From Concepts to Conditions: Bridging the Gap in AI-Based Maintenance Systems

Adam Lycksam

Mittuniversitetet
MID SWEDEN UNIVERSITY

Department of Computer and Electrical Engineering
Mid Sweden University

Doctoral Thesis No. 431
Sundsvall, Sweden
2025

Mittuniversitetet
Institutionen för data- och elektroteknik
SE-851 70 Sundsvall
SWEDEN

ISBN 978-91-90017-26-5
ISNN 1652-893X

Akademisk avhandling som med tillstånd av Mittuniversitetet framlägges till offentlig granskning för avläggande av teknologie doktorsexamen måndagen den 16 juni 2025 i C312, Mittuniversitetet, Holmgatan 10, Sundsvall.

Tryck: Tryckeriet Mittuniversitetet

Abstract

The importance of preventing machine failures and reducing costly unplanned downtime has led to extensive research aiming to develop methods that predict maintenance needs. In this context, data-driven and particularly Deep Learning (DL) based methods for anomaly detection, fault diagnosis, and health prognosis have been studied extensively because of their ability to handle the complexity of the sensor data describing the health state of machines. However, many challenging factors exist before it is possible to utilize these methods in practice. These primarily include the lack of labelled failure events, the heterogeneous nature of the data, and the occurrence of multi-component fault scenarios. Currently, most studies ignore these aspects and focus on scenarios limited to a laboratory environment, which means there is a need to develop methods that can be deployed in practice. Therefore, this thesis suggests methods for these challenges and gives insight, aiming to reduce the gap between research-defined scenarios and scenarios found in industrial environments. To achieve this, different areas in the context of DL for Predictive Mainenence (PdM) are examined, including multivariate anomaly detection, fault diagnosis, and Remaining Useful Life (RUL) prediction methods.

One of the contributions is a threshold-setting procedure that optimizes anomaly detection models with the user's support and a novel separate scoring method, and outperforms state-of-the-art alternatives for deployments in industrial applications. A published dataset of bearing faults from an industrial environment is also described, which is beneficial when developing and evaluating methods. In addition, a novel DL method for fault diagnosis of bearings using vibration data constructed with knowledge enrichment, time-based contextual enrichment, and a transfer learning technique is suggested. This method can be deployed on any machine without historical faults and outperforms state-of-the-art methods. Lastly, the most significant contribution is a prognostic hybrid framework for multi-component fault scenarios in rotating machines using vibration data that utilizes advancements in methods for anomaly detection, fault diagnosis, and RUL prediction of machines.

In summary, this thesis suggests novel methods for PdM adapted for industrial applications that can be used on a general basis and provides insights that lower the gap between research-defined scenarios and scenarios found in industrial environments.

Sammanfattning

Vikten av att förhindra haverier i maskiner och reducera kostsamma oplanerade stopp har lett till omfattande forskning med syfte att utveckla metoder för att prediktera underhållsbehov. Inom detta område har datadrivna metoder, och framför allt djupinlärning blivit framträdande på grund av dess förmåga att hantera komplexiteten i sensordata som beskriver hälsotillståndet hos maskinerna. För att använde dessa metoder krävs dock att ett antal olika utmaningar hanteras. Dessa inkluderar primärt bristen på historiska uppmärkta händelser av haverier, att data är heterogen och förekomsten av fel från flera olika komponenter. Nuvarande forskning ignorerar de flesta av dessa utmaningar och fokuserar på scenarier som bara återfinns i labbmiljö. På grund av detta finns det ett stort behov av att utveckla metoder som är ämnande åt den industriella miljön. Därför föreslår den här avhandlingen metoder för dessa utmaningar och ger insikter som kan minska gapet mellan scenarier framtagna i forskningen och de som förväntas inträffa i praktiken. För att uppnå detta utforskas olika områden med hjälp av djupinlärning inom kontexten prediktivt underhåll innefattande avvikelsedetektering, feldiagnostisering och metoder för att prediktera återstående livslängd.

Ett av bidragen är en metod för att sätta tröskelvärden för avvikelsedetekteringsmetoder med hjälp av återkoppling från användare och en ny metod för av att avgöra graden av avvikelse som presterar bättre än befintliga metoder. Vidare beskrivs egenskaperna hos ett dataset med vibrationsdata besående av lagerfel från den industriella miljön som också publicerats, vilket är av stort värde vid utveckling och utvärdering av metoder. Dessutom presenteras en nyskapande metod för feldiagnostisering av lager baserat på djupinlärning som berikas med kunskap och historisk kontext. Metoden går att applicera på alla maskiner utan att behöva ha tillgång till historiska fel och presterar bättre än de bästa befintliga metoderna. Slutligen är det mest värdefulla bidraget ett ramverk för att hantera fel från flera olika komponenter i roterande maskiner med hjälp av vibrationsdata som utnyttjar framsteg hos metoder inom avvikelsedetektering, feldiagnostisering och prediktering av återstående livslängd.

Sammanfattningsvis presenterar den här avhandlingen innovativa metoder för prediktivt underhåll anpassat för industriella applikationer som har en hög generaliserbarhet och presenterar insikter som reducerar gapet mellan scenarier definierade inom forskningen och de som återfinns i praktiken.

Contents

List of Figures

List of Tables

Abbreviations

AE Autoencoder

AI Artifical Intelligence

CBM Condition-Based Maintenance

CNN Convolutional Neural Network

CWT Continuous Wavelet Transform

DANN Domain-Adversarial Neural Network

DL Deep Learning

FFT Fast Fourier Transform

GAN Generative Adversarial Network

GNN Graph Neural Network

LSTM Long-Short-Term Memory

ML Machine Learning

PdM Predictive Mainenence

RMS Root Mean Squared

RMSE Root Mean Squared Error

RNN Recurrent Neural Network

RUL Remaining Useful Life

XAI Explainable AI

List of Papers

Acknowledgments

Pursuing a PhD has been an interesting, challenging, and humbling experience, and I'm grateful for the knowledge and insights I have gained in the process. I want to acknowledge that I owe much of my progress to the support, inspiration, and guidance of others.

First, I'm grateful to my main supervisor, Professor Mattias O'Nils, for your guidance, encouragement, support, and insights. I could not have asked for a better supervisor. I also like to thank my co-supervisor, Professor Faisal Z. Qureshi, for continuously helping me grow as a researcher and for your contributions. Additionally, thanks to my colleagues at Mid Sweden University, who have supported and inspired me.

I express my gratitude to SCA for investing in me and my research. I also want to thank my colleagues at SCA, especially my industrial supervisor, Jenny Klöver, and colleague Lars Ericsson, for your support, encouragement, and invaluable insights, which enabled and inspired me to improve my work.

I'm grateful for the financial support from the Knowledge Foundation within the industrial graduate school Smart Industry Sweden. I also want to express my appreciation for the research school's support and all the insightful and inspiring interactions with other industries and fellow PhD students I've been fortunate to have, with their help.

I am deeply grateful to my parents, Sara and Stefan, and my two sisters, Lisa and Stina, for always believing in me and supporting my aspirations. Lastly, my most heartfelt thanks to my wife, Frida, for your unwavering support and the sacrifices you have made. You and our son mean the world to me.

