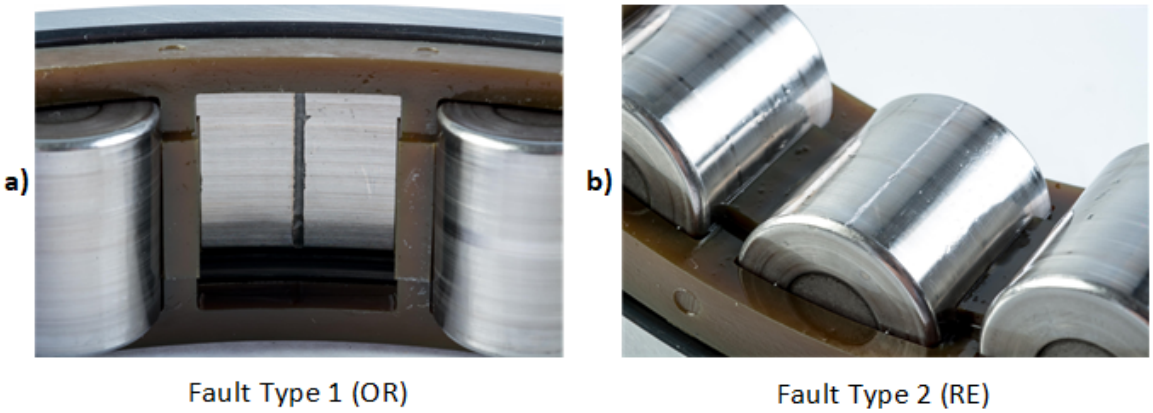


Figure D.4: Faults in the test bearings: (**a**) outer race (OR) fault and (**b**) rolling element (RE) fault.

Impact energy is reduced as the load decreases, making fault detection more challenging. According to the recommendation of the bearing manufacturer, the minimum radial load for the test bearing is 2.09 kN at 1000 rpm. Data chosen for analysis were collected

with a constant radial load of 5 kN to test the proposed method under the most difficult conditions while complying with the bearing operating specification. The test operating conditions are described in Table D.4, from which 560 CWT spectrograms were produced: 25% belonged to the damaged case, while 75% were in the healthy case. Twenty five percent of the dataset was reserved for validation of the algorithms, and the remaining 75% was used for training. Further, as shown in Table D.5, the proposed algorithms were also trained with 25% and 50% data.

Table D.4: Radial test data.

rpm / **Load (kN)**	100 / 5	200 / 5	300 / 5	400 / 5	500 / 5	750 / 5	1000 / 5	**Total**
No. of HE	60	60	60	60	60	60	60	420
No. of RE	10	10	10	10	10	10	10	70
No. of OR	10	10	10	10	10	10	10	70
Total	80	80	80	80	80	80	80	560

Table D.5: Radial training/validation data distribution.

		Training			**Validation**
Label	**Number of Images**	**75%**	**50%**	**25%**	**25%**
Healthy	420	315	210	105	105
Damage	140	105	70	35	35
Total	560	420	280	140	140

D.3.2 Dataset 2: Axial Roller Bearing Test Rig

The axial bearing test rig was designed and built at the University of Agder. A schematic drawing and picture of the test setup are shown in Figure D.5. This testbed was built based on an observations from an offshore drilling machine. It was found that wear occurred in the rib-roller area of a large, tapered rolling element bearing with a characteristic arc-shape. The developed test rig was to investigate the fault signature of such scratches occurring during operation. A spherical tapered rolling element bearing was placed inside an enclosed test unit containing a lubricant. The bearing was loaded using a hydraulic cylinder and rotated using a direct-drive permanent magnet synchronous motor. A damage mechanism allowed the roller to be scratched during operation.

The rib-roller contact area is prone to wear due to the relative sliding motion. Consider a simplified example where a particle large enough to break the oil film is stuck between the rib and roller end, stationary with respect to the rib. It can be shown that the trace

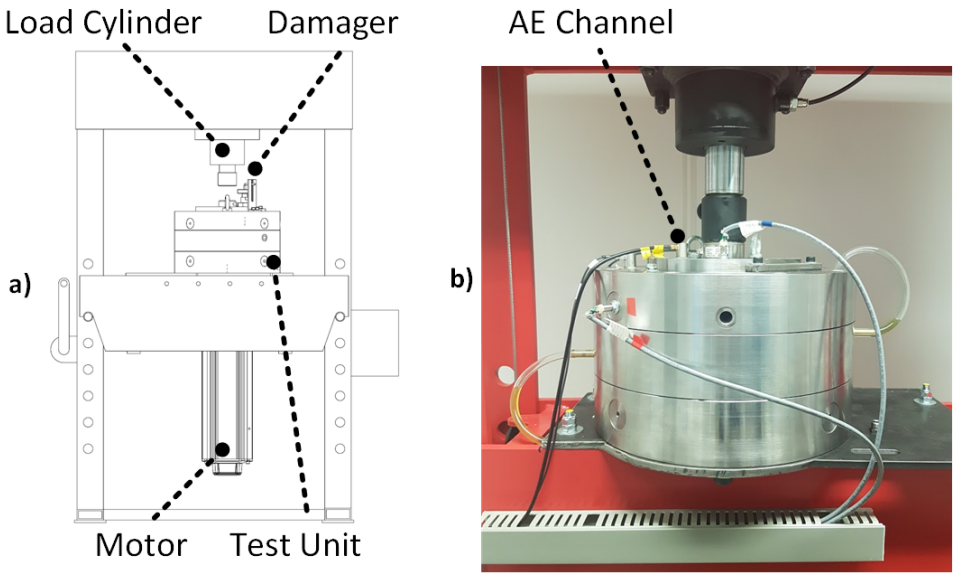


Figure D.5: (**a**) Schematic drawing of the in-house axial bearing test rig. (**b**) Acoustic emission sensor placement.

of the particle on the roller will resemble a segment of an epitrochoid as the roller passes, which complies with the observations from the offshore drilling machine. This type of damage differs from localized damages such as spalling. In the experiments on the axial bearing test rig, we were able to replicate this damage using a hardened needle, as shown in Figure D.6. To detect faults on the rolling element bearing, one channel of acoustic emission data, shown in Figure D.5, was used to collect time-series data (Dataset 2) for analysis.

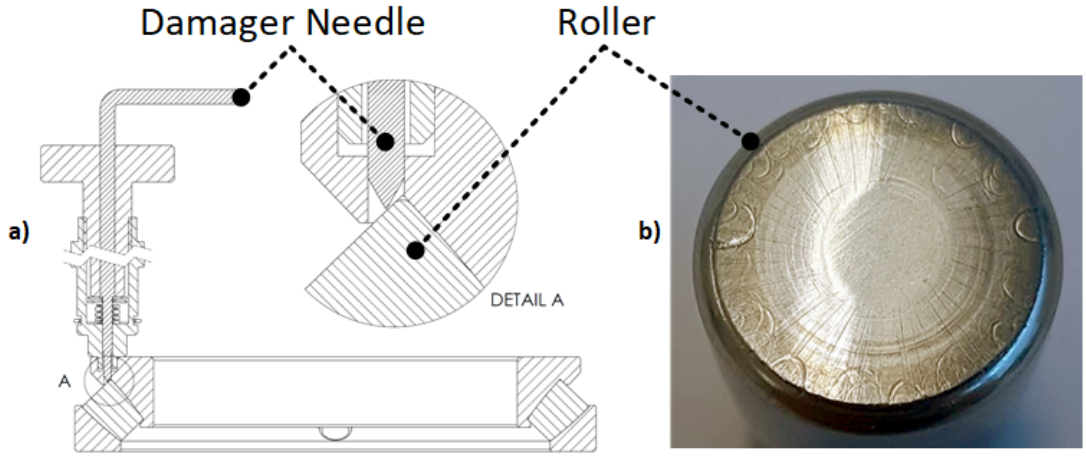


Figure D.6: (**a**) Schematic drawing of the damage mechanism. (**b**) Scratches on the axial bearing roller end.

Dataset 2 contains two labels: healthy (HE) and scratch (SC). A total of 600 images were generated, of which 200 were from the damaged class. Again, 25% of the data was

reserved for validation and the rest used for training. Table D.6 shows the number of images used for training and validation in the different cases. This dataset was collected at a very low speed of 1 rpm, as the current test setup did not allow consistent scratching at higher rotational speeds due to possible breakage of the needle. Axial load was kept constant at 50 kN during the tests.

Table D.6: Axial training data.

		Training			**Validation**
Label	**Number of Images**	**75%**	**50%**	**25%**	**25%**
Healthy	400	300	200	100	100
Damage	200	150	100	50	50
Total	600	450	300	150	150

D.3.3 Preprocessing

Collected vibration data are in time domain. If supplying the time domain signal directly to the CNN, the CNN pattern recognition is constrained by the 1-dimensional convolution [24]. To detect signals associated with faults in both the time and frequency domains, input data of the CNN should be converted to 2D spectrograms with 2D convolution. The AlexNet architecture deals with the 2D convolution, being fed with $227 \times 227 \times 3$ RGB images. Continuous wavelet transform (CWT) is used in the preprocessing to convert the collected vibration signals to time-frequency representations or spectrograms. Short-time Fourier transform (STFT) can be an alternative method, but requires an extra parameter. Selecting the best method to produce proper spectrograms is out of scope of this work. CWT is applied to both Datasets 1 and 2 in the following sections. To study the effect of training data on the performance of the proposed fault classifiers, namely CNN, SVM, and SAE-SVM, the training data were reduced by steps of 25% from 75%–25% while keeping the same amount of validation data of 25%, as shown in Tables D.5 and D.6. Analysis results on the datasets, from two different test setups of radial and one axial bearings will be described in more detail in Sections D.3.1 and D.3.2 below.

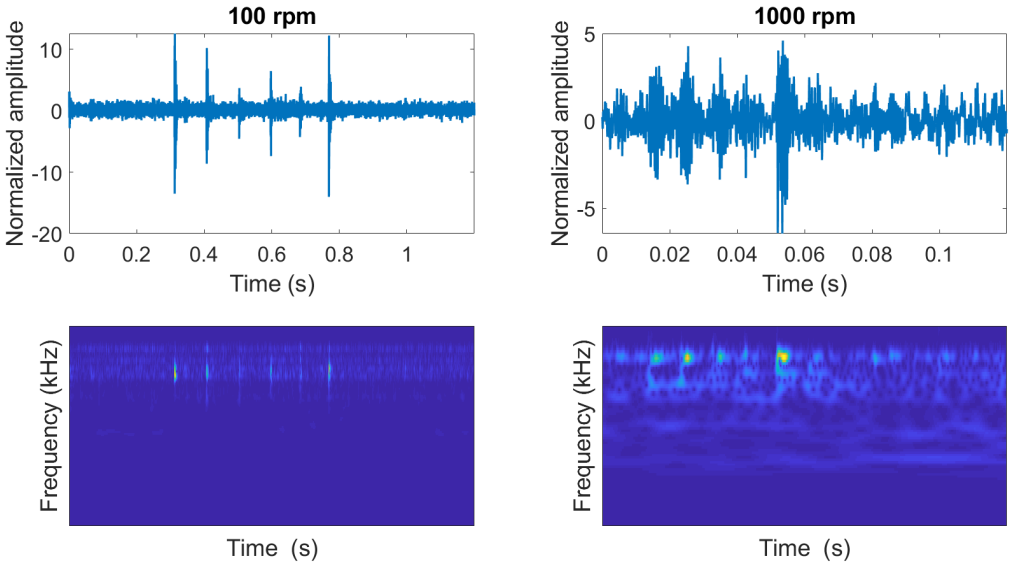
The data in Dataset 1 were collected in the time domain under different operating conditions, with the speed ranging from 100 rpm–1000 rpm. Normalization of root mean square (RMS) was performed to make time series more comparable. To increase the amount of training data, each time series was split into segments of two revolutions. This transforms the signal from the time domain to the order domain. The number of samples per revolution f_{order} was calculated as the sample frequency f_{sample} divided by the shaft

frequency f_{shaft}, as written in Equation (D.1). To test the robustness of the proposed method against noise, white Gaussian noise (WGN) was added to the time-waveforms in Dataset 1. The raw signal (SNR1) was considered as a reference for WGN power at other SNR levels.

$$f_{order} = \frac{f_{sample}}{f_{shaft}} \tag{D.1}$$

The number of samples in each segment N_{seg} was then calculated as in Equation (D.2), where n_{rev} is the number of revolutions to be included in the segment. $n_{rev} = 2$ was used in this work. The modification of the time scale for each segment is effectively a transformation to the order domain, which allows the spacing between transients to be kept constant at different speeds. Figure D.7 shows the difference in time and time-frequency representations between two speeds of 100 and 1000 rpm in the case of the outer-race fault in Test Setup 1. While the transient peaks were more or less equally spaced, they showed up as lines at 100 rpm, but more like "blobs" at 1000 rpm.

$$N_{seg} = n_{rev} \cdot f_{order} \tag{D.2}$$



(**a**) (**b**)

Figure D.7: Comparison of the time-waveform and CWT at (**a**) 100 rpm and (**b**) 1000 rpm.

Noise was added to original vibration signals of Dataset 1 to test the robustness of fault classifiers based on CNN, SVM, and SAE-SVM. Table D.7 summarizes the SNR levels added to the vibration signals. Figure D.8 shows the exemplary original and vibration

signals under different SNRs specified by 0, −3 and −10 dB. Impacts due to outer race defects could be easily observed in the original signals (top figure). Increasing noise to the signal causes the fault impacts to be mixed with the noise, resulting in a big challenge of fault detection if using signal processing techniques alone. Figure D.9 shows a detailed flowchart of preprocessing for Dataset 1 with an outer race defect on the axial roller bearing. After the rms normalization, the original signal and signal under noise were subdivided into segments of two revolutions, which were converted to spectrograms by using CWT.

Table D.7: SNR levels.

SNR1	Reference (no added noise)
SNR2	0 dB
SRN3	−3 dB
SNR4	−10 dB

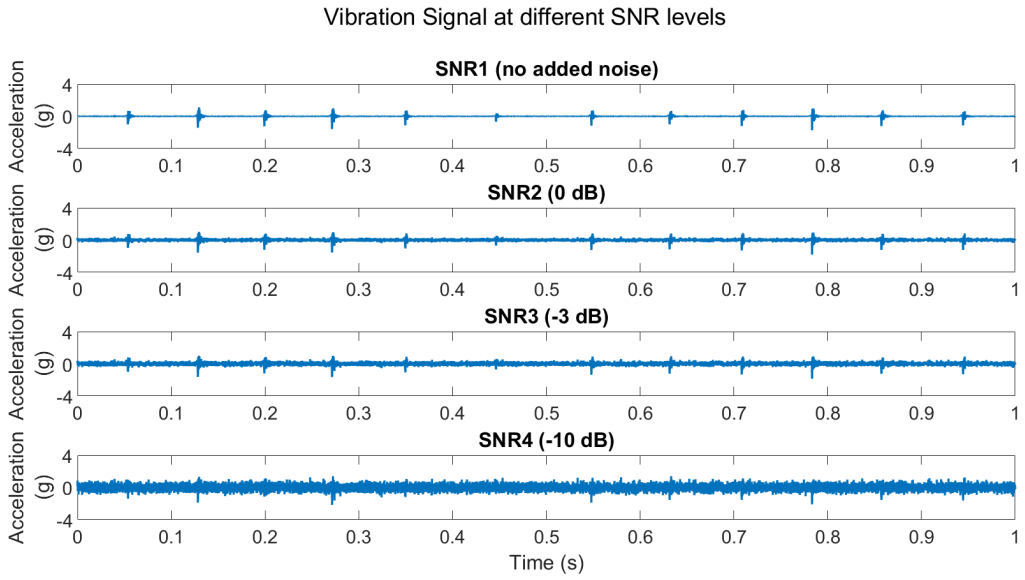


Figure D.8: Fault Type 1 vibration signal with an outer race fault at 4 increasing SNR levels.

Dataset 2 consists of 10-s acoustic emission waveforms, which were recorded at a sampling frequency $f_{sample} = 1$ MHz. Unlike Dataset 1, no noise was added to the acoustic emission data. Bursts collected in acoustic emission data occurred with inconsistent amplitudes throughout time; thus, adding white noise would cause a large share of the bursts to be undetectable. Order normalization is also not necessary, as all data were collected at the same low speed. Each 10-s of data was then subdivided into segments of 1 s, which were converted to spectrograms by CWT. Figure D.10 shows the process

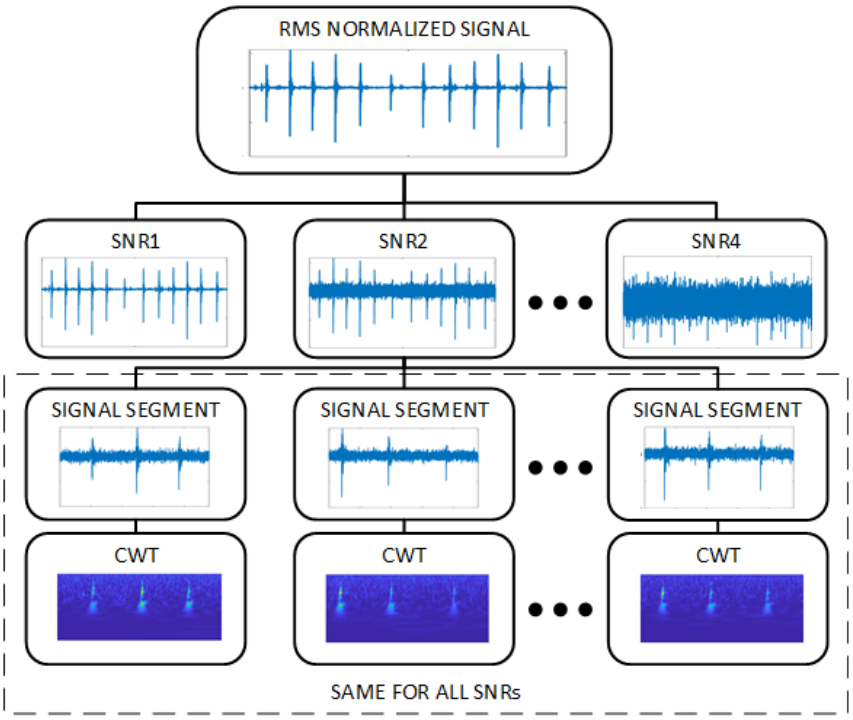


Figure D.9: Dataset 1 preprocessing.

of converting time series data to spectrograms in the case of scratches in the bearing, in which Segment 3 (2–3 s) and Segment 9 (8–9 s) are shown as exemplary spectrograms. Weak vertical lines are visible, corresponding to bursts of acoustic emission energy.

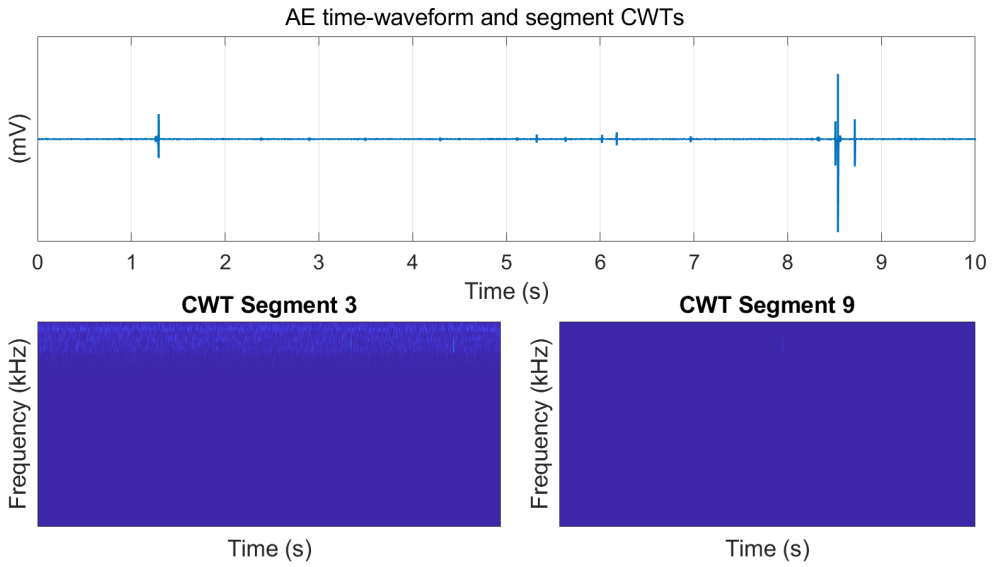


Figure D.10: Acoustic emission signal during damage and CWT spectrograms of Segments 3 and 9.

D.4 Results of Roller Bearing Fault Classifications

This section first presents the results of preprocessing, converting time-series signals to images or spectrograms for CNN algorithm. Further, the process of including noise in the original vibration signals is described. The fault classification results from the radial bearing test rig are presented in Section D.4.1 and the fault diagnosis or scratch detection in the axial tapered bearing in Section D.4.2.

D.4.1 Fault Classification for the Radial Bearing Based on Vibration Signals

In this section, the performance of classifiers, namely CNN, SVM, and SAE-SVM, is evaluated based on their accuracy of classifying faults in a radial bearing. Features fc6 and fc7 generated from AlexNet using Dataset 1 were fed into three fault classifiers. The dataset included two fault types (FT): outer race (OR) and rolling element (RE). Four SNR levels were investigated in each fault type, in which SNR1 was without noise, while other SNRs are shown in Table D.7. The classification results at 75%, 50%, and 25% training data are summarized in Table D.8. The CNN classifier was a fine-tuned version of the pretrained network, while the SVM and SEA-SVM classifiers were implemented on both fc6 and fc7 features, as mentioned in Figure D.1, resulting in CNN, SVM fc6, SVM fc7, SAE-SVM fc6, and SAE-SVM fc7 in Table D.8.