Adam Lycksam
Sundsvall, May 2025

Introduction

Machine failures are a significant issue for industries because they induce the risk of unplanned downtime, which can come at a substantial cost due to the reduced possibility of production and repairs, but also pose a safety hazard [1]–[3]. This has been the motivating factor for Condition-Based Maintenance (CBM) and particularly its extension, Predictive Maintenance (PdM), which aims at predicting maintenance needs and thereby reducing the risk of downtime [1], [2], [4]. With technological advancements, increased availability of data, and the improvements of Artifical Intelligence (AI) algorithms, data-driven methods have gained popularity, which typically means that data describing the health states of machines are collected from sensors and used to detect faults and predict the Remaining Useful Life (RUL) [2], [3], [5]. Several different methods for PdM using a data-driven strategy have been suggested and range from statistical to Machine Learning (ML) [1]–[3]. Recently, Deep Learning (DL) has become one of the most prominent approach for developing anomaly detection, fault diagnosis, and RUL prediction methods, where multiple studies have successfully shown the possibility for intelligent PdM solutions by capturing the complex patterns found in the data [2], [5]. Despite this, many challenges exist to deploy these methods in practice, including data heterogeneity, the lack of labeled historical failures, and multi-component fault scenarios [2], [5]. In addition, many studies make unrealistic assumptions when developing and evaluating methods, considering the scenarios expected in industrial applications. Therefore, this thesis proposes methods based on DL for PdM that specifically target these challenges and provide insight that lowers the gap between research-defined scenarios and those found in industrial environments. This is achieved by exploring different approaches towards PdM on a system level, including anomaly detection, fault diagnosis, and RUL prediction, depicted in Figure 1.1, that enables users to get support in maintenance actions.

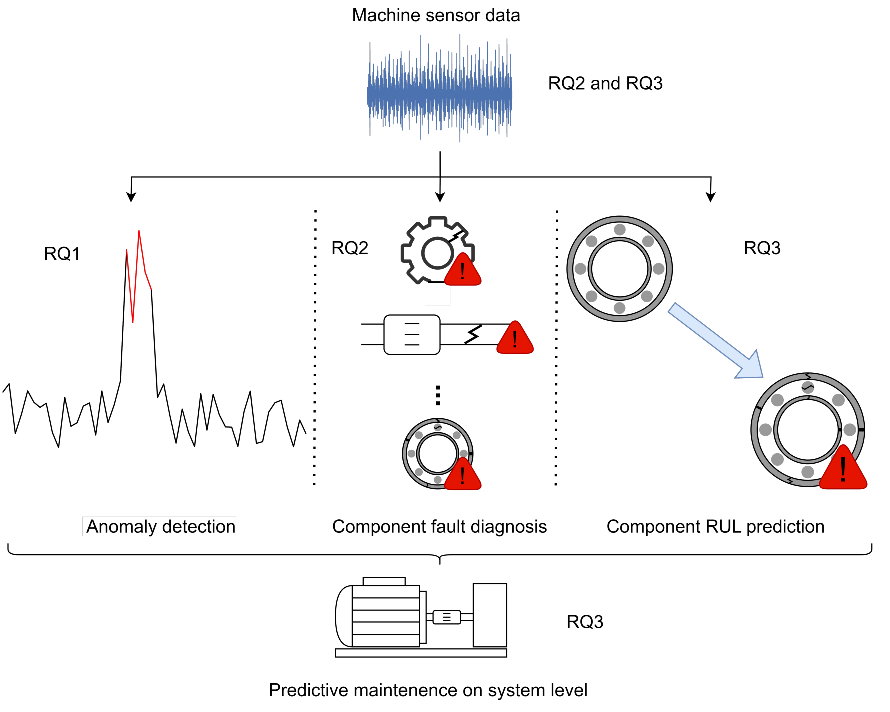


Figure 1.1: Summary of the different research directions in PdM based on data-driven approaches.

1.1 Background and Problem Motivation

Considering the possibility of capturing complex patterns in data, DL has the potential to generate accurate decision support using sensor data in PdM solutions. Despite this, many challenges face the development of DL methods. In practice, labelled data is challenging to collect in an industrial environment due to, for example, a lack of historical failures [2], [3]. This means that methods must be able to be deployed with a limited amount or without any historical examples of faults from the target machine. In addition, a high level of discrepancy exists in the sensor data from industrial environments due to differences in, for example, measurement settings, operational conditions, noise, and component or machine type, which further complicates the usage of DL methods [1], [2], [6]. This can, for example, mean that a solution that works well for an application to monitor a component or machine in one part of the factory might not work well in another. Lastly, considering that machines consist of multiple components, each of which may fail, it is of great importance to develop methods for a system-level approach that manages different fault scenarios [1], [5], [7]. Based on this, it is possible to summarize three of the most significant challenges towards the practical application of DL methods for PdM to:

C1 The management of limited historical faults from the target machine.

C2 The discrepancies between the training data that describe the health state of a machine or component and the test data from a new machine.

C3 Developing methods that can handle multi-component fault scenarios.

Much research has been conducted using DL-based methods for anomaly detection, fault diagnosis, and RUL prediction, including the usage of unsupervised, semi-supervised, and supervised learning [8], [9]. However, significant gaps exist in the current body of research when considering industrial applications. Firstly, most of the current studies evaluate their method on data from the laboratory, which does not necessarily capture the varying conditions found in industrial environments [1], [2], [10]. Secondly, many studies make assumptions, such as the availability of historical faults from the target machine [11], [12], and that the deployment of models is limited to the same machine as considered during training [13]. This means they ignore factors that can significantly affect the method's performance and are expected in practice. Lastly, most studies for fault diagnosis and RUL prediction only consider a single component in developing and evaluating methods [3], [5], [7]. This is problematic because sensor readings from a machine are difficult to limit to a single component. In addition, for a system-level support system based on PdM, there is a need to distinguish between different fault scenarios [5]. These aspects suggest a wide gap between research-defined and industrial scenarios that needs to be bridged, which has been the motivating factor of this research.

1.2 Purpose and Research Questions

This thesis aims to propose methods that can be deployed on a general basis and provide insights that lower the gap between research-defined scenarios and those found in industrial environments. In particular, its purpose is to address challenges C1-C3. To achieve this, the thesis explores different approaches, including improvements to multivariate anomaly detection, domain generalization for fault diagnosis, and developing a PdM framework considering a multi-component system. To support this, it also examines the limitations of state-of-the-art methods and the characteristics of industrial data. Specifically, the following research questions are targeted:

RQ1 How can multivariate anomaly detection methods based on DL be improved to increase the performance compared to state-of-the-art in industrial applications?

RQ2 How can a DL method for domain generalization for bearing fault diagnosis be developed to outperform state-of-the-art methods and demonstrate high accuracy on data from industrial environments?

RQ3 How can state-of-art multi-component PdM system for rotating machines be achieved?

1.3 Delimitations

This thesis presents methods that can be used in many applications, especially considering the methods related to anomaly detection, but the focus is on rotating machines and their components. To achieve this, most of the work is based on vibration measurements due to being effective in displaying faults and degradation of rotating components [7]. Furthermore, this research has considered multiple fault scenarios but has had a particular focus on bearing fault scenarios, considering that it is one of the most common components in machines to fail [14], and the high availability of public data [5].

1.4 Contribution

The authors' contribution to the papers included in the thesis is summarized in Table 1.1. It should be noted that Adam Lycksam's previous name was Adam Lundström, which is referred to in all papers except paper VI.

Table 1.1: The papers and the contribution from the authors.

Paper	Author	Role	Contributions
I	Adam Lycksam	Main author	Conceptualization, methodology, data curraction, software & writing
	Mattias O'Nils	Co-author	Supervision, contribution to conceptualization, methodology & text
	Faisal Z. Qureshi	Co-author	Contribution to conceptualization, methodology & text
	Axel Jantsch	Co-author	Contribution to conceptualization, methodology & text
II	Adam Lycksam	Main author	Conceptualization, methodology, data curraction, software & writing
	Mattias O'Nils	Co-author	Supervision, contribution to conceptualization, methodology & text
	Faisal Z. Qureshi	Co-author	Contribution to conceptualization, methodology & text
III	Adam Lycksam	Main author	Conceptualization, methodology, data curraction, software & writing
	Mattias O'Nils	Co-author	Supervision, contribution to conceptualization, methodology & text
IV	Adam Lycksam	Main author	Conceptualization, methodology, data curraction, software & writing
	Mattias O'Nils	Co-author	Supervision, contribution to conceptualization, methodology & text
	Faisal Z. Qureshi	Co-author	Contribution to conceptualization, methodology & text
V	Adam Lycksam	Main author	Conceptualization, methodology, data curraction, software & writing
	Mattias O'Nils	Co-author	Supervision, contribution to conceptualization, methodology & text
	Faisal Z. Qureshi	Co-author	Contribution to conceptualization, methodology & text
VI	Adam Lycksam	Main author	Conceptualization, methodology, data curraction, software & writing
	Mattias O'Nils	Co-author	Supervision, contribution to conceptualization, methodology & text
	Faisal Z. Qureshi	Co-author	Contribution to conceptualization, methodology & text

1.5 Outline

The outline of the thesis is as follows:

- Chapter 2 describes related work and theory. It summarizes methods found in state-of-the-art research and underlines some of the existing weaknesses and gaps. This chapter is used as background for Chapter 3.
- Chapter 3 summarizes the methodology, focusing on the approach to answer the research questions. It describes the key aspects of the methods used in the different papers and explains how they are evaluated.
- Chapter 4 summarizes the results and addresses the research questions. To achieve this, it accounts for some key results and conclusions of the different papers.
- Chapter 5 discusses the impact of the thesis, describes some of the limiting factors, and suggests future work.
- Chapter 6 concludes the thesis by summarizing the key findings and provides some concluding remarks.

Related Work

This section introduces related work and theory associated with the areas in the context of PdM targeted by the research questions, which include anomaly detection, fault diagnosis, and RUL prediction. It also highlights some of the main limitations of the current research based on scenarios excepted in industrial applications, which acts as a motivating foundation for this thesis.

2.1 Anomaly Detection

Anomaly detection, aiming to separate abnormal events from normal, has demonstrated significant value in several areas such as manufacturing and maintenance [15]. Considering the interdependencies of signals from a single machine, the importance of multivariate anomaly detection has become apparent [15]. Because of the complexity, DL has become prominent because it can capture complex patterns in the data related to the temporal aspects and intercorrelation between signals, which can be difficult to achieve using methods based on statistics or shallow ML [16]. One of the most common methods for these solutions is to utilize the healthy data, which is relatively easy to collect and describes the normal state of the different signals. Commonly, the algorithm is trained to recreate, generate, or predict future values based on the normal data, which means that it will be good at understanding the normal behaviour of machines but bad at predicting abnormal data [17]. This can then be used to separate anomalous from normal data. Several different types of approaches have been suggested, including the usage of Convolutional Neural Network (CNN), Autoencoder (AE), Long-Short-Term Memory (LSTM), transformers, Generative Adversarial Network (GAN), and Graph Neural Network (GNN) [15], [17]. Some methods are USAD [18], which employs an adversarial learning procedure with multiple fully connected layers, MAD-GAN [19], which uses a GAN structure, OmniAnomaly [20] with a variational AE, GDN [21], which is based on a GNN, and VTT based on a transformer [22]. These methods use normalized or standardized data, usually structured in windows so that the network can capture the temporal interdependencies. Based on the model's prediction, an anomaly score is applied that describes how anomalous the current observation is. Several different scoring methods exist, but a common factor is that they employ an aggregated score [23] as illustrated in Figure 2.1. The potential issue with using an aggregated scoring method is that the aggregation procedure might lose valuable information from the output for different channels, which correspond with each sensor. It is therefore valuable to explore methods that capture all information from the different channels in the scoring, which was the motivation for paper I.

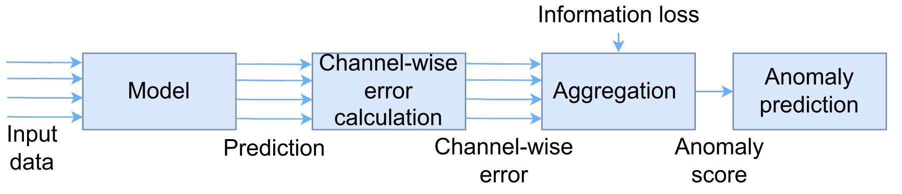


Figure 2.1: Overview of the different parts of current multivariate anomaly detection methods.

A threshold must be used based on the output of the scoring method based on the anomaly detection model to define what is anomalous and what is not [17]. In most research, such as [15], [18], [19], [22], [24], when evaluating a new anomaly detection method, the thresholds are optimized based on the ground truth labels of the test data. Since the thresholds only affect the sensitivity of the anomaly prediction and not the model's ability to separate abnormal from normal data, it can effectively show the model and scoring methods' capability. In addition, because this can be achieved for all methods and is a fair comparison approach, it can be practical when benchmarking a method against the state-of-the-art. However, since these labels do not exist in practical applications, it is necessary to have a method to create these thresholds [17]. Several attempts at employing automated threshold-setting procedures have been suggested, such as Nonparametric Dynamic Thresholding (NDT) [25] and the Peaks-Over-Threshold (POT) method proposed by [20], which have shown promising results. However, to what extent a user-supported threshold-setting method can outperform these automated methods has not been evaluated, which was the aim in paper II.

2.2 Bearing Fault Diagnosis

The inherent limitation of anomaly detection methods is their inability to distinguish between unusual events that are normal and faults. To manage this, fault diagnosis methods can be used, and one of the most common applications is bearing fault diagnosis because of its commonness in industrial machines and leading cause of failure [8], [14]. Several methods have been suggested, including physics-based [26] and ML-based methods developed for vibration data [5]. Recently, solutions based on DL using for example CNN, Recurrent Neural Network (RNN) such as LSTM and attention mechanisms, have become popular because of their ability to learn from the complexity of the vibration signal, which can be challenging to achieve using shallow ML [27]–[29]. In addition, many publicly available datasets exist that are used for benchmark evaluations. However, as seen in Table 2.1, most of these datasets are retrieved from a laboratory environment. This means it is still unclear to what extent the methods currently developed based on these will work well in practice [28], [29]. Given the heterogeneous nature of the data from different machines, it is also important to publish and describe features of the data from the industrial environment that might be important for future work, which was the purpose of paper III.

Table 2.1: Overview of public datasets of bearing faults based on vibration data.

Dataset	**Environment**	**Fault type**
IMS [30]	Laboratory	Natural
CWRU [31]	Laboratory	Artificial
Pronostia [32]	Laboratory	Natural
PU [33],[34]	Laboratory	Artifical and natural
HUST [35],[36]	Laboratory	Artifical
MPFT [37]	Laboratory and three cases with industrial data	Artificial and natural
SEU [38],[39]	Laboratory	-
Ottawa [40], [41]	Laboratory	Artificial

Because of the lack of historical faults from components in industrial environments and data discrepancies across domains, meaning a unique operational setting or machine, transfer learning methods have been suggested to solve these challenges [11], [13]. There is an inconsistency in the literature regarding the definition of transfer learning and related terms such as domain adaptation and domain generalization, for example, between [11] and [42]. Still, this work views it as a training technique where one or several tasks are used to develop a model that can be deployed on a related task, similar to the definition in [11]. This typically means that knowledge is learned by training on data from labelled source domains and then transferred to a target domain where a domain shift exists and historically available examples of faults are lacking. This is exemplified in Figure 2.2, which shows a scenario where knowledge is extracted from several source domains (could also be labeled or unlabeled samples from the target domain if available) using a transfer learning method and used to predict the state in a target domain where the ground truth is unknown.