We can see from Table D.8 that the overall trend is that the accuracy dropped with the decrease in SNR and the reduction in training data. As shown in the confusion matrices in Figure D.11, the misclassification was not consistent. Healthy (HE) was considered as a negative class, while damaged was considered as a positive class. Target class refers to the ground truth, while output class was the classifier output. At low noise levels, the algorithms tended to give more false alarms than missed detections. For Fault Type 2, RE damage, this pattern changed between SNR3 and SNR4. The number of misclassifications increased more than false alarms.

D.4.2 Fault Classification for the Axial Roller Bearing Based on Acoustic Emission Signals

Validation accuracy for the axial data, shown in Table D.9, was in general above 95%, except for SAE-SVM at fc7. These results are shown in more detail in Figure D.12. The confusion matrices show that for 75 and 25% training data, the classifier was not able to separate classes and labeled all data as healthy. In the 50% case, all healthy data were misclassified as damaged. Additionally, some damaged cases were misclassified as healthy,

Table D.8: Validation accuracy: Dataset 1. FT, fault type.

		FT1-OR				FT2-RE			
Training Data		SNR1	SNR2	SNR3	SNR4	SNR1	SNR2	SNR3	SNR4
75%	CNN	99.3%	100.0%	100.0%	97.9%	100.0%	98.6%	92.9%	89.3%
	SVM fc6	100.0%	100.0%	100.0%	96.4%	100.0%	97.9%	94.3%	87.1%
	SVM fc7	100.0%	100.0%	100.0%	95.7%	99.3%	97.1%	95.0%	82.1%
	SAE-SVM fc6	100.0%	98.6%	100.0%	93.6%	100.0%	85.7%	95.0%	82.1%
	SAE-SVM fc7	100.0%	100.0%	100.0%	90.0%	99.3%	95.7%	93.6%	85.0%
50%	CNN	100.0%	100.0%	100.0%	95.7%	99.3%	97.1%	97.1%	85.0%
	SVM fc6	100.0%	100.0%	100.0%	97.1%	98.6%	97.1%	95.0%	86.4%
	SVM fc7	100.0%	100.0%	100.0%	95.7%	98.6%	95.0%	92.1%	80.0%
	SAE-SVM fc6	100.0%	100.0%	100.0%	95.0%	99.3%	95.0%	94.3%	78.6%
	SAE-SVM fc7	100.0%	100.0%	100.0%	95.0%	99.3%	92.9%	90.0%	79.3%
25%	CNN	98.7%	100.0%	98.6%	86.4%	100.0%	91.4%	92.1%	86.4%
	SVM fc6	100.0%	100.0%	100.0%	95.7%	99.3%	96.4%	95.0%	85.0%
	SVM fc7	100.0%	100.0%	100.0%	95.0%	98.6%	96.4%	94.3%	85.7%
	SAE-SVM fc6	100.0%	100.0%	99.9%	82.1%	98.6%	97.1%	92.1%	69.3%
	SAE-SVM fc7	100.0%	100.0%	99.3%	87.1%	97.9%	93.6%	90.7%	78.6%

dropping the overall accuracy to only 22.7%.

Table D.9: Validation accuracy: Dataset 2.

Training Data	**CNN**	**SVM fc6**	**SVM fc7**	**SAE-SVM fc6**	**SAE-SVM fc7**
75	97.3%	98.7%	98.7%	97.3%	66.7%
50	98.7%	100.0%	99.3%	94.7%	22.7%
25	97.3%	98.0%	96.7%	95.3%	66.7%

D.5 Discussions

Table D.10 shows accuracy averaged across SNR levels and results for different amounts of test data, the overall accuracy in both datasets. The SVM classifier had the highest accuracy, followed by CNN and SAE-SVM, respectively. The SAE-SVM score was heavily affected by poor performance in analyzing features fc7 in Dataset 2.

While accuracy is an indicator of classifier performance, detection rate and of false alarm rate will further justify performance evaluation. Table D.11 shows the probability

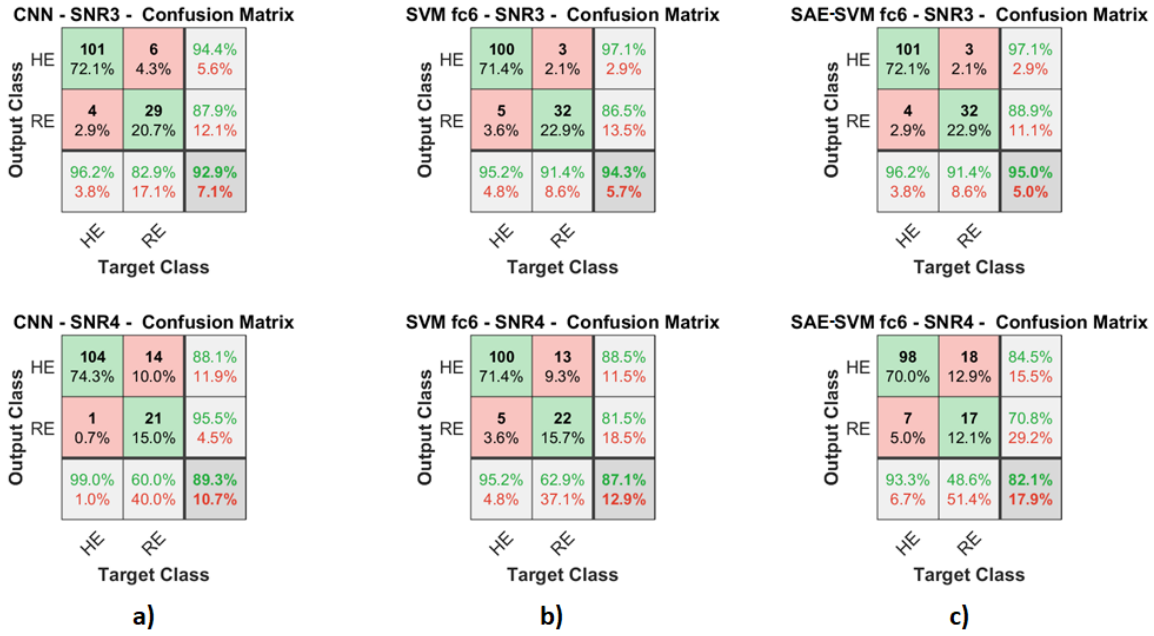


Figure D.11: Confusion matrices for Dataset 1, 75% training data, Fault Type 2 at SNR3 and SNR4. (**a**) CNN classifier, (**b**) SVM classifier at fc6, and (**c**) SAE-SVM classifier at fc6.

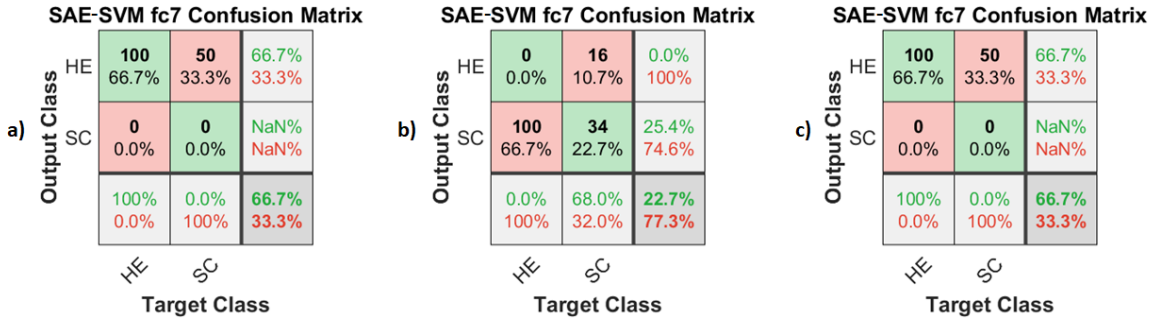


Figure D.12: Confusion matrices for Dataset 2 with (**a**) 75%, (**b**) 50%, and (**c**) 25% training data.

Table D.10: Mean accuracy for each classifier.

	Accuracy
CNN	96.93%
SVM fc6	97.81%
SVM fc7	97.04%
SAE-SVM fc6	94.89%
SAE-SVM fc7	73.23%

of false alarm (PFA) and probability of detection (POD) averaged across SNR level and training data size. Additionally, the mean value across both datasets is included. These

metrics are summarized and color coded using dark green (best), light green, yellow, light red, and dark red (worst) for each dataset. Ideally, a classifier has a high POD combined with a low PFA.

Table D.11: Mean probability of false alarm (PFA) and probability of detection (POD) across SNR and training data size.

	PFA				**POD**			
	Radial FT1	Radial FT2	Axial	Mean	Radial FT1	Radial FT2	Axial	Mean
CNN	1.19%	1.98%	2.00%	1.72%	95.71%	82.38%	96.00%	91.36%
SVM fc6	0.79%	2.54%	1.00%	1.44%	98.81%	85.00%	98.67%	94.16%
SVM fc7	0.87%	3.81%	2.33%	2.34%	98.10%	82.86%	99.33%	93.43%
SAE-SVM fc6	2.22%	4.92%	0.67%	2.60%	96.19%	77.14%	88.67%	87.33%
SAE-SVM fc7	2.22%	4.52%	33.33%	13.36%	97.14%	78.81%	22.67%	66.21%

Evaluation of the classifiers was done qualitatively with respect to classification performance, robustness, ease of implementation, and computational demand. When ranking performance between classifiers, the notation of $X/5$ was used, as five different variations have been tested where $X$is the ranking of the performance, e.g., $X = 1$ indicates the best performance.

D.5.1 CNN Classifier

As seen in Table D.10, the CNN classifier had an overall accuracy of 96.93%, which was ranked 3/5 of the tested cases. Closer examination of Tables D.8 and D.12 reveals that accuracy with 25% training data in Dataset 1 had the most negative impact on overall score. Its PFA ranked 2/5, while its POD 3/5. Its POD ranked 5/5 for Dataset 1, FT1, but still over 95%. The implementation was easy, but required training the network. The performance with 25% training data suggests that more training data are required compared to other classifiers. The CNN classifier also scaled well to multi-class classification problems by simply increasing the number of neurons in the final layers.

D.5.2 SVM Classifier

The SVM classifier was the easiest to implement. Filters and weights from the pretrained network were not modified, and the features input to the SVM were available without any fine-tuning of the network. The major tuning parameter was from which layer to extract the features. In this paper, features at layers fc6 and fc7 were used, but fc6

showed a better accuracy, lower PFA, and higher POD for all datasets except POD for Dataset 2. Additionally, both mean accuracy, PFA, and POD ranked 1/5 overall. Tuning of SVM parameters and different kernels will affect performance, but training the SVM is less computationally heavy than training the CNN or autoencoders. Overall, the SVM classifier on features fc6 had the best performance among the tested classifiers.

D.5.3 SAE-SVM Classifier

The sparse autoencoder added to the SVM classifier became an SEA-SVM classifier. While unsupervised extraction of important parameters seems favorable, the methods showed no consistent advantage over the SVM classifier in terms of classification performance. As illustrated by the results in Dataset 2, the autoencoder might even fail to extract useful features for discriminating between classes where the SVM classifier succeeds. Extracting features at fc7 ranked 5/5 in accuracy, PFA, and POD, mainly due to performance on the axial roller bearing dataset. In contrast, the SAE-SVM using fc6 was ranked 1/5 in PFA in fault classification for the axial roller bearing, but this result was accompanied by a 4/5 rank in POD. Introducing the autoencoder in addition to the SVM adds complexity in terms of tuning parameters and requires time and computational power to train. Combined with the results, this method is not recommended for classifying faults in roller bearings if using simple transfer learning.

Table D.12 gives a comparative evaluation of the proposed fault classifiers for roller bearings, where + and ++ indicate good and very good relative performance, while − and −− are the negative equivalents.

Table D.12: Classifier evaluation.

	Accuracy	**Robustness**	**Implementation**	**Computational Burden**
CNN	+	+	+	−
SVM	+	+	++	+
SAE-SVM	−	−	−	−−

D.5.4 Comparison with Envelope Analysis

Envelope analysis has been commonly used in detecting bearing faults in industry. The performance of the proposed algorithms using machine learning was compared to those of using envelope analysis. In Dataset 1, the roller element fault (FT2) was more difficult to detect than the outer race fault; thus, we provide an example where FT2 at the lowest speed (100 rpm) was analyzed using envelope analysis. In the proposed classifiers, segments of two revolutions were used. To improve resolution in the envelope spectrum,

five segments were combined so that total segment length was extended to 10 revolutions. As shown in Figure D.13, even though transients were visible in the time domain waveform, no clear peak was visible in the envelope spectrum without further processing. However, the proposed method, here illustrated by the CNN classifier, can predict the correct class with above 99% probability. Envelope analysis would in this case require a certain expertise to perform further analysis.

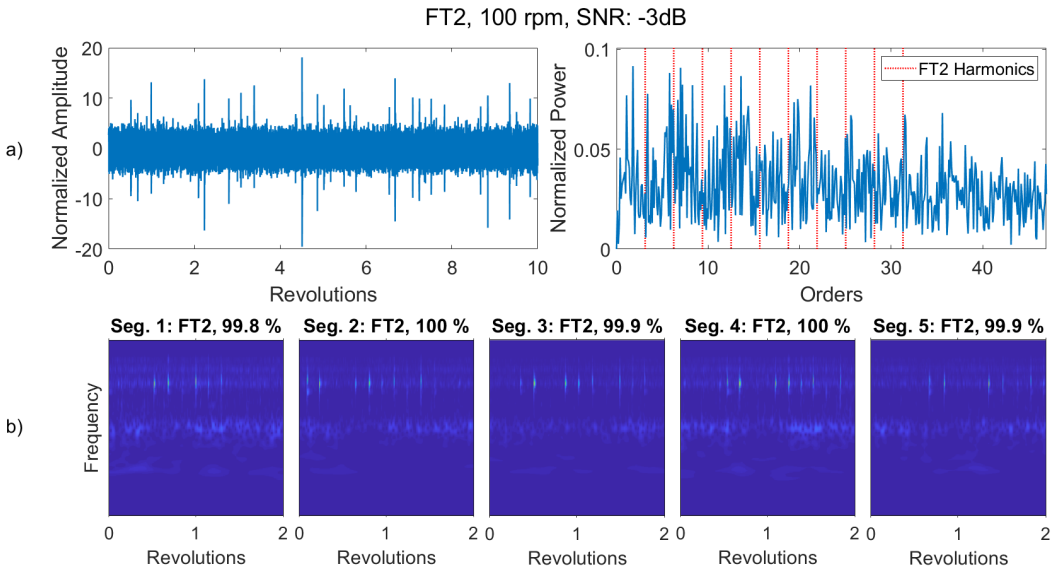


Figure D.13: (**a**) Time domain signal and the envelope spectrum for the faulty roller element at 100 rpm. (**b**) Classification of the corresponding spectrograms using the CNN classifier.

D.6 Conclusions

A transfer learning approach to bearing fault classification using a pretrained convolution neural network (CNN) was proposed in this work. It was shown that the pretrained network can be fine tuned, or used to generate features for detecting bearing faults by other machine learning-based classifiers. Three classifiers based on CNN, support vector machine (SVM) and combined sparse autoencoder (SAE) and SVM algorithms were used to classify faults in axial and radial roller bearings using both vibration and acoustic emission signals.

The performance and robustness of the proposed method were investigated under different fault types, operating speed, and noise levels. The investigation shows that extracting features from the pretrained CNN directly, then using the SVM for classification, is the best option to detect faults in roller bearings in terms of robustness, easy implementation, and computational burden. Fine-tuning of the CNN scales well to multiclass

classification problems, but yields lower accuracy than the SVM classifier. Combined with increased computational burden and more tunable hyperparameters, the CNN-based classifier is ranked as the second best option. Unsupervised dimensionality reduction using SEA to the extracted features from the pretrained CNN increases the computational burden and complexity of the SAE-SVM classifier for this application. It also has a negative effect on robustness and thus the accuracy of the classification.

Funding

The research presented in this paper has received funding from the Norwegian Research Council, SFI Offshore Mechatronics, Project Number 237896.

References

[1] P. F. Albrecht, J. C. Appiarius, R. M. McCoy, E. L. Owen, and D. K. Sharma. Assessment of the Reliability of Motors in Utility Applications — Updated. *IEEE Transactions on Energy Conversion*, 1:39–46, 1986. doi:10.1109/TEC.1986.4765668.

[2] *ISO 18436-2:2014*. 2014.

[3] A. Jacop, H. V. Khang, K. G. Robbersmyr, and A. J. Cardoso. Bearing Fault Detection for Drivetrains using Adaptive Filters Based Wavelet Transform. In *2017 20th International Conference on Electrical Machines and Systems, ICEMS 2017*, pages 1–6, 2017. doi:10.1109/ICEMS.2017.8055941.

[4] H. Qiu, J. Lee, J. Lin, and G. Yu. Wavelet Filter-based Weak Signature Detection Method and Its Application on Rolling Element Bearing Prognostics. *JOURNAL OF SOUND AND VIBRATION*, 289:1066–1090, 2006. doi:10.1016/j.jsv.2005.03.007.

[5] M. Hemmer and T. I. Waag. A Comparison of Acoustic Emission and Vibration Measurements for Condition Monitoring of an Offshore Drilling Machine. In *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, pages 278–285, 2017.

[6] H. Z. Z. Du X. Chen and R. Yan. parse feature identification based on union of redundant dictionary for wind turbine gearbox fault diagnosis. *IEEE Transactions on Industrial Electronics*, 10:6594–6605, 2015.

[7] D. Jung and C. Sundström. A Combined Data-driven and Model-based Residual Selection Algorithm for Fault Detection and Isolation. *IEEE Transactions on Control Systems Technology*, 27(2):616–630, 2019. doi:https://doi.org/10.1109/TCST.2017.2773514.

[8] X. Dai and Z. Gao. From Model, Signal to Knowledge: A Data-driven Perspective of Fault Detection and Diagnosis. *IEEE Transactions on Industrial Informatics*, 9(4):2226–2238, 2013. doi:https://doi.org/10.1109/tii.2013.2243743.

[9] Z. Gao, C. Cecati, and S. X. Ding. A Survey of Fault Diagnosis and Fault-tolerant Techniques-part I: Fault Diagnosis With Knowledge-based and Hybrid/active Approaches. *Transactions on Industrial Informatics*, 62(6):3757–3767, 2015. doi:https://doi.org/10.1109/tie.2015.2419013.

[10] Z. Gao, C. Cecati, and S. X. Ding. A Survey of Fault Diagnosis and Fault-tolerant Techniques-part Ii: Fault Diagnosis With Knowledge-based and Hybrid/active Approaches. *IEEE Transactions on Industrial Electronics*, 62(6):3768–3774, 2015. doi:10.1109/TIE.2015.2419013.

[11] G. Niu. *Data-driven Technology for Engineering Systems Health Management: Design Approach, Feature Construction, Fault Diagnosis, Prognosis, Fusion and Decision*. Springer, Singapore, 2016. doi:https://doi.org/10.1007/978-981-10-2032-2.

[12] A. B. N. Baydar Q. Chen and U. Kruger. Detection of Incipient Tooth Defect in Helical Gears using Multivariate Statistics. *Mechanical Sys. and Sig. Proc.*, 15:303–321, 2001.

[13] P. Konar and P. Chattopadhyay. Bearing Fault Detection of Induction Motor using Wavelet and Support Vector Machines (SVMs). *Applied Soft Computing*, 11(6):4203–4211, 2011. doi:https://doi.org/10.1016/j.asoc.2011.03.014.

[14] J. S. L. Senanayaka, S. T. Kandukuri, H. V. Khang, and K. G. Robbersmyr. Early Detection and Classification of Bearing Faults using Support Vector Machine Algorithm. In *2017 IEEE Workshop on Electrical Machines Design, Control and Diagnosis (WEMDCD)*, pages 250–255, 2017. doi:10.1109/WEMDCD.2017.7947755.

[15] J. S. L. Senanayaka, H. V. Khang, and K. G. Robbersmyr. Towards Online Bearing Fault Detection using Envelope Analysis of Vibration Signal and Decision Tree Classification Algorithm. In *IEEE International Conference on Electrical Machines and Systems (ICEMS)*, pages 1–6, 2017. doi:https://doi.org/10.1109/icems.2017.8056146.

[16] R. Zhang, H. Tao, L. Wu, Y. Guan, L. W. R. Zhang H. Tao, Y. Guan, R. Zhang, H. Tao, L. Wu, and Y. Guan. Transfer Learning With Neural Networks for Bearing Fault Diagnosis in Changing Working Conditions. *IEEE Access*, 5:14347–14357, 2017. doi:10.1109/ACCESS.2017.2720965.

[17] L. Wu, B. Yao, Z. . Peng, and Y. Guan. Fault Diagnosis of Roller Bearings Based on a Wavelet Neural Network and Manifold Learning. *Appl. Sci.*, pages 1–10, 2017.

[18] G. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi. Machine Learning for Predictive Maintenance: A Multiple Classifier Approach. *IEEE Trans. Ind. Informat.*, 11:14347–14357, 2015.

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet Classification With Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105, 2012.

[20] E. Shelhamer. Alexnet Model. URL: `https://github.com/BVLC/caffe/tree/master/models/bvlc{_}alexnet`.

[21] J. S. L. Senanayaka H. V. Khang and K. G. Robbersmyr. Multiple Fault Diagnosis of Electric Powertrains Under Variable Speeds using Convolutional Neural Networks. In *Proc. IEEE ICEM*, pages 1932–1938, 2018. `doi:10.1109/ICELMACH.2018.8507171`.