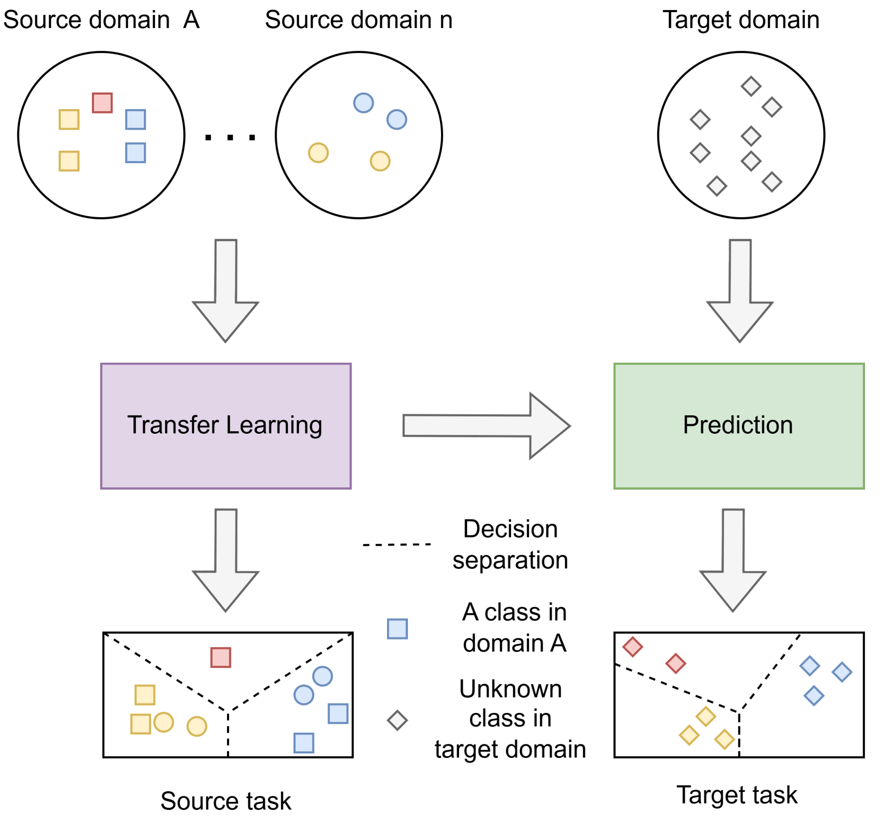


Figure 2.2: An illustration of transfer learning.

In the context of bearing fault diagnosis, several approaches for transfer learning have been suggested, including fine-tuning-based, which means that part of the network is frozen after pre-training and adjusted based on data from the target domain [11], [13]. In addition, adversarial-based and statistic-based methods are commonly used, where features extracted from the feature extractor are used to enforce the network to reduce discrepancies between domains during training [11], [13]. Adversarial-based methods use labels for each source domain to train the feature extractor to deceive a domain classifier, making it capture domain-invariant features [11], [13]. In contrast, statistic-based methods use a distribution metric loss function based on features extracted from the feature extractor to reduce the discrepancies between domains [11], [13]. Additionally, combining these approaches as was done in, for example, [43] and [44], is a common practice.

Furthermore, it is possible to categorize transfer learning methods based on the problem they intend to solve. One of the popular approaches is domain adaptation using faults and normal data, meaning that the features extracted from the feature extractor of an unlabeled or labeled target domain containing one or several faults are adapted to one or several labeled source domains during training [11], [13]. The problem with this approach is that it assumes that historical faults are available in the target domain, which is not necessarily the case.

To solve this issue, another approach called domain generalization is frequently targeted, which, in contrast to domain adaptation, does not need historical faults from the target domain [11], [42]. Generally, several different sources with different working conditions are used in training, but no fault from the target domain is included. Then the model is used to predict the state of the target domain. The issue with most of the current work, such as [45]–[51], is that they do not consider the deployment on a new machine and only between operational settings on the same machine. The primary issue with this is that knowledge transfer between machines is expected in industrial scenarios, which means that the performance of these suggested methods has not been adequately evaluated.

Some studies suggest methods that achieve domain generalization and transfer knowledge between machines. Three of them are [12], [52] and [53]. However, there is a need to improve the performance of methods for these scenarios because they replicate the actual conditions found in industrial scenarios. Therefore, the purpose of paper IV was to develop a new method and compare it against the state-of-the-art.

2.3 Component RUL Prediction

Understanding the life length of rotating machines is important in industries because it allows planning maintenance at the right time, which means the factory can run with reduced unplanned downtime. In this context, RUL is an important concept that at a certain point in time describes the time until the machine or component loses its function [54]. Several approaches exist for RUL prediction of bearings and other components, commonly based on vibration data. When considering these, a distinction can be made between strategies for when the RUL prediction method makes its first estimation. Firstly, one approach is to develop a model for the complete life length, meaning that the model predicts the RUL during both the normal and degradation phases, for example, used in [55] and [56].

Another approach is based on health indicators that are used to define the start of the degradation by applying anomaly detection, and are often used as input for the RUL prediction model. This means that the RUL prediction model only considers the degradation phase of the component. Several approaches exist to construct the health indicator ranging from statistics-based methods [57]–[59], AE [60], methods based on Gaussian Mixture Model (GMM) [61], [62] and to those that targets the fault characteristics of the component using the fault frequencies [63]–[65].

Regarding the development of the RUL prediction model, different approaches exist and include physics-based, statistical, ML-based and, in recent times, DL-based, which, similarly as in the area of anomaly detection and fault diagnosis, have gained popularity due to being able to find complex patterns in data [1], [10]. Much research has been conducted to find methods based on DL for predicting degradation of components using vibration data [66], [67], where the learning objective is usually to map the data to a degradation point from 1, healthy, to 0, failure, or to make a forecase of a health indicator [68]. Different types of preprocessing exist and range from using the raw data [69], frequency transformation using Fast Fourier Transform (FFT) [70], time-frequency transformation using, for example, Continuous Wavelet Transform (CWT) [57], to aggregated features [71]. Similar to the development of methods for fault diagnosis, data discrepancy across domains is a significant challenge. For example, it is not certain that the degradation pattern found in one machine can be seen in another [67]. In addition, the lack of historical examples of failures from the target machine makes it difficult to apply the DL method in practice [2]. To manage these challenges, transfer learning methods have become prominent using techniques similar to those for fault diagnosis. Firstly, adversarial learning is commonly applied, including in [72], [73], where architectures based on Domain-Adversarial Neural Network (DANN) [74] are adopted, illustrated in Figure 2.3. The main idea is to use a gradient reversal layer to train the feature extractor to make features from different domains indistinguishable for the domain classifier. In contrast, the domain classifier is optimized to make this distinction. Combining an objective for the feature extractor and regressor to accurately predict the RUL of the bearing creates a model that can ignore domain discrepancies and simultaneously accurately predict the RUL.