[22] J. A. Suykens and J. Vandewalle. Least Squares Support Vector Machine Classifiers. *Neural Processing Letters*, 9(3):293–300, 1999. `doi:10.1023/A:1018628609742`.

[23] J. M. Z. E. Hosseini-Asl and O. Nasraoui. Eep Learning of Part-based Representation of Data Using Sparse Autoencoders With Nonnegativity Constraints. *IEEE Transactions on Neural Networks and Learning Systems*, 27:2486–2498, 2016.

[24] N. L. F. Jia Y. Lei and S. Xing. Deep Normalized Convolutional Neural Network for Imbalanced Fault Classification of Machinery and Its Understanding via Visualization. *Mechanical Sys. and Sig. Proc.*, 110:14347–14357, 2018.

Paper E

Health Indicator for Low-speed Axial Bearings using Variational Autoencoders

Martin Hemmer, Andreas Klausen, Huynh V. Khang, Kjell G. Robbersmyr and Tor I. Waag

This paper has been submitted as:

Martin Hemmer, Andreas Klausen, Huynh V. Khang, Kjell G. Robbersmyr and Tor I . Waag. Health Indicator for Low-speed Axial Bearings using Variational Autoencoders. Under review at: *IEEE Access*.

Health Indicator for Low-speed Axial Bearings using Variational Autoencoders

Martin Hemmer*,**, Andreas Klausen*, Huynh Van Khang*, Kjell G. Robbersmyr* and Tor I. Waag***

*University of Agder
Department of Engineering Sciences
Jon Lilletunsvei 9, N-4879 Grimstad, Norway

**MHWirth AS AS
Butangen 21, N-4639 Kristiansand, Norway

***NORCE Norwegian Research Centre AS
Jon Lilletuns vei 9 H, N-4877 Grimstad, Norway

Abstract – This paper proposes a method for calculating a health indicator (HI) for low-speed axial rolling element bearing (REB) health assessment by utilizing the latent representation obtained by variational inference using Variational Autoencoders (VAEs), trained on each speed in the dataset. Further, versatility is added by conditioning on the speed, extending the VAE to a conditional VAE (CVAE), thereby incorporating all speeds in a single model. Within the framework, the coefficients of autoregressive (AR) models are used as features. The dimensionality reduction inherent in the proposed method lowers the need of expert knowledge to design good condition indicators. Moreover, the suggested methodology allows for setting the probability of false alarms when encoding new data points to the latent variable space using the trained model. The effectiveness of the proposed method is validated based on two different datasets: from a workshop test of an offshore drilling machine and from an in-house test rig for axial bearings. In both datasets, the HI is exceeding the warning and alarm levels with a probability of false alarm (PFA) of 10^{-6}, and the method is most effective at lower shaft speeds.

E.1 Introduction

Rolling element bearings (REBs) are widely used in heavy industrial machinery such as offshore drilling machines, wind turbines, and paper mills. A defect in such bearings might result in a catastrophic failure in the industrial system. Therefore, condition monitoring (CM) for REBs is important to avoid unplanned downtime and production loss in heavy industry. The majority of bearing condition monitoring techniques focus on detecting the

presence and development of localized damage in bearing raceways or rolling elements [1, 2, 3]. CM of low-speed machinery, with a shaft speed below 10 Hz [4], is more challenging. The energy associated with faults is then smaller, resulting in a low signal-to-noise ratio (SNR). This requires more sensitive sensors and development of advanced signal processing methods to extract fault signatures. Operating conditions tend to be less stationary at lower speeds [4], thus resampling to angular domain is necessary in low-speed applications [5, 6]. Health conditions of large bearings at low speed are usually observed via acoustic emission or vibration measurements [7, 8, 5, 9, 10, 11]. Cyclostationary methods [12, 13, 14], wavelet denoising and filtering [15, 16], and empirical mode decomposition (EMD) [17] have all been successful in low-speed bearing fault detection. Data-driven fault diagnosis methods based on machine learning have also been intensively developed in recent years [18]. Fault classifiers based on decision trees (DT) [19, 20], support vector machine (SVM) [21, 22], k-nearest neighbor (k-NN) [23, 24], convolutional neural network (CNN) [25, 26, 27] and deep belief networks (DBN) [28, 29] are well applied to deal with bearing fault detection. All mentioned machine learning based methods require historical failure data for training, which is hard to obtain in industry. In addition, the authors could not identify previous research dealing with faults on axial bearings, where a characteristic fault frequency might not exist or is inconsistent in spectra. This work aims to develop an anomaly detection method without using historical failure data. Tapered axial roller bearings, e.g. in drilling machines from the offshore industry, have relative sliding motion in the rib-roller contact area. Low speed makes this area particularly susceptible to wear. In [30], wear on the roller ends was observed in a tapered axial bearing, as shown in Fig. E.1.

Figure E.1: Wear on roller end of offshore drilling machine bearing.

However, no characteristic frequency component associated with defects, i.e. roller

frequency, was observed on the axial bearing during tests. This suggests that diagnosis methods based on detection of defect characteristic frequencies alone are ineffective in detecting wear in large and slow axial bearings. Identifying this defect on the axial bearings is currently relying on offline monitoring methods such as lubricant analysis and visual inspection combined with precautionary maintenance actions [30, 25]. This practice requires interruptions of production and may allow failure to progress inconspicuously between inspections. Therefore, development of online, non-intrusive monitoring methods is very important to facilitate condition based maintenance (CBM) for large axial bearings in heavy industry.

Since data from a healthy state is easier to obtain than in a damaged state, a procedure of determining whether or not the observed bearing is normal based on prior knowledge of healthy behavior of the machine, would be very useful to avoid using failure data.

References [31, 32] proposed a method for health threshold setting based on healthy operating characteristics, allowing controlling probability of false alarm (P_{FA}). A whitening transformation was applied to a set of correlated condition indicators (CIs) with Gaussian or Rayleigh distributions. These CIs were then used to calculate a health indicator (HI) with a known probability density function (PDF) and cumulative distribution function (CDF). The HI is normalized by the inverse CDF evaluated at $(1 - \mathrm{P}_{FA})$, and optionally multiplied with a warning factor $w < 1$. In this case, let HI_0 denote an observation from a healthy machine. The probability of observing HI_0 above the warning factor is then equal to the P_{FA}, as shown in (E.1). The consequences of failures and false alarms must be considered when setting the P_{FA} threshold. Additionally, the number of inferences to be done must be considered. Multiple testing increases the risk of false positive samples simply by chance [33].

$$P(\mathrm{HI}_0 > 1) = \mathrm{P}_{FA} \tag{E.1}$$

However, the method in [31] is only effective if CIs are well selected with known probability distributions. The overall goal of conventional approaches is to perform health assessment with a statistical foundation, based on a potentially large set of observed variables. Due to the curse of dimensionality, also known as Hughes' phenomenon [34], a single indicator might "drown" in high dimensional feature space, which reduces accuracy of the model. Thus, a method for dimensionality reduction of the features and maintaining most of the information is required. To implement the HI threshold setting, features are also required to be independent variables. Principal Component Analysis (PCA) can transform a set of variables to linearly uncorrelated features with decreasing contribution to the variance, but it does not account for non-linear dependencies. Machine learning (ML) algorithms can be an alternative solution since they can capture complex dependencies among the observed variables.

Autoencoders are successfully used for dimensionality reduction in fault detection and classification of rotating machinery [35, 36, 37, 38], but lack a probabilistic latent representation. Generative models are capable of estimating complicated PDFs of given data, and can generate new samples, which follow the same distribution as the training data. In [39], it was shown that sequential training of restricted Boltzmann machines could discover hidden dependencies between observed variables and a sparse representation. However, training such networks typically requires an additional statistic method, e.g. Markov Chain-Monte Carlo (MCMC) methods, resulting in computational burden.

To achieve dimensionality reduction and reduce computational burden, this work uses a combination of a Variational Autoencoder (VAE) [40] and a Generative Adversarial Networks (GAN) [41], which is similar to the Adversarial Autoencoder (AAE) [42]. The VAE performs inference of variational parameters using neural networks in an encoder-decoder structure by minimizing the reconstruction error and the Kullback-Leibler Divergence (D_{KL}) [43] between an encoded sample and a Gaussian standard distribution, which is equivalent to maximizing the evidence lower bound (ELBO). This objective can be optimized with gradient descent algorithms through the "reparameterization trick" [40]. These generative models allow imposing a distribution on the latent variables. In [42], the latent distribution in AAEs seems to follow the target distribution closer, which is desirable for the purpose of a HI. However, the adversarial training in GANs and AAEs is often unstable [44]. This problem was also observed in experiments with AAEs while devloping the proposed method. VAEs have been used in ball bearing fault classification by using the latent variables for each data point as input to a classifier [45]. The proposed approach instead utilizes the aggregated distribution of healthy conditions in the latent space of a VAE to calculate a HI for new observations.

The remaining of the paper is organized as follows: In section E.2, the network architecture, training procedure and HI calculation are described. Section E.3 details the data acquisition and pre-processing. Results from two different datasets are presented in section E.4. Section E.5 provides conclusions and discussions.

E.2 Methodology

This section presents the approach for calculating a bearing health indicator, utilizing the latent variables in a VAE. The calculation of a HI limits the selection of CIs to those following known distributions as described in [31] for Gaussian and Rayleigh distributions. It also requires the user to pre-select suitable CIs based on domain knowledge. The proposed method performs unsupervised dimensionality reduction from a set of input features, while simultaneously imposing a Gaussian distribution on the latent variables.

This section provides a review of the network components, the loss functions and training algorithm. The model was implemented in Python using TensorFlow r1.12 [46].

E.2.1 Network Architecture and Losses

The network architecture is shown in Fig. E.2. An encoder (red) and a decoder (green) are connected by the latent representation (yellow).

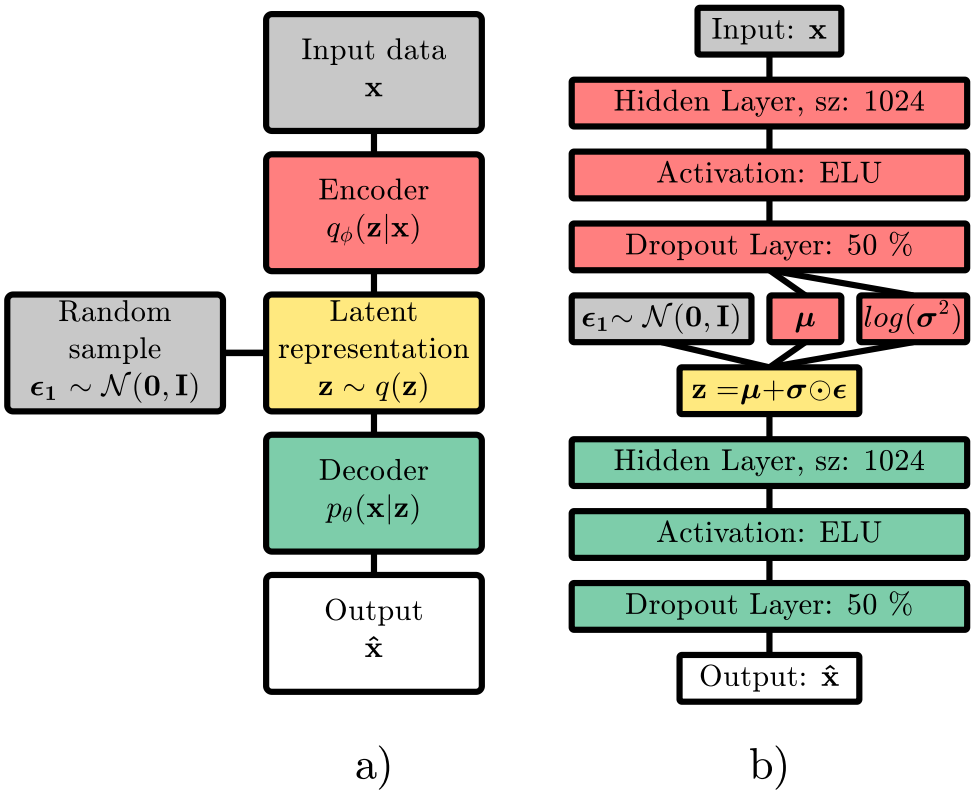


Figure E.2: **a)** VAE architecture overview, **b)** Detailed architecture.

Let $\mathbf{x}$ be the feature input vector and $\mathbf{z}$ be the latent variable vector. The encoder consists of a fully connected layer of size 1024 with weights, biases, an exponential linear unit (ELU) activation function, and 50 % dropout. In this work, the coefficients of an autoregressive model are used as features. The output includes two vectors, containing the parameters of the latent representation for each data point. Let J be the dimension of the latent space. The latent variables are constrained to have a Gaussian distribution with diagonal covariance matrix, so the encoder outputs a vector containing the means, $\boldsymbol{\mu}$, and log of the variances, $log(\boldsymbol{\sigma}^2)$, each of length J. Note that these parameters are for the individual data points, not the aggregated latent distribution $q(\mathbf{z})$. Utilizing the reparameterization trick from [40], samples from a white noise vector $\boldsymbol{\epsilon}$ are used to obtain a random sample $\mathbf{z}$ from the latent representation while still allowing gradients to flow through the network. The decoder has the same architecture as the encoder,

with a fully connected hidden layer, ELU activation, and 50 % dropout. Weights and biases are denoted $\boldsymbol{\phi}$. The desired output is a reconstruction of the input, like a normal autoencoder. Combining these parts of the network results in the VAE. Originally, the VAE was developed as a generative model for producing reconstructions similar to the input by sampling from a given prior distribution $p(\mathbf{z})$. The connection between data and $p(\mathbf{z})$ is in general not known, and must be approximated. Let the training data distribution be $\mathbf{x} \sim p_d(\mathbf{x})$, and VAE output $\hat{\mathbf{x}} \sim p(\mathbf{x})$. Further, $q_\phi(\mathbf{z}|\mathbf{x})$ and $p_\theta(\mathbf{x}|\mathbf{z})$ are the encoding and decoding distributions of the encoder and decoder networks. Subscript ϕ and θ are the encoder and decoder variables. Thus, the aggregated posterior distribution of the latent variable, $\mathbf{z} \sim q(\mathbf{z})$, is defined as in (E.2). To be utilized in the HI calculation, $q(\mathbf{z})$ must approximate the desired prior $p(\mathbf{z})$.

$$q(\mathbf{z}) = \int_{\mathbf{x}} q_\phi(\mathbf{z}|\mathbf{x}) p_d(\mathbf{x}) d\mathbf{x} \tag{E.2}$$

To ensure that the latent representation contains useful information about the input data, the encoder and decoder are trained to minimize the reconstruction loss function $\mathcal{L}_R$, as in (E.3). $\mathcal{L}_R$ is the mean square error between each feature $\mathrm{x}_{i,j}$ and its reconstruction $\hat{\mathrm{x}}_{i,j}$ over a minibatch, $\mathbf{x}_M$, of size M. The number of features per datapoint is denoted N.

$$\mathcal{L}_R(\boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{x}_M, \boldsymbol{\epsilon}) = \frac{1}{M} \sum_{i=1}^{M} \sum_{j=1}^{N} (\mathrm{x}_{i,j} - \hat{\mathrm{x}}_{i,j})^2 \tag{E.3}$$

This encourages similar input data to cluster in latent space, while dissimilar data are separated. Note that the square error is summed over a datapoint and averaged over the minibatch. This gives more weight to reconstruction error, which helps avoid mode collapse, i.e. the latent vector converges to a Gaussian that does not carry information.

While reducing $\mathcal{L}_R$ provides a good reconstruction, the aggregated latent distributions will not take a Gaussian distribution. To make the latent distribution approximate the desired prior, KL divergence is introduced as a regularization on the encoder variables ϕ. Given the assumption of diagonal covariance matrix, and Gaussian prior, the KL divergence for a data point can be calculated in a closed form. The combined KL loss over a minibatch is then calculated as in (E.4).

$$\begin{aligned} \mathcal{L}_{KL}(\boldsymbol{\phi}; \mathbf{x}_M) \; = \frac{1}{2M} \sum_{i=1}^{M} \sum_{j=1}^{J} \big(& 1 + log(\sigma_{i,j}^2) \\ & - \mu_{i,j}^2 - \sigma_{i,j}^2 \big) \end{aligned} \tag{E.4}$$

The objective function to be minimized is the sum of $\mathcal{L}_R$ and $\mathcal{L}_{KL}$, as given in (E.5).

$$\mathcal{L}_{VAE}(\boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{x}_M, \boldsymbol{\epsilon}) = \mathcal{L}_R + \mathcal{L}_{KL} \tag{E.5}$$

Algorithm 1 Training algorithm.

$\phi, \theta \leftarrow x$ Initialize parameters
repeat
  Shuffle training dataset
  repeat
    $\mathbf{x}_M \leftarrow$ Get minibatch from the training dataset
    $\mathbf{g} \leftarrow \nabla \mathcal{L}_{VAE}(\boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{x}_M, \boldsymbol{\epsilon})$ Calculate gradients
    $\theta, \phi \leftarrow$ Update encoder/decoder parameters
  until Epoch is completed
until Total number of epochs is completed
return ϕ, θ

Pseudo-code for the training procedure is given in Algorithm 1. Parameter updates are performed using the Adam optimizer with cosine decay of the learning rate. Hyperparameters used in the experiments are given in Table E.1. Training was repeated 5 times with different random seeds for weight initialization and shuffling. As suggested in [40], the network parameters are set as $M = 100$ and $L = 1$.

Table E.1: Training Parameters

Parameter	**Value**
Optimizer	Adam
Learning rate schedule	Cosine decay
Inital learning rate	1e-4
Final learning rate	1e-6
Minibatch size (M)	100
Hidden layer size	1024
Dropout	50 %
Activation	ELU
Latent dimension (J)	3
Epochs, Dataset 1	30 000
Epochs, Dataset 2	15 000

E.2.2 Conditional Variational Autoencoder

With the described approach, it is required to train a separate VAE for each speed. For machines with multiple operating conditions, this is impractical. Therefore, a conditional

VAE (CVAE) is trained for each dataset. CVAEs utilize the same network structure and loss function $\mathcal{L}_{VAE}$ as VAEs, but can be conditioned on additional information, such as speed. For each datapoint, the speed information is a categorical variable, one-hot encoded into a conditioning vector c. For example, the speed of 100 rpm in dataset 1 is encoded to $c_{100} = [0, 1, 0, 0, 0]$ while the speed of 60 rpm in dataset 2 is encoded to $c_{60} = [0, 1]$. As the model order is different for rpms, x is zero-padded to the largest model order p. VAE training datasets consist of data from a single spped, while the CVAE uses data from all speeds. Except for these differences, VAEs and CVAEs follow an identical training procedure.

E.2.3 Health Indicator

A methodology for threshold setting given CIs with Rayleigh or Gaussian distributions is proposed in [31]. In this work, a Gaussian distribution is chosen for the latent variables. To verify that $q(\boldsymbol{z})$ approximates the standard normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$, the Kullback-Leibler Divergence (D_{KL}) was calculated as given in (E.6) for the aggregated posterior. Σ is the covariance matrix of $\mathbf{z}$, $\boldsymbol{\mu}$ is a vector containing the mean values of $\mathbf{z}$ and J is the number of latent variables.

$$D_{KL} = \frac{1}{2}\left(\mathrm{tr}\left(\Sigma\right) + \boldsymbol{\mu}^T\boldsymbol{\mu} - J + \log_e\left(\frac{1}{\det\Sigma}\right)\right) \tag{E.6}$$

The norm of J Gaussian variables follows a χ distribution with v degrees of freedom. Let $F(\cdot)$ denote the CDF of a χ-distribution. The HI is normalized with a factor that is a function of the P_{FA}. The HI is calculated as shown in (E.7).

$$\mathrm{HI} = \left(\sum_{j=1}^{J} |z_j|^2\right)^{\frac{1}{2}} \left(F^{-1}(1 - \mathrm{P}_{FA})\right)^{-1} \tag{E.7}$$

E.3 Experimental Setup

The proposed algorithm is tested on data from two experiments: Vibration data from a workshop test of an offshore drilling machine, and acoustic emission (AE) data from an in-house test rig for axial bearings. A further description of the experimental setup is given in the following sections.

E.3.1 Dataset 1: Offshore Drilling Machine Workshop Test

Dataset 1 (DS1) was collected from an offshore drilling machine taken out of operation for maintenance as described in [30]. A schematic drawing of the setup is shown in Fig. E.3.