In addition to adversarial-based methods, statistics-based methods (sometimes referred to as metric-based) that share the same idea, such as Coral and Maximum Mean Discrepancy (MMD) used in [55], but instead uses statistical metrics to reduce the domain discrepancies between features across domains based on the output of the feature extractor are common. Despite these efforts, the vast majority of the current studies either only consider knowledge transfer between operational settings on the same machine, such as [72], [75], [76], or between machines but use run-to-failure data from the target domain during training [69], [77], [78]. Because of this, it is unclear how these methods will handle scenarios where knowledge must be transferred between machines and labelled data is unavailable.

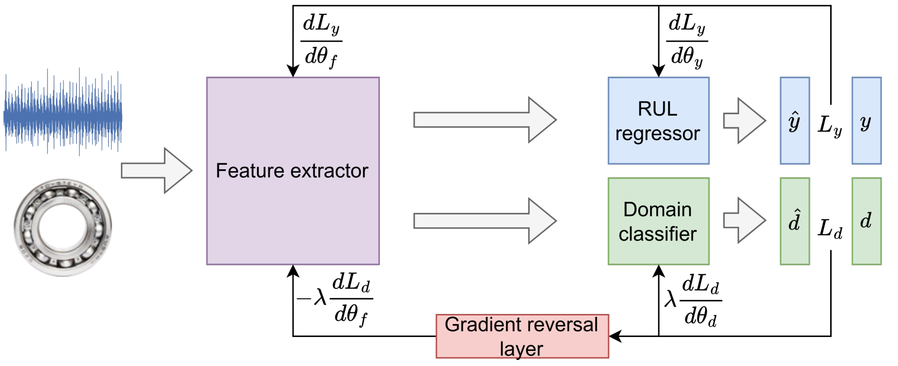


Figure 2.3: The architecture of DANN [74] for bearing RUL prediction.

Therefore, the motivation behind paper IV was to evaluate the methodology used in these methods in scenarios found in industrial applications.

2.4 Multi-Component Systems

Apart from the domain discrepancies and a lack of labelled data from the target machine, a significant challenge in data-driven PdM using ML is multi-component fault scenarios. Limiting the vibration measurements to a single component in the rotating machine to capture faults is difficult, if not impossible. For example, vibration measurements on a motor's bearing can show symptoms of bearing faults and other faults, such as gear faults, shaft imbalances, or misalignments [7]. In this context, a major issue also acknowledged in, for example [1], [2], [5], [7], is that there is a lack of studies that consider a multi-component system where several faults can occur on the same machine. For example, in studies where methods for RUL prediction and fault diagnois of bearings are considered, the datasets being used are from an isolated laboratory environment where only the bearing is failing, including the IMS dataset [30], the XJTU-SY dataset [79], and the PRONOSTIA dataset [80] meaning that the model only need to to be able to separate between healthy and a single unhealthy state. This strictly means that the application of these methods is limited to an environment where no faults apart from those of the component considered during training exist, which can't be assumed in industrial environments. In addition, it is of great importance to be able to separate fault scenarios because of the possible differences in degradation and severity, and thereby the action needed by the user. It is therefore highly problematic that an understanding of how a method for multi-component systems should be developed for the industrial scenario is lacking. Because of this, the motivation for paper VI was to create a framework for this scenario.

2.5 Positioning

In summary, a substantial amount of previous work has developed methods based on DL using novel preprocessing techniques, model architectures, or training procedures for PdM. The central issue is that the scenarios used for developing and evaluating these methods often do not replicate conditions found in practice. In contrast, the primary focus of this thesis is adjusting state-of-the-art DL methods and related techniques to adapt to the scenarios expected in industrial applications, considering challenges C1-C3. To achieve this, the main strategy is utilizing healthy data, which is easy to collect in contrast to historical faults, and incorporating domain knowledge, such as known fault characteristics in training and user expertise.

Methodology

The methodology, displayed in Figure 3.1, based on quantitative research and knowledge development, was used to answer the research questions. Considering the complexity of challenges C1-C3, some papers, for example I, III, and VI, only served to develop methods and gather knowledge used to target research questions in a succeeding paper. Consistent with the research questions, three areas were considered in this work to reach the goal of industry-adapted methods for PdM. These were multivariate anomaly detection related to RQ1, fault diagnosis to RQ2, and the PdM system for rotating machines considering multi-component fault scenarios to RQ3. This chapter summarizes the methods used for each paper corresponding to the research questions and how they were evaluated.

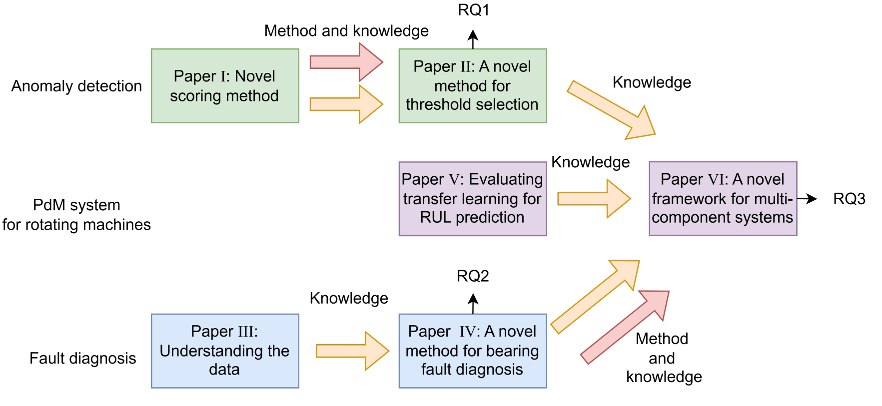


Figure 3.1: Overview of the methodology.

3.1 RQ1: Anomaly Detection

Two different methods related to RQ1 were developed, and they specifically targeted the possibility of using the healthy data to construct robust methods for monitoring the health of machines without needing to rely on historical faults. In addition, because of the availability of healthy data, a model can be developed for each new case, mitigating the challenge of domain discrepancies. For RQ1, paper II is the main contribution, but it is based on the method used and knowledge gained in paper I.

3.1.1 Method for Anomaly Detection Based on Separated Channels

The first method was constructed to support the usage of the complete information of the anomaly detection models' output, which is a limitation of the current work, as highlighted in section 2.1. A separate anomaly scoring method was developed, which is shown in Figure 3.2. It displays how a model is developed using training data, consisting of data describing a healthy machine, where the learning objective is to minimize the mean squared distance from the recreated data and the actual data. This was achieved using a Denoising Convolutional AE (DCAE), consisting of an encoder that compresses the data and a decoder that recreates the data. Additionally, noise is injected into the data input during training to make the algorithm target the most essential features, thereby increasing the model's generalizability. Then, in an online scenario, a scoring method is constructed based on all channels in the model's output, enabling a superior separation between anomalous and normal segments. Firstly, the error for each channel is calculated individually using the mean of a sliding window. Secondly, a simple normalization procedure is applied based on the relation between the score of each channel and its threshold. Lastly, the maximum value of the anomaly score of each channel at a certain point in time can be used to visualize the anomaly prediction to the user, similar to aggregated scoring used in state-of-the-art, but without information loss.

The method was compared against state-of-the-art methods on five benchmark datasets for multivariate anomaly detection, including DL and traditional ML methods. The evaluation considered multiple metrics to capture the model's performance.