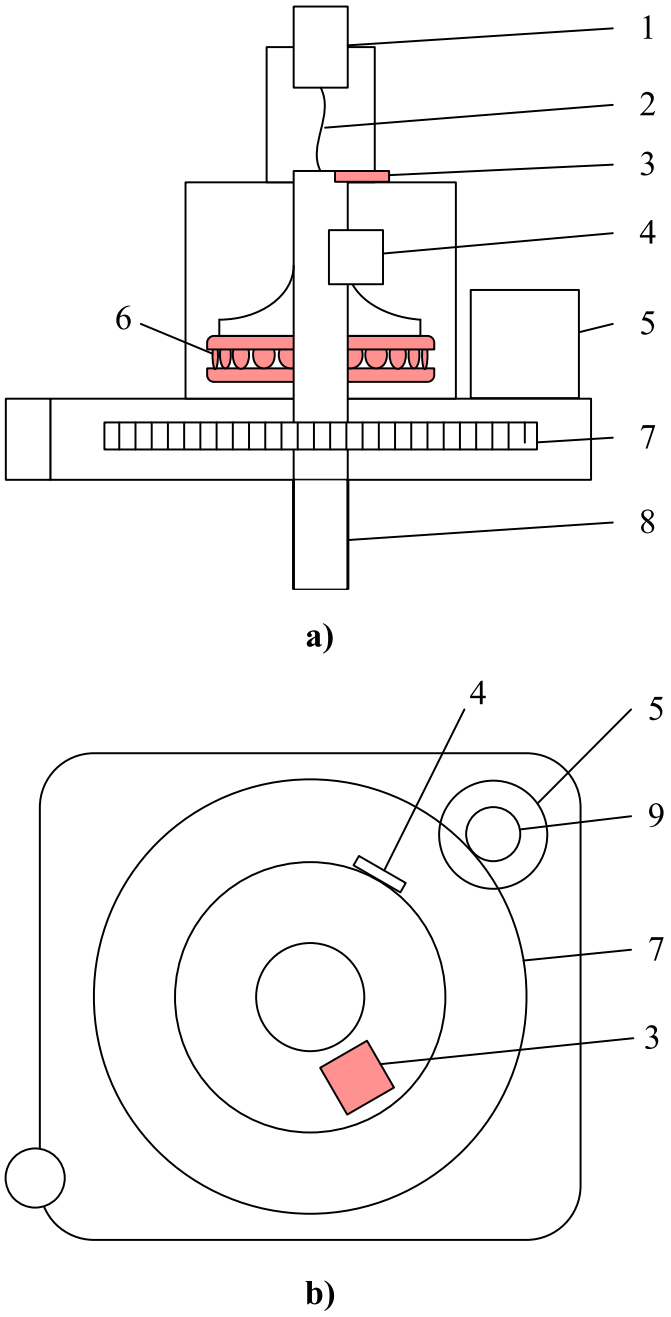


Figure E.3: Conceptual drawing of an offshore drilling machine. **a)** Side view **b)** Top view. 1: Encoder 2: Flexible coupling 3: Axial sensor placement 4: Radial sensor placement, 5: Motor, 6: Axial bearing, 7: Gear, 8: Shaft, 9: Pinion.

Data is collected from an accelerometer mounted in the axial location. Data was sampled at 102.4 kHz and decimated to 81.92 kHz. The axial bearing showed signs of roller end wear as shown in Fig. E.1. Data was first recorded using a healthy bearing, being denoted damage level (DL 0). Then, reassembling the machine with a slightly damaged bearing results in a change of the vibration characteristics and a reduction in root mean square (RMS) [30]. Distinguishing this change from any fault induced change

is not possible. Thus, the slightly damaged condition is selected as the baseline condition (DL 1) for training data. Additional damages in the form of indentations from a carbide tip tool were applied to one of the roller end, producing data at DL 2. For data at DL 3, the bearing was further damaged and also run under poor lubricating conditions. Data was recorded at 50, 100, 150, 200 and 250 rpm. At 50 rpm, only data from DL 1 and DL 3 was recorded. The machine was running unloaded, subject to the gravity by its own weight. A quantitative measurement of damage is not available, but a degradation resulting in a measurable change is expected. However, previous analysis of the vibration signal was not successful in detecting any clear indication of the damage [30]. Damage to a roller was expected to cause amplitude modulation at the roller frequency, but as shown in Fig. E.4, no peak was observed at either one or two times the roller frequency in the envelope spectrum.

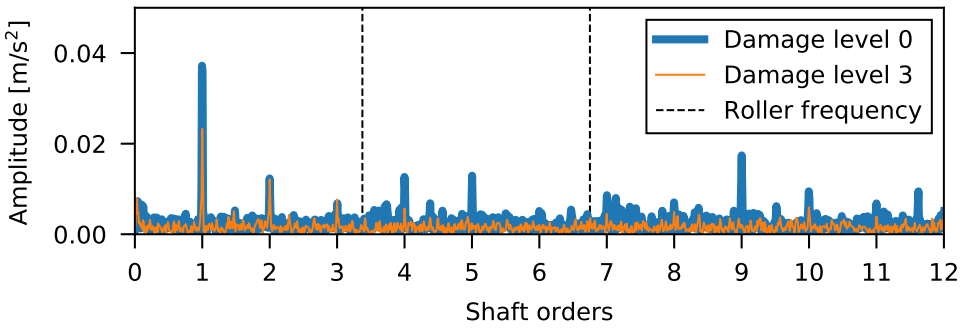


Figure E.4: Envelope spectrum for a healthy bearing (damage level 0) and at damage level 3.

Segments corresponding to approximately 1 revolution are used for calculating the features. To increase the number of data points, an overlap of 75 % is applied. The autocorrelation function (ACF) is examined on a healthy dataset to determine if the signal is stationary or not. If the ACF reduces quickly, the signal is considered stationary [47], otherwise the signal is considered non-stationary. The ACF of a vibration signal acquired at 50 rpm is shown in Fig. E.5 a). The ACF is slowly decreasing, and has a cyclic trend, and the signal is therefore considered stationary. To mitigate trends and cyclic signal components, the signal is differentiated once. Effectively, the jerk (m/s^3) is calculated with this differentiation, and low-frequency components from shaft and gearbox are mitigated, while high-frequency components are enhanced. The resulting ACF after differentiation is shown in Fig. E.5 b), showing that the ACF is now decreasing fast, and only varies randomly after 100 lags. Given this result, the vibration signal acquired on this test rig is differentiated once to make the signal more stationary.

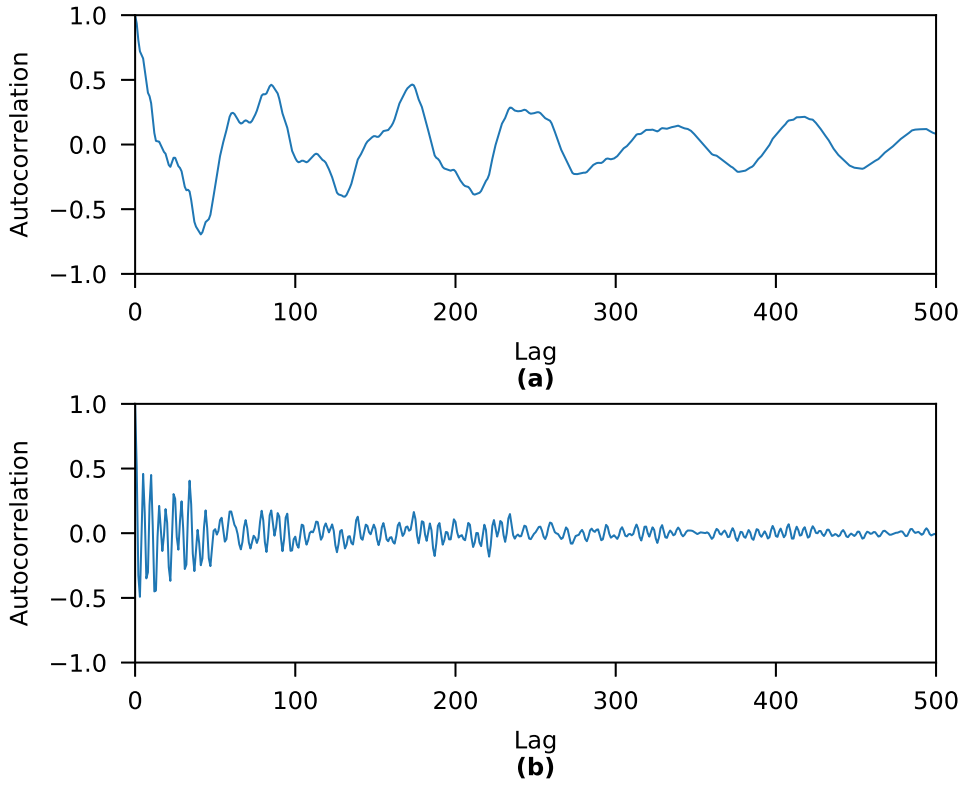


Figure E.5: Autocorrelation function of vibration data captured in Dataset 1. **a)** Raw data **b)** After differentiation.

E.3.2 Dataset 2: Axial Roller Bearing Test Rig

Dataset 2 (DS2) consists of AE data from an in-house test rig, shown in Fig. E.6. The test bearing was of type 29230 M from manufacturer ISB, subject to an axial load of 50 kN. Data was recorded at 30 and 60 rpm, in that order. AE data was collected at 1 MHz sampling rate for 10 seconds. Data was then split into constant length segments of 50 000 samples.

To emulate the distributed abrasive wear shown in Fig. E.1, the rollers were removed, and roller ends were ground with sandpaper of grit size from ISO/FEPA grit grade P400 (finest), P320, P220 and P80 (coarsest), as shown in Fig. E.7. "Heavy" and "Very Heavy" refer to relative degrees of damage using the same sandpaper grade.

The ACFs of the acoustic emission dataset before and after differentiation are shown in Figs. E.8. a) and b), respectively. The ACF of the raw signal in Fig. E.8 a) decreases rapidly, and differentiating the signal has little effect on the ACF as observed in Fig. E.8 b). Therefore, the acoustic emission signal is considered stationary and requires no further differentiation.

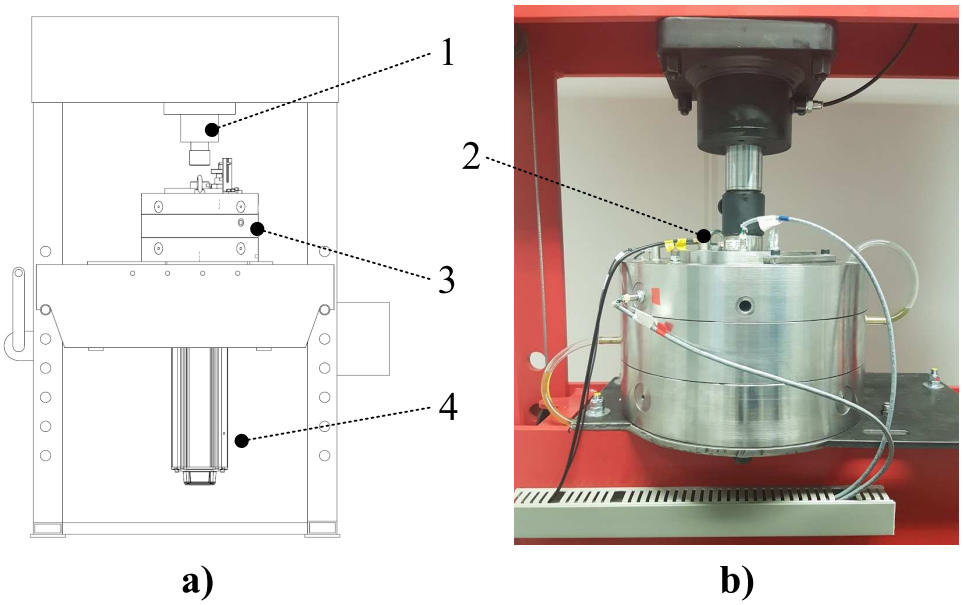


Figure E.6: **a)** Drawing of in-house axial bearing test rig. **b)** Assembled bearing test unit. 1: Load cylinder, 2: AE sensor location, 3: Bearing test unit, 4: Motor.

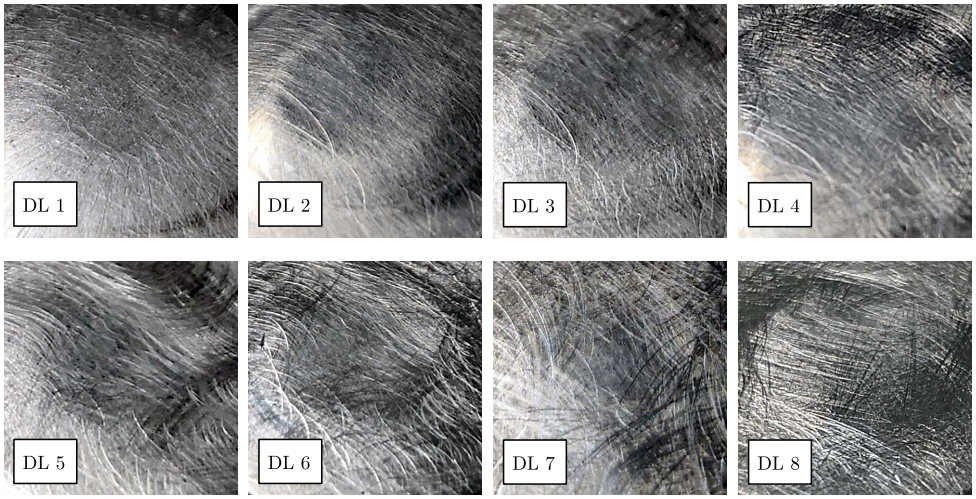


Figure E.7: Damage on roller ends in dataset 2. DL 1: P400, DL 2: P400 Heavy, DL 3: P320, DL 4: P220, DL 5: P220 Heavy, DL 6: P80, DL 7: P8 Heavy, DL 8: P80 Very Heavy

E.3.3 Feature Extraction and Preprocessing

The input $\mathbf{x}$ to the autoencoder network is a feature calculated from the vibration and AE data. In the previous work, vibration energy was not significantly increased when the damage level on an axial bearing was escalated [30]. In addition, energies at specific characteristic frequencies do not increase either. However, the bearing condition degradation is expected to produce a change in frequency content of the associated signal. Therefore, features, which are sensitive to changes in the measured signal, are required to be used

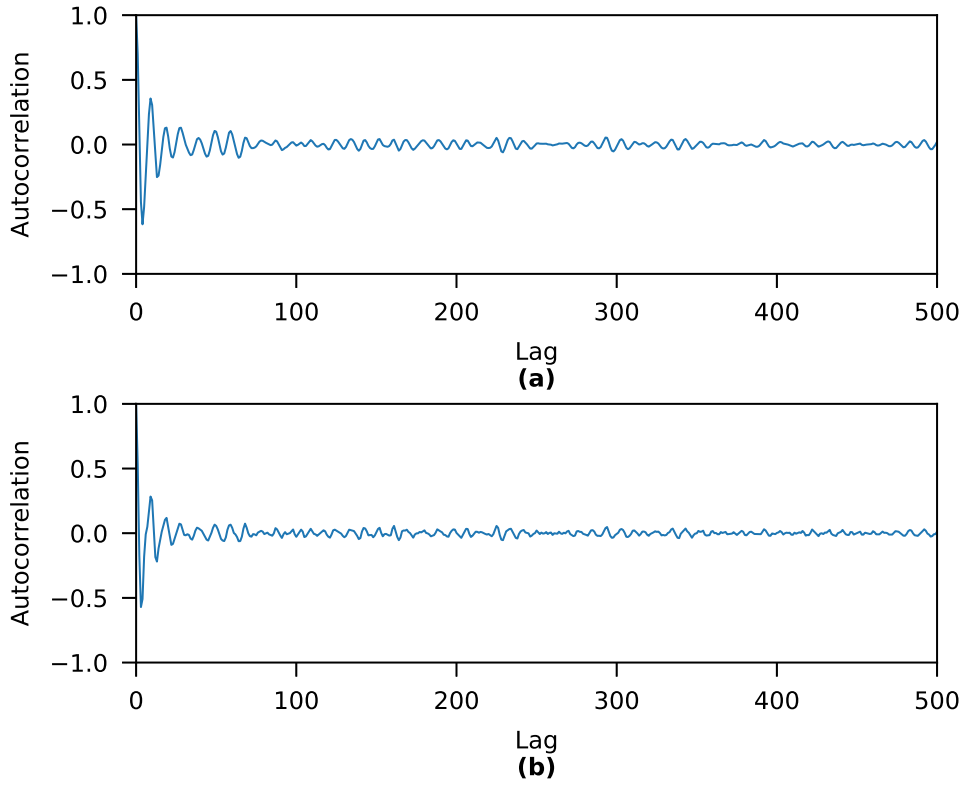


Figure E.8: Autocorrelation function of vibration data captured in Dataset 2. **a)** Raw data **b)** After differentiation.

as input to the autoencoder. An autoregressive (AR) model of order p can predict the next signal sample based on a linear combination of p previous samples, assuming that the signal s is stationary.

Moved and revised: Thus, changes in the AR model parameters should reflect that the vibration signal has changed. The AR coefficients may have arbitrary distributions, which makes it challenging to quantify a change. It is therefore easier to threshold in latent space, where the distribution of healthy latent vectors approximates a Gaussian distribution.

The AR model is depicted as

$$s_i = \nu_i + \sum_{j=1}^{p} a_j s_{i-j} \tag{E.8}$$

where s_i is the signal at i'th time step, ν is the model residual and a_j is the j'th model parameter. The Yule-Walker equations [48, 49] are solved for an input signal s to obtain the AR model parameters.

The order p is determined by calculating the partial autocorrelation function (PACF) [50] for an increasing number of lags. The model order p of a time series with N samples is considered sufficient where PACF at lag p is zero with a 5 % significance level [47], as

given in (E.9) [51].

$$PACF \leq \frac{\pm 1.96}{\sqrt{N}} \tag{E.9}$$

The smallest lag p, which results in a PACF below the 5 % significance level, is determined for each healthy segment. As an example, the PACF of a differentiated vibration signal acquired at 50 rpm using test rig 1 is shown in Fig. E.9.

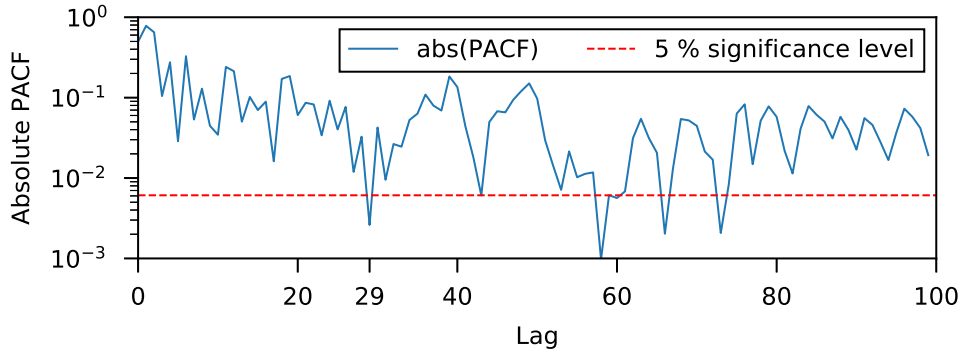


Figure E.9: Absolute value of the partial autocorrelation function of differentiated vibration data acquired using test rig 1. The graph is crossing beneath the confidence interval at lag 29.

At lag 29, the PACF is beneath the 5% significance level. This procedure is repeated for all signal segments, and statistics between all segments within each speed range are calculated afterwards and shown in Table E.2. As seen in the table, the mean value is selected as model order p. Standard deviation (STD) and median are also given for each dataset and speed.

Table E.2: Statistics for calculated model order p.

	DS1-50	**DS1-100**	**DS1-150**	**DS1-200**	**DS1-250**	**DS2-30**	**DS2-60**
Mean (p)	32	12	20	16	15	14	25
STD	9	9	13	9	10	10	11
Median	29	5	23	9	9	7	17

All input data was afterwards standardized using the mean and standard deviation of the remaining training data. Outliers in training data are removed if one AR coefficient differed from the mean value by more than five standard deviations. Baseline data (DL0) was shuffled and split in training (50 %), validation (25 %) and test (25 %) subsets. The remaining DLs were used for testing only. Table E.3 shows sample rate, number of samples in the raw data, and size of the datasets at each DL.

Table E.3: Dataset details.

	DS1-50	**DS1-100**	**DS1-150**	**DS1-200**	**DS1-250**	**DS2-30**	**DS2-60**
Sample rate	81.92 kHz	81.92 kHz	81.92 kHz	81.92 kHz	81.92 kHz	1 MHz	1 MHz
Samples	98304	49152	32768	24576	19661	50000	50000
DL0 size	988	568	596	328	312	2099	2098
DL1 size	-	536	480	280	372	420	420
DL2 size	964	496	388	304	280	420	420
DL3-8 size	-	-	-	-	-	420	420

E.4 Results

This section presents the results of the experiments, evaluating the calculated HI using both VAE and CVAE. The validity of the required assumptions of a Gaussian-distributed latent variable is also discussed. The presented results are the aggregate of the 5 models trained with different random initialization.

E.4.1 Health Indicator Evaluation

In the first dataset, DS1, an increase in HI with damage level is observed at all speeds. The alarm level (HI=1) is calculated with $P_{FA} = 10^{-6}$. Boxplots of the calculated HI from VAEs and a CVAE are shown in Fig. E.10. Whiskers are set to 2.5th and 97.5th percentile. In the following discussion, the median (orange line inside boxes) is considered as the HI value. In dataset 1, HI at DL2 exceeds the warning level 0.75 in all speeds except at 150 rpm for VAE (HI=0.63) and 200 rpm for CVAE (HI=0.69). Data for DL2 was not recorded at 50 rpm. At DL2, the HI exceeds alarm value of 1 at all speeds. Results from VAE and CVAE differ more as damage level increases, but the overall results are well aligned with an increase in HI with damage level at all rpms.

The HI calculated for dataset 2 with VAE and CVAE is shown in Fig. E.11. At 30 rpm, the HI is above the alarm level from DL2. However, there is no monotonic increase in HI level with damage level. Still, this result should be considered as a clear indication of anomalous behaviour. HI for 60 rpm follow a similar trend, but the HI values are lower, exceeding the warning level in DL3-5 only. As in DS1, the HI values calculated using the standard VAE and CVAE are very similar.

Compared to dataset 1, there is less consistency in the HI with increasing HI, and larger differences between speeds. The inconsistency between damage levels may be caused by removing the bearing for applying damage. This procedure introduces differences in the mechanical assembly that may affect the results. Also, the damage was applied manually,

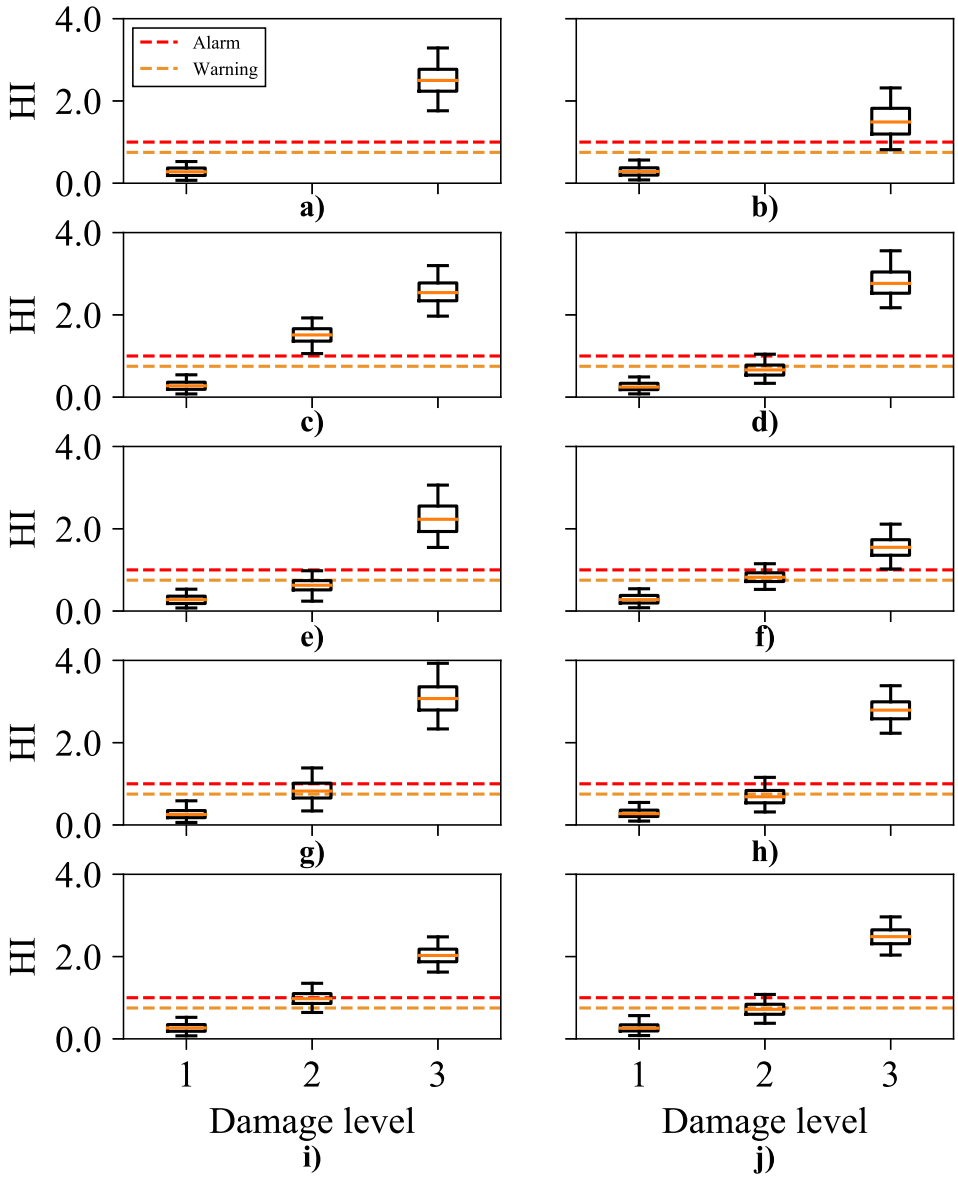


Figure E.10: HI for dataset 1. Left column: Standard VAE. Right column: Conditional VAE. **a-b)** 50 rpm, **c-d)** 100 rpm, **e-f)** 150 rpm, **g-h)** 200 rpm, **i-j)** 250 rpm. Whiskers are set to 2.5th and 97.5th percentile.

which gives room for more variations between damage levels. Finally, data for increasing speeds were recorded consecutively. The seeded damage may therefore be smoothed over time during acquisition. This is a possible explanation for the differences between 30 rpm and 60 rpm. If the smoothing effect differs with damage severity, this will also contribute to the HI inconsistency between damage levels. Further, higher speed may generate high energy frequency components, which dominate the AR coefficients but are not associated with the bearing damage.

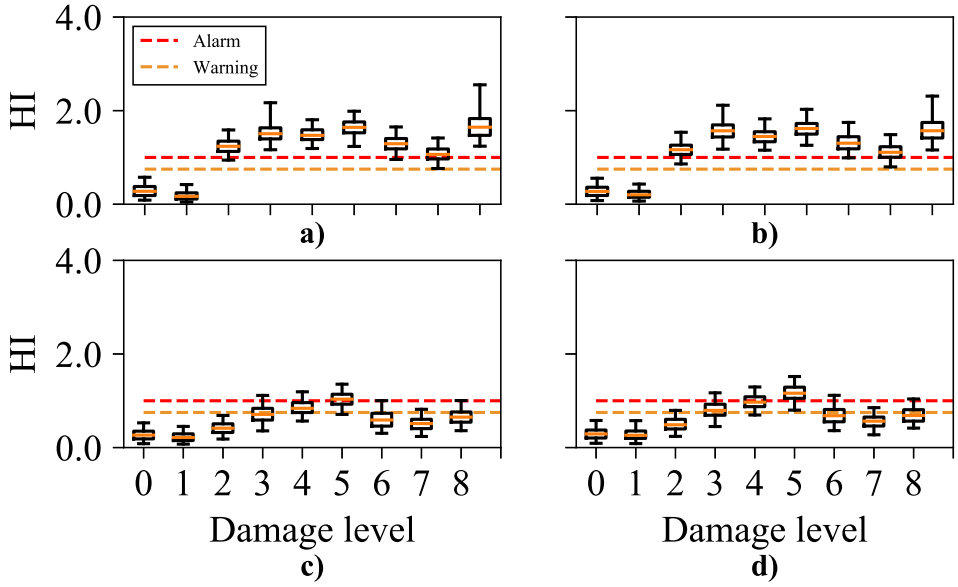


Figure E.11: HI for dataset 2. Left column: standard VAE. Right column: Conditional VAE. **a-b)** 30 rpm, **c-d)** 60 rpm. Whiskers are set to 2.5th and 97.5th percentile.

E.4.2 Model Properties

A summary of final training, validation and test losses for the VAE are shown in Table E.4, including the median values for the 5 models. The ability of latent representations carrying useful information is measured by the reconstruction loss $\mathcal{L}_R$. Examining $\mathcal{L}_R$ in Table E.4 reveals that the value is correlated with model order p, which is expected from the square error summation per datapoint in (E.3). The reconstructed AR coefficients for DS1-100 are shown in Fig. E.12. This speed has the lowest number of features ($p = 12$) in the dataset, and also the lowest reconstruction loss. Still, we observe that reconstructions of coefficient 8 and 9 are skewed. It is likely that further tuning of hyperparameters such as hidden layer size, number of hidden layers and latent dimension can improve reconstruction, but a systematic investigation of parameters search was not performed due to the associated computational cost of training.

The statistical properties of the HI assumes that $q(\mathbf{z})$ approximates a multivariate standard Gaussian distribution $p(\mathbf{z}) \sim \mathcal{N}(0, \mathbf{I})$. The Gaussian latent space is imposed by $\mathcal{L}_{KL}$, which takes values between 1.838 (DS1-250) and 2.106 (DS1-50) in the test dataset. The loss values are more stable than $\mathcal{L}_R$, as the latent dimension J is constant.

However, $\mathcal{L}_{KL}$ describes the mean KL divergence of each datapoint rather than the aggregated distribution of $q(\mathbf{z})$. Therefore, the KL divergence D_{KL} between the aggregated distribution (after sampling) and $p(\mathbf{z})$ are calculated as in (E.6). The value is bounded to $D_{KL} \geq 0$, and a value of zero means that $q(\mathbf{z})$ and $p(\mathbf{z})$ are identical distributions. In the test datasets, D_{KL} takes values between 0.007 (DS2-30) and 0.112 (DS2-30). Fig.

Table E.4: Final loss values after training of the VAEs.

	$\mathcal{L}_R$			$\mathcal{L}_{KL}$			KL_D		
	train	val	test	train	val	test	train	val	test
DS1-50	20.668	23.130	22.607	3.484	3.266	2.106	0.033	0.017	0.025
DS1-100	4.432	5.375	5.400	2.765	2.550	1.950	0.033	0.033	0.050
DS1-150	9.814	10.397	11.623	3.205	2.991	2.085	0.038	0.044	0.021
DS1-200	6.987	8.778	9.189	3.001	2.853	2.071	0.028	0.091	0.112
DS1-250	5.328	6.090	6.215	3.178	2.739	1.838	0.020	0.067	0.054
DS2-30	5.255	5.670	5.569	2.844	2.803	2.099	0.050	0.010	0.007
DS2-60	7.418	7.658	7.236	3.586	3.452	1.943	0.041	0.019	0.017

Table E.5: Final loss values after training of the CVAE.

	$\mathcal{L}_R$			$\mathcal{L}_{KL}$			KL_D		
	train	val	test	train	val	test	train	val	test
DS1	14.453	14.080	13.977	2.910	2.825	1.988	0.068	0.007	0.008
DS2	5.302	6.303	8.389	2.702	3.015	2.748	0.038	0.005	0.020

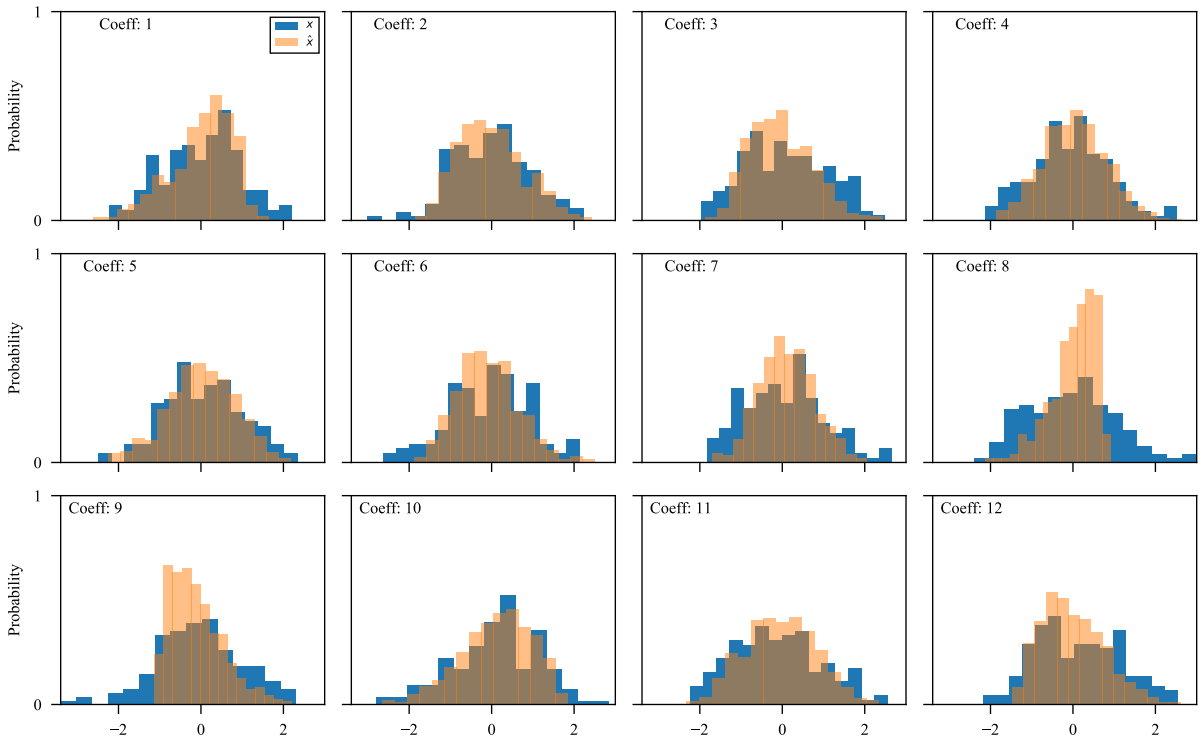


Figure E.12: Histogram of test data $\mathbf{x}$ and reconstructions $\hat{\mathbf{x}}$ for the 12 coefficients in DS1-100.

E.13 shows histograms of each dimension of $\mathbf{z}$ for DS1-100, which has $D_{KL} = 0.05$. A qualitative evaluation confirms that it approximates a Gaussian distribution.

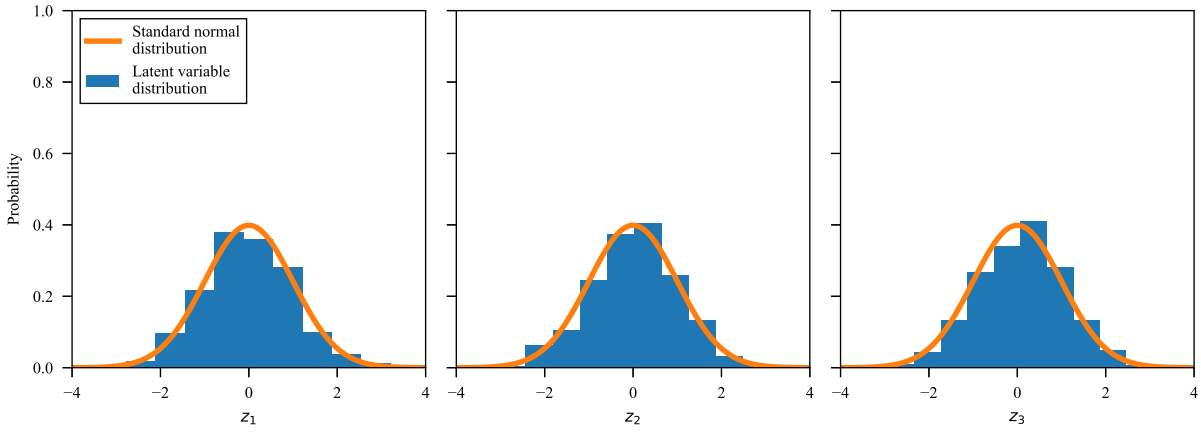


Figure E.13: Histogram of the aggregated posterior $\mathbf{z} = [z_1, z_2, z_3]$ for the test data of DS1-100.

Table E.5 lists $\mathcal{L}_R$, $\mathcal{L}_{KL}$ and KL_D from the network trained as CVAEs, where all speeds in the dataset are used simultaneously in training. $\mathcal{L}_R$ in the test datasets is higher than the average for the separate speeds in Table E.4. This is reasonable, as the same number of neurons in the network must learn to reconstruct data from 5 and 2 rpms in DS1 and DS2 respectively, instead of just one. However, we see that values for $\mathcal{L}_{KL}$ and KL_D are similar to the VAE. This indicates that the assumption of a Gaussian latent space is valid for the CVAE as well.

E.5 Conclusions

This paper proposes a method for unsupervised learning of a Health Indicator (HI), aiming to detect defects in large, slow-rotating axial bearings, by performing variational inference using a variational autoencoder (VAE) and a conditional variational autoencoder (CVAE). Within the framework, coefficients from autoregressive (AR) models were used for both vibration and acoustic emission measurements. The proposed method is proven to be effective using both vibration and acoustic emission (AE) measurements. Using vibration measurements, as opposed to acoustic emission data, allows the proposed method to be cost-effective. In contrast, the previous work of dataset 1 was not able to reveal any degradation of the bearing using vibration measurements. The HI calculated from AE data in dataset 2 was less consistent with the applied damage. However, the experimental design may have had an impact on the calulated HI, in particular at 60 rpm. In both datasets, the proposed method was able to uncover and quantify a significant change in machine operation through the HI. The possibility to calibrate the HI to a desired level

of Probability of False Alarm (PFA) allows the alarm setting to adapt to the criticality of the equipment.

Challenges of detecting defects on axial, large bearings at low speeds were discussed in this study. The effectiveness of the proposed method for axial bearing fault detection at low speeds is validated by data from 2 test rigs. As the proposed method does not rely on detection of fault frequencies, changes in machine operation can be detected regardless of failure mode and fault location. In future studies, the methodology can be extended to include other types of feature input, such as time series data. Other target distributions can also be investigated, in conjunction with tuning of hyperparameters for improving reconstruction while maintaining the desired distribution. The HI is capable of capturing changes in the condition of the axial bearing, so a logical next step is to incorporate it in prognostics and remaining useful life estimation.

Acknowledgment

The research presented in this paper has received funding from the Norwegian Research Council, SFI Offshore Mechatronics, Project Number 237896.

References

[1] N. Tandon and A. Choudhury. A Review of Vibration and Acoustic Measurement Methods for the Detection of Defects in Rolling Element Bearings. *Tribology International*, 32(8):469–480, 1999. doi:https://doi.org/10.1016/s0301-679x(99)00077-8.

[2] R. B. Randall and J. Antoni. Rolling Element Bearing Diagnostics - A Tutorial. *Mechanical Systems and Signal Processing*, 25(2):485–520, 2011. doi:10.1016/j.ymssp.2010.07.017.

[3] M. Cerrada, R.-V. Sánchez, C. Li, F. Pacheco, D. Cabrera, J. Valente de Oliveira, and R. E. Vásquez. A Review on Data-driven Fault Severity Assessment in Rolling Bearings. *Mechanical Systems and Signal Processing*, 99:169–196, 2018. doi:10.1016/J.YMSSP.2017.06.012.

[4] Z. Stamboliska, E. Rusiński, and P. Moczko. *Proactive Condition Monitoring of Low-speed Machines*. Springer International Publishing, 1 edition, 2014. doi:10.1007/978-3-319-10494-2.

[5] E. Bechhoefer, R. Schlanbusch, and T. I. Waag. Fault Detection on Large Slow Bearings. In *European Conference of the Prognostics and Health Management Society 2016*, volume 7, pages 1–8, 2016.

[6] Y. Pan, R. Hong, J. Chen, Z. Qin, and Y. Feng. Incipient Fault Detection of Wind Turbine Large-size Slewing Bearing Based on Circular Domain. *Measurement*, 137:130–142, 2019. doi:10.1016/J.MEASUREMENT.2019.01.033.

[7] N. Jamaludin and D. Mba. Monitoring Extremely Slow Rolling Element Bearings: Part I. *NDT and E International*, 35(6):349–358, 2002. doi:10.1016/S0963-8695(02)00006-3.

[8] N. Jamaludin and D. Mba. Monitoring Extremely Slow Rolling Element Bearings: Part II. *NDT and E International*, 35(6):359–366, 2002. doi:10.1016/S0963-8695(02)00006-3.

[9] B. V. Hecke, J. Yoon, and D. He. Low Speed Bearing Fault Diagnosis Using Acoustic Emission Sensors. *Applied Acoustics*, 105:35–44, 2016. doi:10.1016/j.apacoust.2015.10.028.

[10] W. Caesarendra, B. Kosasih, A. K. Tieu, H. Zhu, C. A. S. Moodie, and Q. Zhu. Acoustic Emission-based Condition Monitoring Methods: Review and Application for Low Speed Slew Bearing. *Mechanical Systems and Signal Processing*, 72-73:134–159, 2016. doi:10.1016/J.YMSSP.2015.10.020.

[11] Z. Mo, J. Wang, H. Zhang, X. Zeng, H. Liu, and Q. Miao. Vibration and Acoustics Emission Based Methods in Low-speed Bearing Condition Monitoring. In *2018 Prognostics and System Health Management Conference (PHM-Chongqing)*, pages 871–875, 2018. doi:10.1109/PHM-Chongqing.2018.00156.

[12] D. Abboud, J. Antoni, S. Sieg-Zieba, and M. Eltabach. Envelope Analysis of Rotating Machine Vibrations in Variable Speed Conditions: A Comprehensive Treatment. *Mechanical Systems and Signal Processing*, 84:200–226, 2017. doi:10.1016/j.ymssp.2016.06.033.

[13] E. Bechhoefer, R. Schlanbusch, and T. I. Waag. Techniques for Large, Slow Bearing Fault Detection. *International Journal of Prognostics and Health Management*, 7(1):1–12, 2016.

[14] S. Kass, A. Raad, and J. Antoni. Self-running Bearing Diagnosis Based on Scalar Indicator Using Fast Order Frequency Spectral Coherence. *Measurement*, 138:467–484, 2019. doi:10.1016/J.MEASUREMENT.2019.02.046.

[15] C. Mishra, A. Samantaray, and G. Chakraborty. Rolling Element Bearing Fault Diagnosis Under Slow Speed Operation using Wavelet De-noising. *Measurement*, 103:77–86, 2017. doi:10.1016/J.MEASUREMENT.2017.02.033.

[16] H. Nguyen, J. Kim, J.-M. Kim, H. N. Nguyen, J. Kim, and J.-M. Kim. Optimal Subband Analysis Based on the Envelope Power Spectrum for Effective Fault Detection in Bearing Under Variable, Low Speeds. *Sensors*, 18(5):1389, 2018. doi:10.3390/s18051389.

[17] Q. Xiong, Y. Xu, Y. Peng, W. Zhang, Y. Li, and L. Tang. Low-speed Rolling Bearing Fault Diagnosis Based on EMD Denoising and Parameter Estimate With Alpha Stable Distribution. *Journal of Mechanical Science and Technology*, 31(4):1587–1601, 2017. doi:10.1007/s12206-017-0306-y.