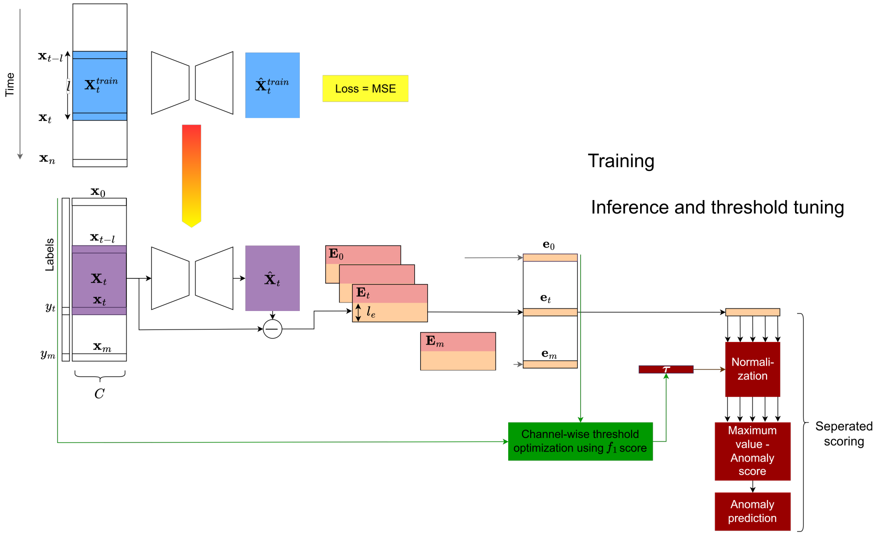


Figure 3.2: An overview of the anomaly detection framework based on separate channel scoring used in paper I.

3.1.2 Threshold-Setting Procedure Based on User Interaction

Because the optimal thresholds for anomaly detection methods are unknown and there are discrepancies in data across machines, an industrial application using an automated threshold-setting procedure becomes highly challenging, which currently is the suggested approach by state-of-the-art as described in section 2.1. As an alternative, a user interaction threshold-setting method called User Validated Thresholds (UVT) was developed based on separate channels, considering the knowledge gained from paper I. The purpose of doing this was to acquire accurate thresholds iteratively with the user's expertise. This method is depicted in 3.3 and is initiated by creating an anomaly detection model using a separate scoring method. Then, a user is shown segments, a collection of anomalous points related in time, and specifies which are anomalous and which are not. The procedure's objective is to provide data to optimize the threshold, but with a limited number of interactions from the user. Firstly, segments are clustered considering four characteristics: time proximity, signal amplitude, anomaly score amplitude, and which sensors show anomalous readings. Then, clustered segments are prioritized based on their attributes so that the user is shown the most important first, ensuring that the most valuable information from the user is extracted before termination. In addition, an automated inclusion strategy for clustered segments similar to the ones the user has already denoted is used to limit the number of interactions without losing information. Finally, when the procedure is terminated based on the number of clustered segments the user denotes as not anomalous, the threshold values are optimized, enabling the method to be used in an online scenario.

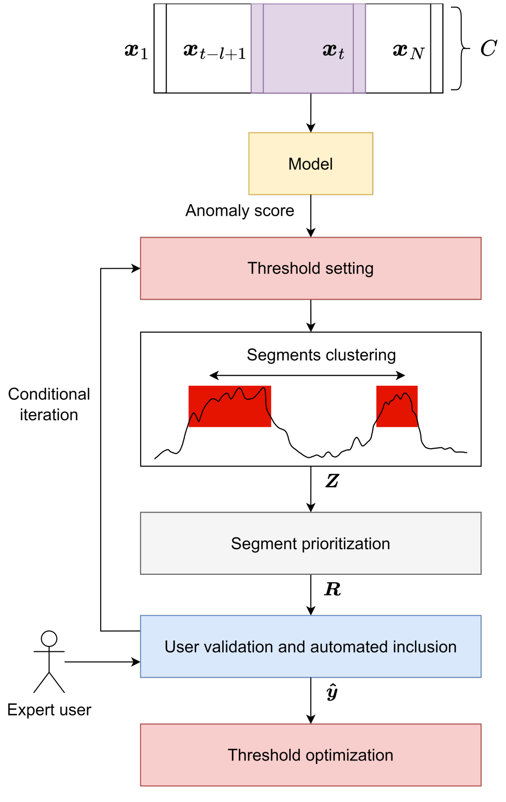


Figure 3.3: A description of the method used in paper II showing how thresholds are optimized based on the user's input.

The method was evaluated using a simulated user and compared against other threshold-setting procedures commonly used in state-of-the-art research on five benchmark multivariate anomaly detection datasets. Two state-of-the-art models, GDN [21] and Omnianomaly [20], were used to increase the generalizability of the findings. Furthermore, the parameters for all methods were kept consistent in the experiment to simulate a deployment in practice. Lastly, multiple metrics were used to capture different aspects of the method's performance.

3.2 RQ2: Fault Diagnosis

The second approach, explicitly targeting RQ2, utilized supervised learning to construct a method using DL for domain generalization for bearing fault diagnosis using vibration data. To achieve this, vibration measurements from an industrial environment were collected and analysed to understand the constraints that need to be considered when developing the method and to be able to validate its performance in practice.

3.2.1 Method for Collecting and Describing a Bearing Dataset from an Industrial Environment

To better understand the characteristics of vibration measurements from an industrial environment and to enable the possibility of validating methods on industrial data, a descriptive study was conducted where vibration measurements of faults were collected, analysed, and published. This data was retrieved from a pulp mill site at SCA. In total, 11 scenarios were extracted, 10 of bearing faults and one of a non-bearing-related fault. The non-bearing-related fault was collected and added to the publicly available dataset to evaluate bearing fault diagnosis methods' ability to distinguish between bearing faults and faults from other components. In addition, for all fault scenarios, a baseline of healthy data was collected for training. Since it is impossible to know precisely when faults occur in an industrial environment (in contrast to the laboratory), the data was labelled manually by transforming the raw signal to the enveloped frequency spectrum using Hilbert transform and FFT, a common method used in previous research [81]. The traits were described considering the amplitude of the raw vibration signal, the time from fault appearance to the bearing being changed, and the rotational speed of the shaft, and compared against the publicly available data from the laboratory environment.

3.2.2 Method for Domain Generalization for Bearing Fault Diagnosis

Based on the increased understanding of the data from an industrial environment, a method for bearing fault diagnosis was developed called Contextual Knowledge-Informed Deep Domain generalization (ConKIDDG), particularly designed to handle challenges C1-C3. The overview of the method is displayed in Figure 3.4. As can be seen, the initial step is to incorporate knowledge about the fault characteristics into the data via a standardization procedure to simplify the training process. This is conducted to reduce the risk of the DL algorithm learning characteristics unrelated to bearing faults. To achieve this, the raw data is first transformed into the enveloped frequency spectrum using Hilbert transform and FFT. Then, as exemplified in Figure 3.5, the data input is standardized using piecewise linear interpolation relating to the fault frequencies of the bearing across domains, which can be calculated based on the bearing type, sampling frequency, and rotational speed of the shaft, which are assumed to be known. The procedure ensures that the fault frequency appears at the same data instance regardless of operational settings and is performed for the different fault scenarios, which are inner ring, outer ring, and rolling element faults, creating three standardized frequency spectrums from the original raw signal.