[18] R. Liu, B. Yang, E. Zio, and X. Chen. Artificial Intelligence for Fault Diagnosis of Rotating Machinery: A Review. *Mechanical Systems and Signal Processing*, 108:33–47, 2018. doi:10.1016/J.YMSSP.2018.02.016.

[19] M. Saimurugan, K. Ramachandran, V. Sugumaran, and N. Sakthivel. Multi Component Fault Diagnosis of Rotational Mechanical System Based on Decision Tree and Support Vector Machine. *Expert Systems with Applications*, 38(4):3819–3826, 2011. doi:10.1016/J.ESWA.2010.09.042.

[20] L. Song, H. Wang, and P. Chen. Vibration-based Intelligent Fault Diagnosis for Roller Bearings in Low-speed Rotating Machinery. *IEEE Transactions on Instrumentation and Measurement*, 67(8):1887–1899, 2018. doi:10.1109/TIM.2018.2806984.

[21] M. Kang, J. Kim, J.-M. Kim, A. C. C. Tan, E. Y. Kim, and B.-K. Choi. Reliable Fault Diagnosis for Low-speed Bearings Using Individually Trained Support Vector Machines With Kernel Discriminative Feature Analysis. *IEEE Transactions on Power Electronics*, 30(5):2786–2797, 2015. doi:10.1109/TPEL.2014.2358494.

[22] A. Soualhi, K. Medjaher, and N. Zerhouni. Bearing Health Monitoring Based on Hilbert–huang Transform, Support Vector Machine, and Regression. *IEEE Transactions on Instrumentation and Measurement*, 64(1):52–62, 2015. doi:10.1109/TIM.2014.2330494.

[23] J. Tian, C. Morillo, M. H. Azarian, and M. Pecht. Motor Bearing Fault Detection Using Spectral Kurtosis-based Feature Extraction Coupled With K-nearest Neighbor Distance Analysis. *IEEE Transactions on Industrial Electronics*, 63(3):1793–1803, 2016. doi:10.1109/TIE.2015.2509913.

[24] P. Nguyen, M. Kang, J. Kim, and J.-M. M. Kim. Reliable Fault Diagnosis of Low-speed Bearing Defects Using a Genetic Algorithm. In *PRICAI 2014: Trends in Artificial Intelligence 2014*, pages 248–255. Springer, Cham, 2014. doi:10.1007/978-3-319-13560-1.

[25] M. Hemmer, H. Van Khang, K. G. Robbersmyr, T. I. Waag, and T. J. J. Meyer. Fault Classification of Axial and Radial Roller Bearings Using Transfer Learning Through a Pretrained Convolutional Neural Network. *Designs*, 2(4), 2018. doi:10.3390/designs2040056.

[26] D. Peng, Z. Liu, H. Wang, Y. Qin, and L. Jia. A Novel Deeper One-dimensional CNN With Residual Learning for Fault Diagnosis of Wheelset Bearings in High-speed Trains. *IEEE Access*, 7:10278–10293, 2019. doi:10.1109/ACCESS.2018.2888842.

[27] M. M. Islam and J.-M. Kim. Automated Bearing Fault Diagnosis Scheme Using 2d Representation of Wavelet Packet Transform and Deep Convolutional Neural Network. *Computers in Industry*, 106:142–153, 2019. doi:10.1016/J.COMPIND.2019.01.008.

[28] J. Xie, G. Du, C. Shen, N. Chen, L. Chen, and Z. Zhu. An End-to-end Model Based on Improved Adaptive Deep Belief Network and Its Application to Bearing Fault Diagnosis. *IEEE Access*, 6:63584 – 63596, 2018. doi:10.1109/ACCESS.2018.2877447.

[29] Z. Chen and W. Li. Multisensor Feature Fusion for Bearing Fault Diagnosis using Sparse Autoencoder and Deep Belief Network. *IEEE Transactions on Instrumentation and Measurement*, 66(7), 2017. doi:10.1109/TIM.2017.2669947.

[30] M. Hemmer and T. I. Waag. A Comparison of Acoustic Emission and Vibration Measurements for Condition Monitoring of an Offshore Drilling Machine. In *Proceedings of the Annual Conference of the Prognostics and Health Management Society*, pages 278–285, 2017.

[31] E. Bechhoefer and A. P. F. Bernhard. A Generalized Process for Optimal Threshold Setting in Hums. In *IEEE Aerospace Conference Proceedings*, pages 1–9, 2007. doi:10.1109/AERO.2007.352867.

[32] E. Bechhoefer, D. He, and P. Dempsey. Gear Health Threshold Setting Based On a Probability of False Alarm. In J. R. Celaya, S. Saha, and A. Saxena, editors, *Annual Conference of the Prognostics and Health Management Society*, pages 275–281, 2011.

[33] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995. doi:10.1111/j.2517-6161.1995.tb02031.x.

[34] G. F. Hughes. On the Mean Accuracy of Statistical Pattern Recognizers. *IEEE Transactions on Information Theory*, 1968. arXiv:0-387-31073-8, doi:10.1109/TIT.1968.1054102.

[35] S. Tao, T. Zhang, J. Yang, X. Wang, and W. Lu. Bearing Fault Diagnosis Method Based on Stacked Autoencoder and Softmax Regression. In *Chinese Control Conference, CCC*, 2015. doi:10.1109/ChiCC.2015.7260634.

[36] H. Shao, H. Jiang, H. Zhao, and F. Wang. A Novel Deep Autoencoder Feature Learning Method for Rotating Machinery Fault Diagnosis. *Mechanical Systems and Signal Processing*, 95:187–204, 2017. doi:10.1016/j.ymssp.2017.03.034.

[37] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu. Stacked Sparse Autoencoder-based Deep Network for Fault Diagnosis of Rotating Machinery. *IEEE Access*, 5, 2017. doi:10.1109/ACCESS.2017.2728010.

[38] C. Lu, Z. Y. Wang, W. L. Qin, and J. Ma. Fault Diagnosis of Rotary Machinery Components using a Stacked Denoising Autoencoder-based Health State Identification. *Signal Processing*, 130:377–388, 2017. doi:10.1016/j.sigpro.2016.07.028.

[39] M. . A. Ranzato, Y.-L. Boureau, and Y. Lecun. Sparse Feature Learning for Deep Belief Networks. In *Advances in neural information processing systems*, pages 1185–1192, 2008.

[40] D. P. Kingma and M. Welling. Auto-encoding Variational Bayes. *arXiv preprint*, 2014. arXiv:1312.6114v10.

[41] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014. arXiv:arXiv:1406.2661v1.

[42] A. Makhzani, J. Shlens, N. Jaitly, G. Brain, I. G. Openai, and B. Frey. Adversarial Autoencoders. *arXiv preprint*, 2016. arXiv:1511.05644v2.

[43] S. Kullback and R. A. Leibler. On Information and Sufficiency. Technical Report 1, 1951.

[44] M. Arjovsky and L. Bottou. Towards Principled Methods for Training Generative Adversarial Networks. *arXiv preprint*, 2017. arXiv:1701.04862v1.

[45] G. San Martin, E. López Droguett, V. Meruane, and M. das Chagas Moura. Deep Variational Auto-encoders: A Promising Tool for Dimensionality Reduction and Ball Bearing Elements Fault Diagnosis. *Structural Health Monitoring*, 2018. doi:10.1177/1475921718788299.

[46] Mart\'\in˜Abadi, Ashish˜Agarwal, Paul˜Barham, Eugene˜Brevdo, Zhifeng˜Chen, Craig˜Citro, Greg˜S.˜Corrado, Andy˜Davis, Jeffrey˜Dean, Matthieu˜Devin, Sanjay˜Ghemawat, Ian˜Goodfellow, Andrew˜Harp, Geoffrey˜Irving, Michael˜Isard, Y. Jia, Rafal˜Jozefowicz, Lukasz˜Kaiser, Manjunath˜Kudlur, Josh˜Levenberg,

Dandelion˜Mané, Rajat˜Monga, Sherry˜Moore, Derek˜Murray, Chris˜Olah, Mike˜Schuster, Jonathon˜Shlens, Benoit˜Steiner, Ilya˜Sutskever, Kunal˜Talwar, Paul˜Tucker, Vincent˜Vanhoucke, Vijay˜Vasudevan, Fernanda˜Viégas, Oriol˜Vinyals, Pete˜Warden, Martin˜Wattenberg, Martin˜Wicke, Yuan˜Yu, and Xiaoqiang˜Zheng. Tensorflow: Large-scale Machine Learning on Heterogeneous Systems, 2015. Software available from tensorflow.org. URL: `https://www.tensorflow.org/`.

[47] K. Gairaa, A. Khellaf, Y. Messlem, and F. Chellali. Estimation of the Daily Global Solar Radiation Based on Box-jenkins and ANN Models: A Combined Approach, 2016. `doi:10.1016/j.rser.2015.12.111`.

[48] G. U. Yule. Vii. On a Method of Investigating Periodicities Disturbed Series, With Special Reference to Wolfer's Sunspot Numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226(636-646):267–298, 1927. `doi:10.1098/rsta.1927.0007`.

[49] S. G. T. Walker. On Periodicity in Series of Related Terms. *Proceedings of the Royal Society of London*, 131(818):518–532, 1931. `doi:10.1098/rspa.1931.0069`.

[50] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung. *Time Series Analysis : Forecasting and Control.* John Wiley & Sons, Inc., 5 edition, 2015.

[51] Y. Vardanyan and M. R. Hesamzadeh. Modeling Regime Switching in Day-ahead Market Prices using Markov Model. In *IEEE PES Innovative Smart Grid Technologies Conference Europe*. IEEE Computer Society, 2017. `doi:10.1109/ISGTEurope.2016.7856316`.

Paper F

Simulation-driven Deep Classification of Bearing Faults from Raw Vibration Data

Martin Hemmer, Andreas Klausen, Huynh Van Khang, Kjell G. Robbersmyr and Tor I. Waag

This paper has accepted for publication as:

Martin Hemmer, Andreas Klausen, Huynh van Khang, Kjell G. Robbersmyr, and Tor I. Waag. Simulation-driven Deep Classification of Bearing Faults from Raw Vibration Data. Accepted for publication in: *International Journal of Prognostics and Health Management (IJPHM)*. Accepted date: 15.08.2019.

Simulation-driven Deep Classification of Bearing Faults from Raw Vibration Data

Martin Hemmer*,**, Andreas Klausen*, Huynh Van Khang*, Kjell G. Robbersmyr* and Tor I. Waag***

*University of Agder
Department of Engineering Sciences
Jon Lilletunsvei 9, N-4879 Grimstad, Norway

**MHWirth AS AS
Butangen 21, N-4639 Kristiansand, Norway

***NORCE Norwegian Research Centre AS
Jon Lilletuns vei 9 H, N-4877 Grimstad, Norway

Abstract – The industry is moving towards maintenance strategies that consider component health, which require extensive collection and analysis of data. Condition monitoring methods that require manual feature extraction and analysis, become infeasible on an industrial scale. Machine learning algorithms can be used to automatically detect and classify faults, however, obtaining sufficient data for training is required for deep learning and other data-driven classification approaches. Data from healthy machine operation is generally available in abundance, while data from representative fault- and operating conditions is limited. This limits both development and deployment of deep learning-based CM systems on an industrial scale. This paper addresses both the challenges of automated analysis and lack of training data. A deep learning classifier architecture utilizing 1-dimensional dilated convolutions is proposed. Dilation of the convolution kernel allows for analysis of raw vibration signals while simultaneously maintaining the receptive field of the classifier enough to capture temporal patterns. The proposed method performs classification in time domain on signal segments of 1 second or shorter. With knowledge of the bearing specification, artificial vibration signals with similar characteristics as an actual bearing fault can be created. In this work, generated fault signals are combined with healthy operational data to obtain training data for a deep classifier. Parameters of the vibration model is chosen as distributions rather than fixed values. By using a range parameters in the vibration model, the classifier learns to recognize temporal features from the training data that generalize to unseen data. The effectiveness of the proposed method is demonstrated by training classifiers on generated data and testing on real signals from faulty bearings at both low and high speed. One dataset containing

seeded faults and three run-to-failure tests are used for the demonstration.

F.1 Introduction

Rolling element bearings (REBs) play a fundamental part in most types of rotating machinery by reducing friction, transferring forces and constraining motion. Even under operation within design specification, REBs have a finite lifetime due to fatigue induced by cyclic loading from the moving rolling elements. Expected fatigue lifetime can be estimated accurately for large populations of bearings under identical operating conditions. However, the lifetime of a single bearing might be shorter or longer than the estimated value. Therefore, performing maintenance based on elapsed time or operating hours is not optimal. Shorter maintenance intervals do not utilize component lifetime, while longer intervals may result in unexpected failures and downtime. This motivates the development of condition monitoring (CM) methods that are able to diagnose and evaluate component health. Condition based maintenance (CBM) and prognostics and health management (PHM) maintenance regimes aim to utilize the knowledge of machine health to perform maintenance when required.

While time domain features such as root mean square (RMS), peak-to-peak and kurtosis are useful for fault detection through trending, they cannot be used for diagnosing fault types. Single-point surface defects in bearings can be modeled as periodic excitation of the system resonance frequency [1]. Amplitude demodulation is effective for detecting such defects, and a fast implementation is possible using the Hilbert envelope [2]. To improve the effectiveness of envelope analysis, the signal should be preprocessed. The fast Kurtogram [3] utilizes spectral kurtosis to identify a frequency band for demodulation [4], and order tracking reduce the amount of spectral leakage due to speed variations [5, 3]. Bearing vibration is random in nature [6], and is thus separable from other vibration sources using time synchronous averaging (TSA) [7, 6], cepstral editing [8, 50] and other methods [9]. While these tools are effective for bearing fault diagnosis at any speed, diagnosis of low speed applications is more challenging. There is less energy in the system and the fault signature is easily masked in noise. Characteristic fault frequencies are also closer in absolute frequency, requiring longer acquisition time to achieve a spectral resolution where faults are distinguishable [10]. In any case, a skilled data analyst is required to process data and diagnose the bearing.

With the emergence of industry 4.0, the amount of available sensor data is increasing rapidly [11]. Proper usage of the aforementioned tools requires feature engineering and

skilled manpower, which leaves manual analysis infeasible on a large scale and motivates a more data-driven, automated approach. Deep learning has proven to be well suited to data-driven feature extraction, classification and prognostics for rotating machinery [12]. Autoencoders [13], deep neural networks [14], support vector machines [15], deep belief networks [16, 82], self-organizing maps [17] and convolutional neural networks (CNNs) [18, 76, 77] are all applied to bearing fault detection and diagnostics. Time series analysis using deep learning has applications in speech recognition and modeling [19], translation [20], and audio recognition and generation [21]. In high resolution time series, it is necessary to consider dependencies that are further apart than the kernel size. One strategy is to use larger kernels and add more convolutional and pooling layers, but this is computationally heavy, and the weight training may be difficult due to vanishing gradients. Long short-term memory recurrent networks are also commonly applied in analysis and modeling of speech and audio [22]. The memory of previous states in recurrent network has to propagate through the entire time series, which is a limitation for retaining long-term memory.

However, a common problem is the need for representative fault data during training. To counteract this, it has been proposed to use simulated bearing faults for training classifiers [23, 24, 25]. A wide range of analytical and FEM-based dynamic models for bearing faults exist, taking clearances, elastohydrodynamic lubrication effects, race waviness, defect size and several other parameters into consideration [15]. However, the results presented in this paper show that modeling the impulse responses and pseudo-cyclostationary behavior [26, 3] is sufficient to capture the characteristics of a real vibration signal.

This paper proposes to train a deep convolutional neural network on vibration time series data based on combinations of healthy and simulated fault data. The architecture is based on hierarchically dilated 1D convolutions. A dilated architecture was applied to bearing fault detection in [27]. However, the proposed method in this paper includes low-level feature extraction without dilation, and a global average pooling layer to replace fully connected layers. Additionally, training in [27] was done on actual fault data as opposed to simulated data in this paper. Varying the parameters of the simulation model allows the classifier to learn features that are generalized enough to detect and diagnose actual bearing faults. The approach reduces the need for manual feature extraction and allows the analysts to focus the attention on detected faults for closer evaluation. The proposed method is applied to shorter signal segments of raw vibration data. This is particularly beneficial for low speed bearings that require longer acquisition time for spectral resolution. The proposed method reduces the need for data processing, storage and transfer by performing classification in time domain.

The remaining of the paper is organized as follows: Section F.2 describes the generation and processing of data that leads to the classification. The datasets used in analysis are presented in Section F.3. Analysis results and discussion are provided in Section F.4. Final conclusions are given in Section F.5.

F.2 Methodology

It is assumed that single-point surface defects excite transient pulses of amplitude modulated resonance frequencies of the support structure when internal bearing parts roll over a fault. Modelling a signal requires several assumptions of unknown or uncertain variables. The proposed approach acknowledges this, and generates data from parameters in a range rather than fixed values. Section F.2.1 defines how these uncertainties are used to generate the fault signal.

F.2.1 Bearing Vibration Model

Faults in outer race (OR), inner race (IR) and rolling element (RE), result in periodic impacts with ball pass frequency outer race (BPFO), inner race (BPFI) and ball spin frequency (BSF), respectively. The fundamental train frequency (FTF) is mainly presented as a modulating frequency for rolling element (RE) faults. The nominal period T is the inverse of the characteristic fault frequency of the fault type and is denoted T_{OR}, T_{IR}, T_{FTF} and T_{RE}. Formulas for calculating the periods are provided in Eqs. (F.1) through (F.4). The number of rolling elements is denoted by n_r, d and D are the roller and pitch diameter, respectively, and ϕ is the contact angle.

$$T_{OR} = BPFO^{-1} = \left(\frac{f_r n_r}{2} \left(1 - \frac{d}{D} \cos \phi \right) \right)^{-1} \tag{F.1}$$

$$T_{IR} = BPFI^{-1} = \left(\frac{f_r n_r}{2} \left(1 + \frac{d}{D} \cos \phi \right) \right)^{-1} \tag{F.2}$$

$$T_{FTF} = FTF^{-1} = \left(\frac{f_r}{2} \left(1 - \frac{d}{D} \cos \phi \right) \right)^{-1} \tag{F.3}$$

$$T_{RE} = BSF^{-1} = \left(\frac{f_r D}{2d} \left(1 - \left(\frac{d}{D} \cos \phi \right)^2 \right) \right)^{-1} \tag{F.4}$$

In reality, the impact period is slightly random due to slip, typically around have 1-2 % [28]. By considering slip as a random process without memory, the bearing signal becomes pseudo-cyclostationary. The slip is accounted for by modeling time between impacts as

a random variable $\varepsilon \sim \mathcal{N}(0, \sigma_T^2)$, where $3\sigma_T = 0.02$ [29]. This keeps the impact periods within approximately $\pm 2\%$ of the nominal period. Thus, the k-th impact period is simply calculated as in Eq. (F.5).

$$T_k = T(1 + \varepsilon) \tag{F.5}$$

In a discrete time series, the sample number of the K-th impact, n_K, is calculated as the cumulative sum of previous impact periods multiplied by the sample frequency, and rounded to the nearest integer with the $nint\,(\cdot)$ operator. The location of the first impact n_0 is determined by sampling a uniform distribution for the first impact time T_0, where $T_0 \sim U\,(0, T)$. Then, the nearest corresponding sample is calculated as $n_0 = nint\{F_s T_0\}$, where F_s is the sample frequency.

$$n_k = n_0 + nint\{F_s f_r \sum_{k=1}^{K-1} T_k\} \tag{F.6}$$

The pulse amplitude is also considered a random variable, where the randomness is modeled by $\xi \sim \mathcal{N}\left(0, \sigma_p^2\right)$. The discrete pulse train $p[n]$ is created by setting $p[n_k] = 1 + \xi$ for $k \in [0, K)$ pulses and 0 elsewhere. IR and RE faults have additional amplitude modulation as the fault passes through the load zone. This phenomenon is modeled by multiplying a periodic function with the pulse train. As the modulation index m_1 is unknown, it is sampled from an interval for generation of each time series. Rolling element faults typically appear at $2 \cdot BSF$, as the fault strike both the inner and outer per roller revolution. This also creates additional amplitude modulation at BSF with modulation factor m_2. The impulse response is modeled as bandpass-filtered white Gaussian noise $w_{bp}[n]$, with center frequency f_c and bandwidth bw. A Butterworth filter of order 5 is used to make the band-pass filter. The impulse response modulation function for a time series is obtained by convolving the pulse train with a window function $h[l]$ where $l \in [0, L)$, where the pulse is given in Eq. (F.7). The number of samples L are determined by sampling $L = nint(F_s t_h)$, where $t_h \sim U(0.5\,\text{ms}, 5\,\text{ms})$ is a uniformly distributed variable to model different pulse durations.

$$h[l] = e^{-5l/L} \tag{F.7}$$

The resulting generated fault signal x_F is given as in Eq. (F.8).

$$x_F = \left((m_1 + m_2)\, p * h\right) w_{bp} \tag{F.8}$$

F.2.2 Data Preprocessing

The following section describes how the generated fault signal x_F and healthy signal x_{HE} are processed before used for training and testing. Each dataset consists of N records

with duration t_r. In this paper, the records are either used directly, or segmented using a rectangular rolling window with duration t_w and a stride of t_s. Depending on the available data, either complete samples or shorter segments can be extracted. Because the simulated data is combined with real, healthy data, it is necessary to know which records are healthy. In seeded fault datasets this information is available, but in run-to-failure tests, true condition is not known. This is solved by using the first N_{HE} records as a reference for healthy condition. As this paper utilizes data that has also been analyzed by other researchers, it is possible to choose N_{HE} small enough to be confident that the bearing is actually healthy. All records selected as healthy are normalized to unit RMS. From the normalized healthy data, the desired number of samples are drawn with replacement for used in training. As the fault data is simulated, it is possible to generate an arbitrary amount of training data. Table F.1 shows the number of generated samples for each dataset. Healthy data use for fault simulation was drawn randomly with replacement.