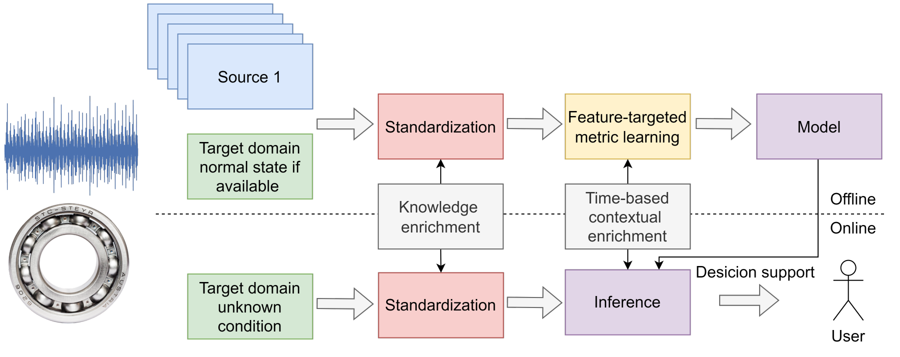


Figure 3.4: An overview of the method suggested in paper IV showing how the model is developed using knowledge and time-based contextual enrichment and a transfer learning technique.

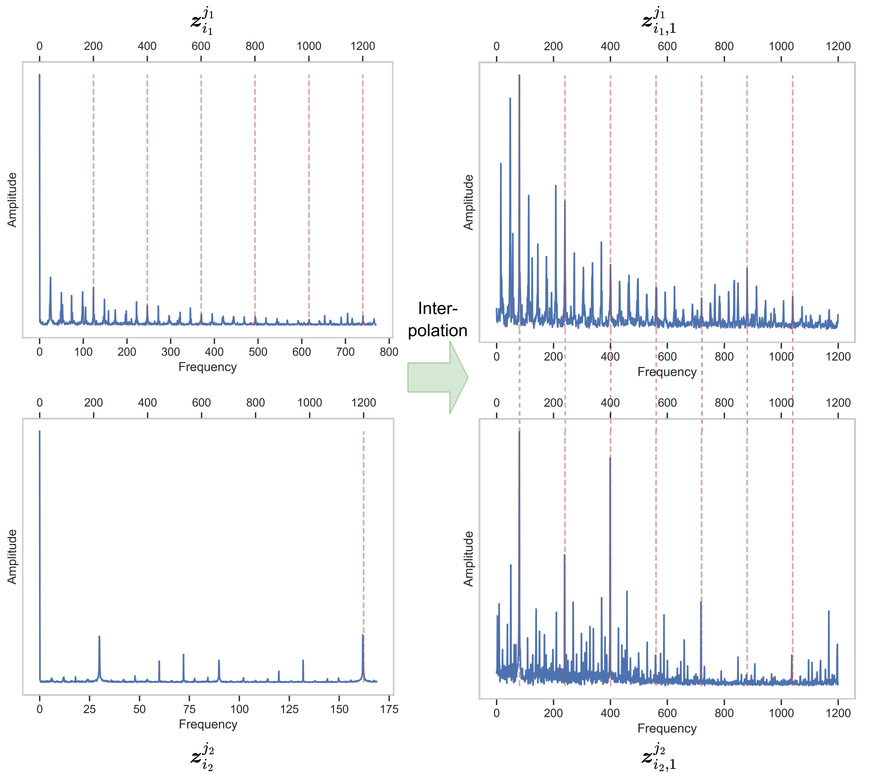


Figure 3.5: Illustration of the knowledge enrichment from paper IV based on the inner ring frequency. The fault frequency and its harmonics are shown as the dashed red line.

The DL model uses three feature extractors to match the number of standardized frequency spectrums to target the different fault scenarios, as seen in 3.6. To further reduce the impact of discrepancies between domains, a statistics-based transfer learning strategy is employed, which minimizes the distance between features extracted from the bottleneck belonging to the same class and simultaneously maximizes the distance between features of different classes, regardless of domain. Lastly, to reduce false positives considering the impact of randomness, a time-based contextual model is incorporated that uses samples from previous measurements using the output from the bottleneck of model 1 to make the final prediction.

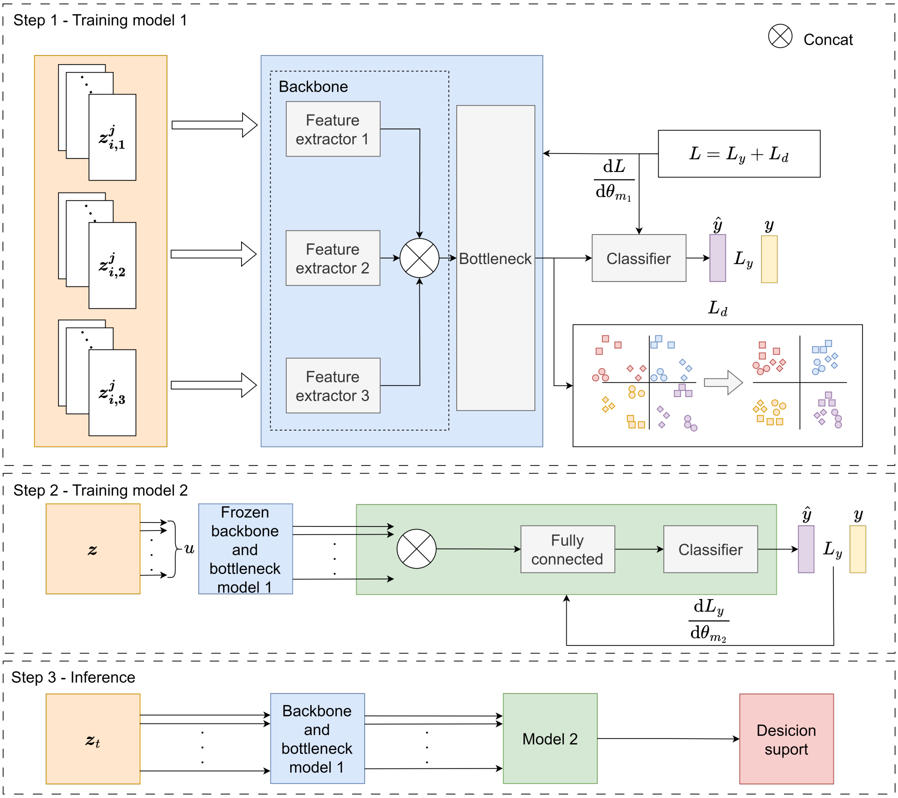


Figure 3.6: The model developed in paper IV for domain generalization for bearing fault diagnosis.

The method was compared against state-of-the-art methods in three cases, including 27 tasks from multiple datasets, such as the published bearing dataset from paper III, expected in industrial scenarios. These covered the scenarios of cross-machine deployment and the unavailability of historical faults from the target machine. The parameters for the method were kept consistent throughout the experiment to simulate an industrial application.

3.3 RQ3: PdM System for Rotating Machines

Much of the knowledge gathered from methods for RQ1 and RQ2 could be used when targeting RQ3. However, to develop a PdM system for multi-component fault scenarios, it is also essential to understand what methods are suitable for RUL prediction of mechanical components, considering challenges C1-C3. Therefore, a study targeting the state-of-the-art methodologies for transfer learning was tested in paper V. The knowledge from this paper and the previous papers was used to form a framework in paper VI.