		CWRU	**IMS**	**UiA**
HE	Real	967	256	920
HE	Simulated	2048	2048	2048
IR	Simulated	2048	2048	2048
RE	Simulated	2048	2048	2048
OR	Simulated	2048	2048	2048

Table F.1: Number of training records

In the seeded fault datasets, all training and test data was normalized to unit RMS to let the network learn patterns from data in a predictable range. First, the simulated faulty data x_F is normalized with a factor $\lambda\,\mathrm{RMS}\{x_F\}$ to model differences in damage severity. Each segment, x_{HE}, is augmented with additive white noise $w \sim \mathcal{N}(0, 0.05^2)$ and normalized to unit RMS. The sum of these signals is denoted x'_S, as shown in Eq. (F.9). This signal is again normalized to unit RMS, as shown in Eq. (F.10).

$$x'_S = \frac{x_{HE} + w}{\mathrm{RMS}\{x_{HE} + w\}} + \frac{x_F}{\lambda\,\mathrm{RMS}\{x_F\}} \tag{F.9}$$

$$x_S = \frac{x'_S}{\mathrm{RMS}\{x'_S\}} \tag{F.10}$$

In run-to-failure datasets, all records are normalized with the mean RMS of the N_{HE} reference records. Again, each healthy segment, x_{HE}, is augmented with additive white noise $w \sim \mathcal{N}(0, 0.05^2)$ and normalized to the RMS of the original segment. The simulated faulty data x_F is also normalized with the same factor $\lambda\,\mathrm{RMS}\{x_F\}$. This time, the two parts are summed directly to obtained the simulated signal x_S, as shown in Eq. (F.11).

This approach also lets the classifier associate higher RMS with a fault, and was included to help the classifier discriminate noise originating from increasingly severe faults and background noise in healthy records. Figure F.1 shows the components of a generated rolling element fault and an actual fault signal.

$$x_S = \frac{(x_{HE} + w)\,\mathrm{RMS}\{x_{HE}\}}{\mathrm{RMS}\{x_{HE} + w\}} + \frac{x_F}{\lambda\,\mathrm{RMS}\{x_F\}} \tag{F.11}$$

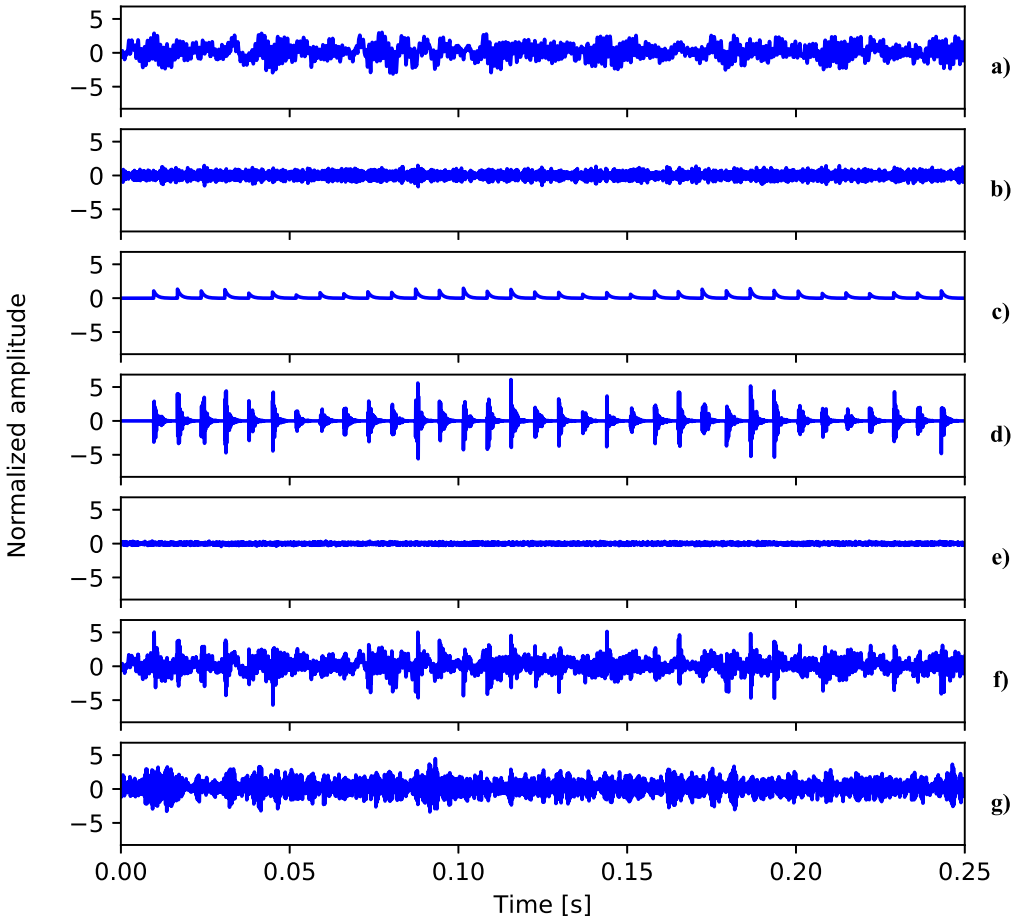


Figure F.1: Components of generated rolling element fault compared to actual fault signal. **a)** Healthy signal x_{HE}. **b)** Band-pass filtered noise w_{bp}. **c)** Modulating envelope. **d)** Generated fault impacts x_p. **e)** Additive noise w. **f)** Generated fault signal x_s. **g)** Actual fault signal.

F.2.3 Classification Network

The classifier is aimed at raw vibration time series input from one or more channels. In this section, we consider single-channel measurements for simplicity. The network consists of three main parts: A low-level feature extraction part, an intermediate feature extraction part with dilated convolutions for capturing long term feature dependencies. Finally, a classification part weights the features and predicts an output. This hierarchical configuration allows the network to first learn long-term relationships with less computational demand. The network configuration and parameters used in this paper are given in Table 3.10. Development was done in Keras with Tensorflow backend [30], therefore Keras layer names are used in the table.

No.	Type	Filters	Size	Stride	Padding	Dilation	Activation
0	Input						
1	Conv1D	32	11	1	valid	1	ReLU
2	MaxPooling1D		3	2			
3	BatchNormalization						
4	Conv1D	32	5	1	valid	1	ReLU
5	MaxPooling1D		3	2			
6	BatchNormalization						
7	Conv1D	8	5	1	same	1	ReLU
8	Conv1D	8	5	1	same	2	ReLU
9	Conv1D	8	5	1	same	4	ReLU
10	Conv1D	8	5	1	same	8	ReLU
11	Conv1D	8	5	1	same	16	ReLU
12	Conv1D	8	5	1	same	32	ReLU
13	Conv1D	8	5	1	same	64	ReLU
14	Conv1D	8	5	1	valid	128	ReLU
15	GlobalAveragePooling1D						
16	Dense	4					Softmax

Table F.2: Network parameters

The low-level feature extraction part consists of two blocks with same structure, as shown in Figure F.2, each containing a 1D convolutional layer, a max pooling layer and a batch normalization layer. This layer extracts features that are close in time by convolving the signal with a kernel that is small compared to the input. Rectified linear unit (ReLU) activation functions [31] are used throughout the network, except for final class output. Max pooling is used to reduce the dimension while maintaining the most prominent features. Batch normalization speeds up training by reducing internal covariate shift [32]. In this paper, separate classifiers were trained for all channels. Principally, the

network structure can also handle multiple data channels simultaneously, but this was not investigated in the experiments.

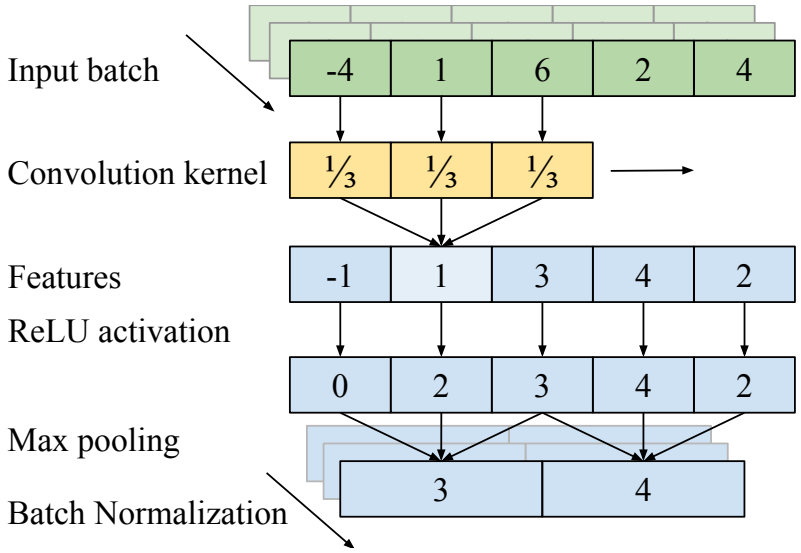


Figure F.2: Low-level feature extraction: Input data points (green) are convolved with 1D kernel (yellow), followed by a ReLU activation function and max pooling layer. The batch normalization is applied to the output of the max pooling.

Dilated convolutions is an efficient way to increase the receptive field of the network. By dilating the convolution kernel, smaller filter sizes can capture long-term dependencies in the data with relatively few layers. The approach has been successful in modeling of high resolution time series [21]. The concept is shown in Figure F.3. Input data is shown in green, active intermediate features in blue, and the output feature in red. By doubling the dilation rate for each new layer, the receptive field grows exponentially.

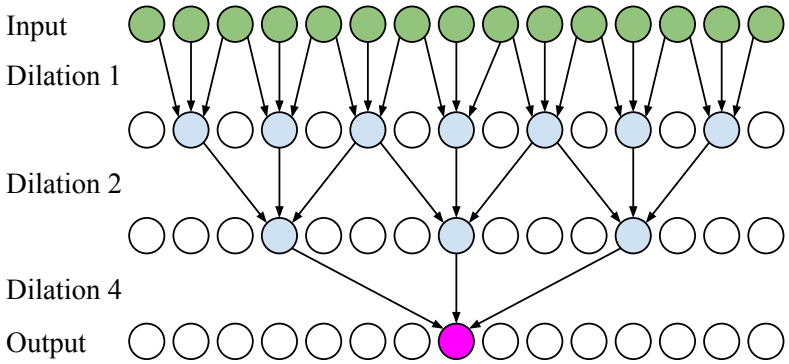


Figure F.3: Intermediate level feature extraction: The receptive field (green) and intermediate features (blue) for computing feature output (pink). Dilation allows a wide receptive field with few layers.

One-dimensional global average pooling is applied to the output of each filter after

all convolutions. This is in contrast to the fully connected layers, which were typically found in the final layers of a CNN. Global average pooling reduces overfitting and makes classification less sensitive to the temporal translation that is introduced by windowing the time series [33]. The output is then fed directly to the final, fully connected classification layer with softmax activation. The objection function to be optimized is categorical crossentropy loss between predicted and true label. The networks were trained for 10 epochs with batch size of 32, using an Adam optimizer with learning rate $1e^{-4}$, except dataset 2 which used a batch size of 16 due to GPU memory constraints.

There is randomness involved in multiple stages of the training process. Network weights are initialized randomly, so two identical models with different random seeds may yield different results. Random shuffling of training data may also affect results. To improve generalization and reduce the effect of randomness, five folds are created from the available data. Each fold contains all the data, but the distribution of data in training, validation and testing splits are different. Each fold is used to train a separate model, resulting in an ensemble of five classifiers. In cases where a record is split in multiple segments, each classifier outputs a decision per record based on a plurality voting scheme. If the record consists of a single segment, there is no voting in this step. The final ensemble classification is also determined through a plurality vote over the individual decisions.

F.3 Experimental Data

Vibration data from three datasets have been used in this study. Dataset 1 contains healthy (HE) and seeded fault data from IR, RE and OR, used to verify that the methodology is capable of detecting and diagnosing the different fault types. The other datasets contain run-to failure data. The following sections give a brief description of the test rigs and data that are utilized in this paper. Results from reference publications are used as a baseline for evaluating classifier performance.

F.3.1 Dataset 1: Case Western Reserve University (CWRU)

Dataset 1 is provided by the Case Western Reserve University (CWRU) and includes vibration records at the drive end, fan end and foundation of a motor. Both the drive end and fan end bearing were seeded separately with OR, IR and RE faults. Data from four loads and four damage severities are available, with a sample frequency of 12 kHz. Additionally, data sampled at 48 kHz are available for drive end faults only. The CWRU test setup is shown in Figure F.4. The test motor (left) is connected to a dynamometer (right) through an encoder and torque sensor (center) [34].

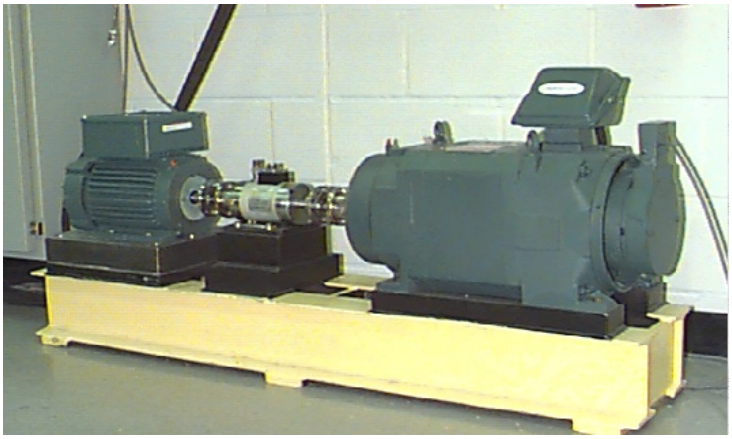


Figure F.4: Case Western Reserve University test rig, used for dataset 1.

In a reference paper for analysis [35], authors applied three fault diagnosis metohds to the data: Envelope analysis of the raw signal, cepstrum prewhitening forllowed by envelope analysis, and envelope analysis of a bandpass filtered signal, where spectral kurtosis was used to calculate the optimal passband. Faults in OR and IR were successfully diagnosed using one or more of the methods. However, the RE faults were not diagnosable. Numerous papers using CWRU data report better performance than the reference paper. A review of deep learning algorithms trained on the CWRU datset show that the majority of research papers report 95-100 % accuracy [36]. However, supervised classification algorithms displaying such high accuracy may be a sign of overfitting and poor generalization performance [35].

This paper uses data from the drive end bearing, running with unloaded motor, sampled at 48 kHz. This was chosen to allow the classifier to capture higher frequency amplitude modulation. The smallest fault size, measuring 0.18 mm in diameter and 0.28 mm in depth, was used for all fault types. Healthy data in only available sampled at 12 kHz for 20 seconds. This record is upsampled by a factor of 4 to obtain 48 kHz healthy data. Faulty bearing records are otherwise 5 seconds long. To obtain more training data, a window of 1 second with a stride of 0.02 seconds is applied to the original record. The files used are listed in table F.3.

	HE	**IR**	**RE**	**OR**
File number	97	109	122	135

Table F.3: Files from CWRU dataset used in this paper.

F.3.2 Dataset 2: Center for Intelligent Maintenance Systems (IMS)

Dataset 2 consists of run-to-failure test data, provided by NSF I/UCR Center for Intelligent Maintenance Systems (IMS). A diagram of the test rig is shown in Figure F.5. The test setup is further described in the reference paper [37]. In this paper, data from test 1 and 2 is used. Test 1 had 8 available channels, however only one channel per bearing was used (channel 1, 3, 5 and 7). Test two was recorded using only one channel per bearing. Data was sampled at 20480 Hz for one second. In both cases, the first 256 records were assumed healthy and used in training. As the available records were only on second, no windowing was performed on this dataset. In test 1, an IR fault was found in bearing

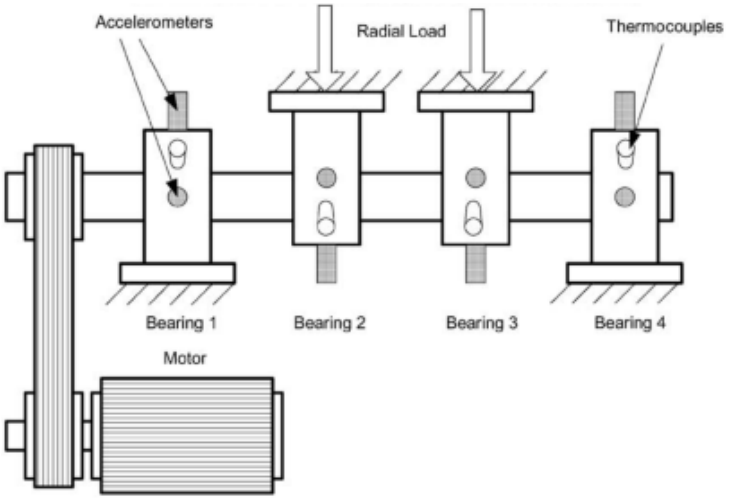


Figure F.5: Schematic drawing of IMS test rig used to collect dataset 2.

3, and an RE fault in bearing 4. An OR fault occurred in bearing 1 in test 2. In both datasets , an increase in RMS and kurtosis is observed towards the end, as shown in Figure F.6. This increase is interpreted as an indication of damage occurrence and progression. In test 1, bearing 4 shows slightly raised kurtosis around record 1435, followed by a large increase from record 1610. Bearing 3 kurtosis increases from record 1800. In test 2, an increase is seen from record 530.

F.3.3 Dataset 3: The In-house Test at University of Agder (UiA)

Dataset 3 was collected using an in-house test rig from the University of Agder as shown in Figure F.7. The test rig was made for run-to-failure tests at low and variable speed conditions under combined radial and axial load. The design and functionality thoroughly described in [38]. Vibration data was collected from an accelerometer perpendicular to

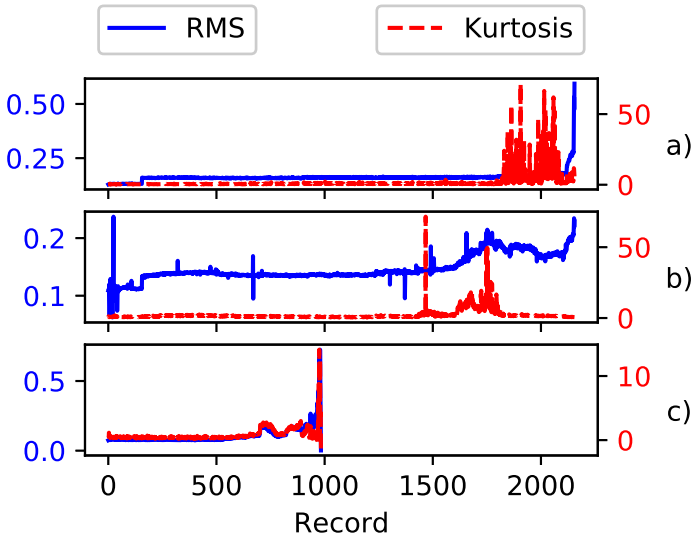


Figure F.6: Plot of RMS and kurtosis for the faulty bearings in dataset 2. **a)** Test 1, bearing 3. **b)** Test 1, bearing 4. **c)** Test 2, bearing 1.

the load zone, sampled at 51200 Hz. Each record contains 100 revolutions. A 2-second window with 1 second stride was applied, resulting in 23 windows per record. 192 records from the last 5 days of testing was used in this paper. Signs of faults were observed from record 163 and onward. Recordings from the first day, 40 records in total, were used for training.

The reference paper [39] reports a pit in the OR, two damaged rollers and a heavily spalled IR at the end of the test. The authors report to observe signs of RE damage first, starting at 29 recordings from the end of life. Later, indications of OR and IR damage were also observed, in that order.

Bearing type and fault frequencies for the datasets are listed in Table F.4.

	CWRU	**IMS**	**UiA**	**Unit**
Make	SKF	Rexnord	SKF	
Model	6205-2RSJEM	ZA-2115	6008-2RS1	
BPFO	107.37	236.38	21.32	Hz
BPFI	162.18	296.90	28.57	Hz
BSF	141.18	139.92	13.86	Hz
FTF	11.92	14.77	1.77	Hz

Table F.4: Bearing specifications for the test datasets

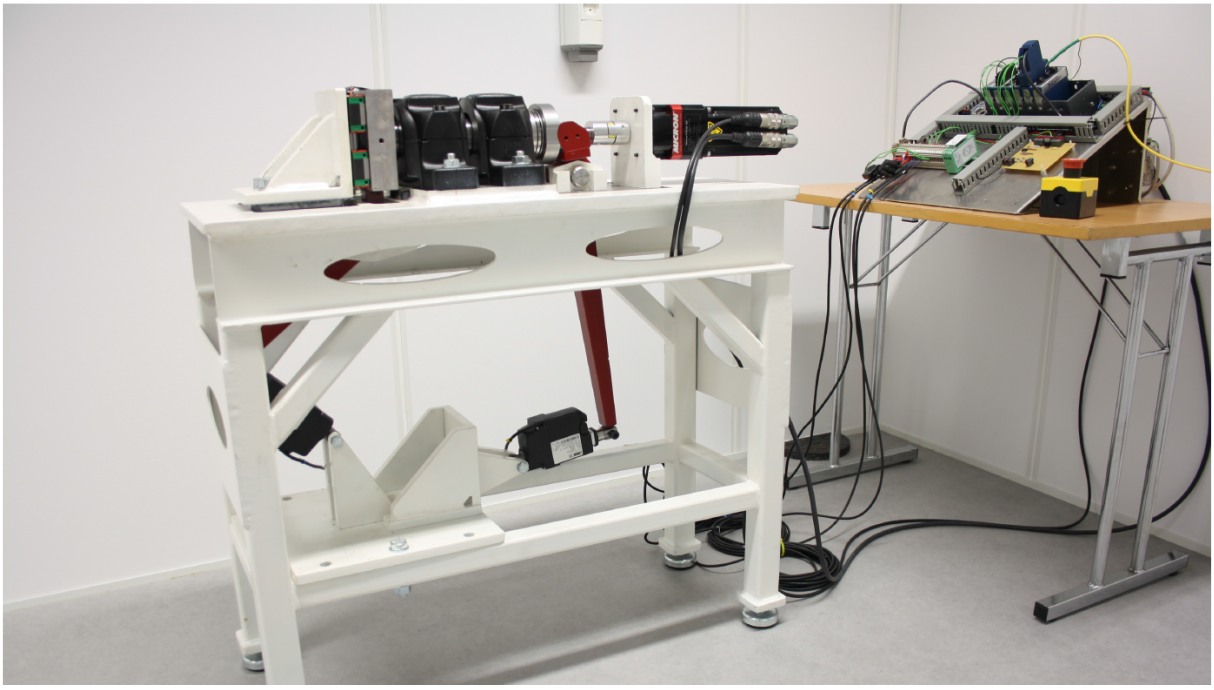


Figure F.7: Test rig used in dataset 3.

F.3.4 Vibration Model Parameters

Fault simulation requires selection of model parameters described in Section F.2.1. Sample frequency F_s, record duration t_r and shaft rate f_s are known system parameters, while window duration t_w and stride t_s are user-defined, limited by sample frequency and the length of available data. For the remaining parameters, optimal values are not known. Therefore, instead of attempting to determine the optimal values, the parameters are defined as either uniform or Gaussian distributions. This papers aims to demonstrate that by defining the distributions wide enough to encompass the assumed true values, a subset of the simulated datapoints will approximate the true faulty condition. Thus, the network is able to classify real, unseen data.

Window duration should be set long enough to capture periodicity from the impacts. The lowest frequency component of interest is typically the FTF, so t_w should at least exceed $1/FTF$. Window length t_w should be set sufficiently long to capture at least a few shaft revolutions, as one would if signal processing were done in the frequency domain. However, it is assumed that longer windows will improve classifier performance. There is no upper limit of t_w, but computational load increase with segment length.

As the test rig in dataset 2 operates at low speed, the window is longer than for the high-speed datasets. Window stride t_s is of less importance, but affects the total number of unique windows. As a rule of thumb, t_s should be selected smaller than t_w to have some overlap of the windows and thus capture more variations of the signal.

The combined center frequency f_c and bandwidth bw must not violate the Nyquist criterion. A conservative limit of $F_s/2.56$ is used. It was found empirically that the impulse response duration t_h should be short enough to not have overlap between impulses in the lower end of the range. A summary of the bearing specifications is shown in Table F.4. Impact arrival time jitter ξ, was set to approximately $\pm 2\%$ to account for slip. Otherwise, the classifier struggles to separate fault types. The upper limit on RMS ratio λ was set to 4 for dataset 1, as it was known that rolling element faults were hard to detect. This did however, not result in successful diagnosis. A low RMS parameter should encourage the network to learn weak signatures, but at the risk of making damage and healthy too similar. Other parameters were set wide across the datasets. A systematic parameter grid search was not performed due to the associated computational load. Table F.5 shows the dataset-specific parameters, while the parameters common for all datasets are listed in Table F.6. Distributions for center frequency and bandwidth are deliberately chosen wide to reduce the chance of achieving good results purely by coincidence. Other parameters are simply set by making a qualified guess, and has not been tuned specifically to improve performance except what is already noted for parameter λ and t_h.

	CWRU	**IMS**	**UiA**	**Unit**
F_s	48	20.48	51.2	kHz
f_s	29.95	33.33	4.17	Hz
t_w	1	1	2	s
t_s	0.02	0.1	1	s
f_c	$U(1.5, 15)$	$U(1.5, 8)$	$U(1.5, 15)$	kHz
bw	$U(0.5, 5)$	$U(0.5, 5)$	$U(0.5, 5)$	kHz
λ	$U(1, 4)$	$U(1, 2)$	$U(1, 2)$	-
N_{HE}	-	256	40	-

Table F.5: Model-specific bearing vibration model parameters.

F.4 Results

This section presents the results of training the classifier on healthy and simulated fault data and testing on actual fault data. For dataset 1, which has seeded faults, performance can be evaluated quantitatively through classification accuracy. In datasets 2 and 3, there is no ground truth available, which makes a quantitative performance evaluation impossible. Classifier outputs are compared to results in reference papers, and evaluated

Parameter	Value	Unit
t_h	$U(0.5, 5)$	ms
ϵ	$\mathcal{N}(0, \sigma_T^2)$	s
$3\sigma_T$	0.02	s
ξ	$\mathcal{N}(0, \sigma_p^2)$	-
$3\sigma_p$	0.1	-
m_1	$U(0.1, 0.5)$	-
m_2	0.25	-

Table F.6: Common vibration model parameters.

qualitatively. All predicted classes shown are the result of a plurality vote in an ensemble of 5 classifiers.

F.4.1 Dataset 1 (CWRU)

Dataset 1 contained samples of single-point faults in OR, IR and OR. The classifier successfully diagnosed the IR and OR fault, but did not manage to diagnose RE faults. Instead, this fault type was consistently misclassified as IR damage by the ensemble. The confusion matrix for is shown in Figure F.8. As seen in the confusion matrix, there were no false alarms, and no damaged bearings were classified as healthy. It's worth noting that in the reference paper, RE faults were not identifiable using any of the applied analysis methods, and did not show the same classical behavior as IR and OR faults [35]. The achieved result of the proposed method is therefore on par with the reference paper. While other deep learning algorithms have been able to diagnose the RE fault as well, the authors of this paper are not aware of any algorithms achieving this with simulated training data.

F.4.2 Dataset 2 - Test 1

In this run-to-failure test, bearing three (B3) and four (B4) were damaged at the end of the experiment with IR and RE damage respectively. In the run-to-failure experiments, each record is classified separately by the ensemble. The kurtosis trend, shown in Figure F.6, has a slight increase around record 1435, so this is the time when damage is expected to start in bearing 4. The outcome of the plurality voting for each record is shown in Figure F.9. Record 1525 in bearing 4 is the first to be classified as faulty. It is classified with a RE fault, which is the damage that is found at the end of life. However, from record

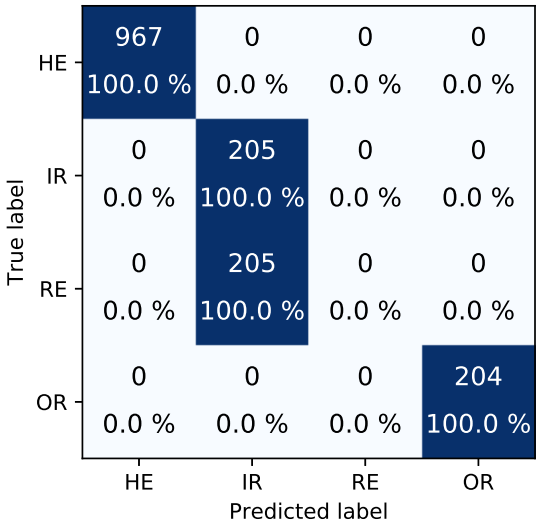


Figure F.8: Confusion matrix for CWRU test data. Predicted label is a result of a plurality voting in the classifier ensemble.

1554, IR damage is indicated, and from that point and onward the classifier outputs both IR and RE damage.

In bearing 3, the classifier also identifies RE and IR damage, with a majority of IR damage from record 2000. Towards the end of life, it is classified as OR fault. The IR fault indication is in accordance with the findings in the reference paper [37]. The authors also note that the IR appeared to be severely spalled, so any impact impulses may not be as prominent as in the training data.

OR damage is indicated in bearing 1 and 2 towards the end of life, but this is assumed to be caused by faults in the other bearings.

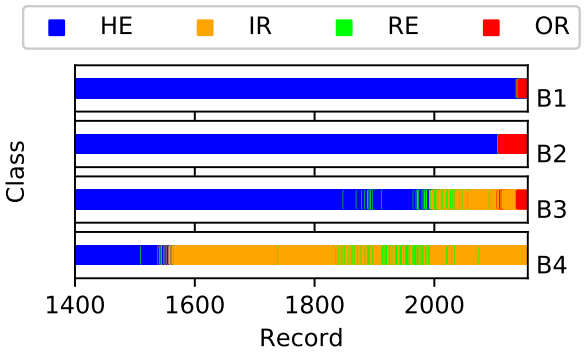


Figure F.9: Dataset 2, test 1, recording 1400 until end of life. Classification by ensemble plurality voting.

F.4.3 Dataset 2 - Test 2

In this test, an OR fault was found in bearing 1 at the end of life, and first signs are expected to appear from approximately record 530. The classifier ensemble correctly outputs OR damage consistently from record 545, as shown in Fig F.10, until the final stages of bearing life, where the classifier changes from OR damage to HE. A possible explanation is that at the end of life, fault size increase, and noise masks any periodic impacts. A healthy bearing is also mainly noisy in some frequency bands. Damage is indicated in the other bearings as well, but as in test 1, this is assumed to be cause by vibration induced by the fault in bearing 1.

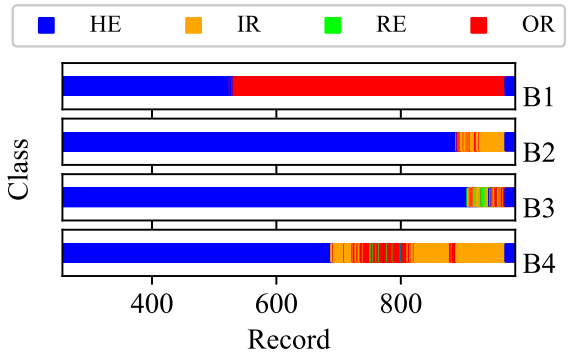


Figure F.10: Dataset 2, test 2. Classification by ensemble plurality voting.

F.4.4 Dataset 3

Dataset 3 is a more challenging case, as the shaft operates at a low speed of 250 rpm. As described in the reference dataset [39], there was extensive bearing damage at the end of the test. The bearing had a small OR pit, two damaged balls, and a larger spalled area in the IR. The authors observed signs of damage in the 20 last measurements, with damage on balls appearing first. Figure F.11 shows RE damage from record 171, then HE from record 179, and RE again from record 188 until record 192.

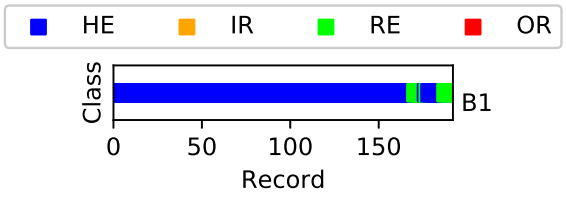


Figure F.11: Dataset 3 test classification by ensemble plurality voting.

These results are partly in accordance with the reference paper. Signs of damage occur at the same time, and the type of damage is coinciding. The classifier does not indicate

other damage than RE, but all fault types were present at the end of testing. This shows that the initial classification was in accordance with the reference paper, but the classifier struggles when multiple faults are present later in the test. This behavior is expected, as training data was only simulated with a single fault at a time.

F.5 Conclusion

In this work, a deep learning classifier for raw vibration signals was trained on simulated data and evaluated on actual fault data. By stacking several layers of 1D convolutions in a hierarchical, dilated structure, the classifier is able to cover a wide receptive field with few layers. This type of architecture has been used successfully in speech and audio modeling, but has yet seen limited application in machine condition monitoring. The presented results show that simulation-driven training of deep neural networks for bearing fault detection and classification has clear potential for industrial applications where automated analysis of time series is beneficial and the access to actual fault data is scarce. Overall, the classifier was able to detect the presence of faults, but tended to misclassify fault types. This may be acceptable in some industrial applications, where detection is more important than diagnosis. In dataset 1, RE damage was classified as IR. In the reference paper, RE damage was also not identifiable, and was said to not exhibit classical fault behavior. This may be the cause of misclassification, as the simulated bearing vibration was simply not representative of this failure mode. Other deep learning classifiers have achieved near perfect accuracy on the same dataset, but not without using real fault data for training. Detection performance is also comparable to results in reference articles in run-to failure dataset 2 and 3, but diagnosis is inconsistent. Training data quality and consequently classifier performance will improve with more information about the system. For example, an estimate of the resonance frequency can be obtained through a bump test. More advanced and accurate models for bearing vibration could also improve performance, and should be investigated further.

Acknowledgment

The research presented in this paper has received funding from the Norwegian Research Council, SFI Offshore Mechatronics, project number 237896.

References

[1] P. McFadden and J. Smith. Model for the Vibration Produced by a Single Point Defect in a Rolling Element Bearing. *Journal of Sound and Vibration*, 96(1):69–82, 1984. doi:10.1016/0022-460X(84)90595-9.

[2] L. Marple. Computing the Discrete-time "Analytic" Signal via FFT. *IEEE Transactions on Signal Processing*, 47(9):2600–2603, 1999. doi:10.1109/78.782222.

[3] J. Antoni. Fast Computation of the Kurtogram for the Detection of Transient Faults. *Mechanical Systems and Signal Processing*, 21(1):108–124, 2007. doi:10.1016/j.ymssp.2005.12.002.

[4] J. Antoni. The Spectral Kurtosis: A Useful Tool for Characterising Non-stationary Signals. *Mechanical Systems and Signal Processing*, 20(2):282–307, 2006. doi:10.1016/j.ymssp.2004.09.001.

[5] K. Fyfe and E. Munck. Analysis of Computed Order Tracking. *Mechanical Systems and Signal Processing*, 11(2):187–205, 1997. doi:10.1006/MSSP.1996.0056.

[6] J. Antoni and R. B. Randall. Differential Diagnosis of Gear and Bearing Faults. *Journal of Vibration and Acoustics*, 124(2):165–171, 2002. doi:10.1115/1.1456906.

[7] E. Bechhoefer and M. Kingsley. A Review of Time Synchronous Average Algorithms. In *Annual Conference of the Prognostics and Health Management Society*, pages 24–33, 2009.

[8] R. B. Randall and N. Sawalhi. Use of the Cepstrum to Remove Selected Discrete Frequency Components from a Time Signal. In *Proc. Int. Conference on Noise and Vibration Engineering (ISMA)*, pages 451–461. Springer, New York, NY, 2011. doi:10.1007/978-1-4419-9428-8_38.

[9] R. B. Randall, N. Sawalhi, and M. Coats. A Comparison of Methods for Separation of Deterministic and Random Signals. *The International Journal of Condition Monitoring*, 1(1):11–19, 2011. doi:https://doi.org/10.1784/204764211798089048.

[10] E. Bechhoefer, R. Schlanbusch, and T. I. Waag. Fault Detection on Large Slow Bearings. In *European Conference of the Prognostics and Health Management Society 2016*, volume 7, pages 1–8, 2016.

[11] A. Diez-Olivan, J. Del Ser, D. Galar, and B. Sierra. Data Fusion and Machine Learning for Industrial Prognosis: Trends and Perspectives Towards Industry 4.0. *Information Fusion*, 50:92–111, 2019. doi:10.1016/J.INFFUS.2018.10.005.

[12] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao. Deep Learning and Its Applications to Machine Health Monitoring. *Mechanical Systems and Signal Processing*, 115:213–237, 2019. doi:10.1016/J.YMSSP.2018.05.050.

[13] X. Li, Z. Liu, Y. Qu, and D. He. Unsupervised Gear Fault Diagnosis Using Raw Vibration Signal Based on Deep Learning. In *2018 Prognostics and System Health Management Conference*, pages 1025–1030. IEEE, 2018. doi:10.1109/PHM-Chongqing.2018.00182.

[14] F. Jia, Y. Lei, L. Guo, J. Lin, and S. Xing. A Neural Network Constructed by Deep Learning Technique and Its Application to Intelligent Fault Diagnosis of Machines. *Neurocomputing*, 272:619–628, 2018. doi:10.1016/J.NEUCOM.2017.07.032.

[15] X. Li, Y. Yang, H. Pan, J. Cheng, and J. Cheng. A Novel Deep Stacking Least Squares Support Vector Machine for Rolling Bearing Fault Diagnosis. *Computers in Industry*, 110:36–47, 2019. doi:10.1016/J.COMPIND.2019.05.005.

[16] M. Gan, C. Wang, and an Zhu. Construction of Hierarchical Diagnosis Network Based on Deep Learning and Its Application in the Fault Pattern Recognition of Rolling Element Bearings. *Mechanical Systems and Signal Processing*, 72-73:92–104, 2015. doi:10.1016/j.ymssp.2015.11.014.

[17] M. He and D. He. Deep Learning Based Approach for Bearing Fault Diagnosis. *IEEE Transactions on Industry Applications*, 53(3):3057–3065, 2017. doi:10.1109/TIA.2017.2661250.

[18] X. Guo, L. Chen, and C. Shen. Hierarchical Adaptive Deep Convolution Neural Network and Its Application to Bearing Fault Diagnosis. *Measurement*, 93:490–502, 2016. doi:10.1016/J.MEASUREMENT.2016.07.054.

[19] G. Hinton, L. Deng, D. Yu, G. Dahl, A. Mohamed, N. Jaitly, and B. Kingsbury. Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal Processing Magazine*, 2012.

[20] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, Ł. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean. Google's Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation. Technical report, 2016. arXiv:1609.08144v2.

[21] A. Van Den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu. Wavenet: A Generative Model for Raw Audio. *arXiv preprint*, 2016. arXiv:1609.03499v2.

[22] H. Sak, A. Senior, and F. Beaufays. Long Short-term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling. In *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pages 338–342, 2014.

[23] D. Ho and R. B. Randall. Optimisation of Bearing Diagnostic Techniques Using Simulated and Actual Bearing Fault Signals. *Mechanical Systems and Signal Processing*, 14(5):763–788, 2000. doi:10.1006/mssp.2000.1304.

[24] N. Sawalhi and R. B. Randall. Simulating Gear and Bearing Interactions in the Presence of Faults Part I. the Combined Gear Bearing Dynamic Model and the Simulation of Localised Bearing Faults. *Mechanical Systems and Signal Processing*, 22(8):1924–1951, 2008. doi:10.1016/j.ymssp.2007.12.001.

[25] C. Sobie, C. Freitas, and M. Nicolai. Simulation-driven Machine Learning: Bearing Fault Classification. *Mechanical Systems and Signal Processing*, 2018. doi:10.1016/j.ymssp.2017.06.025.

[26] J. Antoni. Cyclostationarity by Examples. *Mechanical Systems and Signal Processing*, 23(4):987–1036, 2009. doi:10.1016/j.ymssp.2008.10.010.

[27] M. A. Khan, Y.-H. Kim, and J. Choo. Intelligent Fault Detection via Dilated Convolutional Neural Networks. In *2018 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 729–731. IEEE, 2018. doi:10.1109/BigComp.2018.00137.

[28] R. B. Randall and J. Antoni. Rolling Element Bearing Diagnostics - A Tutorial. *Mechanical Systems and Signal Processing*, 25(2):485–520, 2011. doi:10.1016/j.ymssp.2010.07.017.

[29] J. Antoni. Cyclic Spectral Analysis of Rolling-element Bearing Signals: Facts and Fictions. *Journal of Sound and Vibration*, 304(3-5):497–529, 2007. doi:10.1016/j.jsv.2007.02.029.

[30] F. Chollet and Others. Keras, 2015. URL: https://keras.io.

[31] V. Nair and G. Hinton. Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on Machine Learning*, 2010.

[32] S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. Technical report, 2015. arXiv:1502.03167.

[33] M. Lin, Q. Chen, and S. Yan. Network In Network. Technical report, 2014. arXiv:1312.4400v3.

[34] Case Western Reserve University Bearing Data Center Website. Date obtained: 19.12.2019. URL: http://csegroups.case.edu/bearingdatacenter/home.

[35] W. A. Smith and R. B. Randall. Rolling Element Bearing Diagnostics using the Case Western Reserve University Data: A Benchmark Study. *Mechanical Systems and Signal Processing*, 64-65:100–131, 2015. doi:10.1016/j.ymssp.2015.04.021.

[36] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler. Machine Learning and Deep Learning Algorithms for Bearing Fault Diagnostics-a Comprehensive Review. Technical report. arXiv:1901.08247v2.

[37] H. Qiu, J. Lee, J. Lin, and G. Yu. Wavelet Filter-based Weak Signature Detection Method and Its Application on Rolling Element Bearing Prognostics. *JOURNAL OF SOUND AND VIBRATION*, 289:1066–1090, 2006. doi:10.1016/j.jsv.2005.03.007.

[38] A. Klausen, R. W. Folgerø, K. G. Robbersmyr, and H. R. Karimi. Accelerated Bearing Life-time Test Rig Development for Low Speed Data Acquisition. *Identification and Control*, 38(3):143–156, 2017. doi:10.4173/mic.2017.3.4.

[39] A. Klausen, K. G. Robbersmyr, and H. R. Karimi. Autonomous Bearing Fault Diagnosis Method Based on Envelope Spectrum. *IFAC-PapersOnLine*, 50(1):13378–13383, 2017. doi:10.1016/J.IFACOL.2017.08.2262.