3.3.1 Method for Evaluation Transfer Learning for Bearing RUL Prediction

As mentioned in section 2.3, a common method for RUL prediction of mechanical components such as bearings using vibration measurements is to use transfer learning. However, there is a lack of studies evaluating these methods in industrial scenarios. Because of this, the study implemented a method based on state-of-the-art methodologies and tested it on scenarios based on challenges C1-C3. An adversarial method was used to achieve this, considering it is one of the most common strategies applied in current research. The implementation was based on DANN [74] described in section 2.3, and the source domains used during training were healthy data from the target domain and several scenarios from the laboratory environment of bearing run-to-failure experiments. This means that the data input was associated with a domain label used for the adversarial loss function based on the output of the domain discriminator, and a label for the RUL prediction. The objective of the RUL regressor was to map the input data to the degradation value between 1, healthy, and 0, unhealthy. To increase the generalizability of the study, two different common input data types were used: raw data and time-frequency data based on CWT. In the evaluation, to establish that the implemented method could perform similarly to state-of-the-art and thereby increase the experiment's validity, it was initially compared against experiments in previous research. When confirmed that the method had comparable performance, it was evaluated using multiple datasets on four scenarios: transfer between operational conditions on the same machine, different machines with similar settings, different machines with larger differences in settings, and non-bearing related faults.

3.3.2 A Framework for Multi-Component Systems

Based on the knowledge gained from the previous studies, a hybrid framework called Rotating Machinery Prognostic Framework (RoMaP) was suggested for multi-component fault scenarios that leverage the advancements in anomaly detection, fault diagnosis, and RUL prediction shown in Figure 3.7. In this framework, a model is developed for each component for both fault diagnosis and RUL prediction. The reason for this separation is the variances in fault characteristics and degradation patterns between different components, making it challenging to create an end-to-end solution, especially considering the discrepancies between machines. These models are combined with anomaly detection based on a general health indicator, which means that it is developed to detect all types of faults in the machine. The purpose of using this is to enable system deployment without incorporating a model for each component in the machine, making it highly dynamic. In addition, it is also used to validate the prediction of the component-specific diagnosis model to reduce the false positives and as input for the component-specific RUL prediction model. When more component-specific models are added to the framework, less consideration can be given to the general anomaly detection, reducing the impact of false alarms.

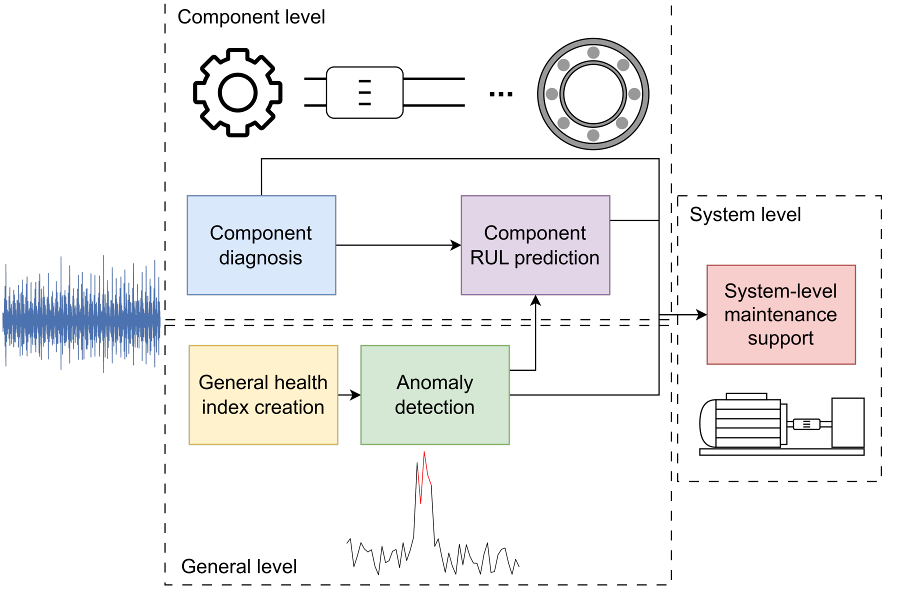


Figure 3.7: The hybrid framework suggested in paper VI using anomaly detection and component-specific models for fault diagnosis and RUL prediction.

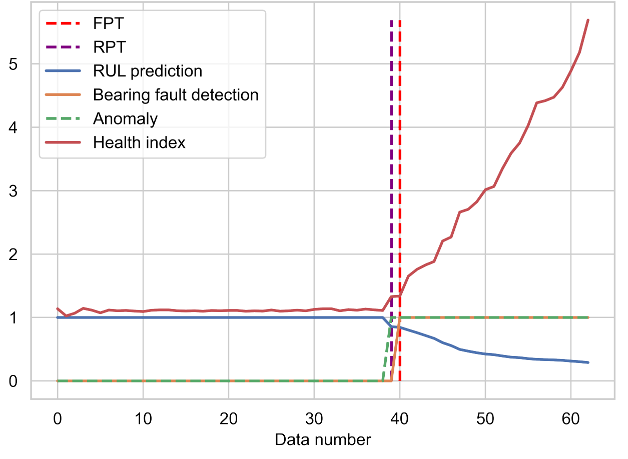


Figure 3.8: An example of the procedure from paper VI showing that when the component-specific diagnosis model finds a fault and an anomaly is predicted, the RUL estimation is initiated.

An example of the procedure considering a bearing fault is shown in Figure 3.8. When the fault diagnosis model for bearings detects a fault and an anomaly is predicted, the component-specific model for bearing RUL is initiated using the general health index as input. At time FPT, the user is notified of a bearing degradation. RPT is the first point that the RUL prediction model uses to estimate the degradation from healthy, 1, to failure, 0, using the health index.

An instance was deployed based on a Convolutional Autoencoder (CAE) for the health index, the method from paper IV, ConnKIDDG for fault diagnosis, and Bidirectional-LSTM (Bi-LSTM) for RUL prediction in a case study to evaluate the framework's potential and compare it to alternative methods based on the state-of-the-art. In total, 27 tasks from multiple datasets were considered, including deploying the method on a new machine, tasks from an industrial environment using the dataset from paper III, no labels of faults from the deployment target, and the occurrence of faults from different components. The component-specific diagnosis and RUL prediction models were developed for bearing fault scenarios.

Results

This chapter summarizes the key findings of the research. An overview of the outcome from each research area and a summarized answer to the research questions is displayed in Table 4.1. These aspects will be described in the following sections.

Table 4.1: Summary of results and the answer to the research questions.

Research area	RQ	Result outcome
Anomaly detection	RQ1	The value of separated anomaly scoring combined with a user interaction method to optimize the thresholds of multivariate anomaly detection models.
	RQ3	Challenge of setting thresholds for anomaly detection methods and limitation of relying on only anomaly detection due to the expectation of false alarms.
Fault diagnosis	RQ2, RQ3	Possibility of using prior knowledge and transfer learning techniques for domain generalization on rotating components using vibration data.
PdM system	RQ3	Challenge of using current transfer learning methodologies for RUL prediction and the value of a hybrid framework for a PdM system for rotating machines utilizing anomaly detection, component-specific fault diagnosis, and RUL prediction.

4.1 RQ1: Anomaly Detection

The studies for multivariate anomaly detection showed that methods based on DL are promising in finding faults, particularly since they only need historical healthy data from the machine. However, they face difficulties when being deployed in practice, considering the challenge of optimal configuration on a general basis. It showed that the accuracy is highly dependent on the scoring method and the threshold, consistent with previous work [23]. This was shown in the result of paper I displayed in Table 4.2, where the developed $DCAE_s$ using a separated anomaly scoring outperformed all compared state-of-the-art methods, which employ an aggregated scoring method. It can also be seen that when an aggregated method was used using the same model, $DCAE_a$, the results were significantly lower, suggesting that the most essential factor for the superiority is the scoring method, not the underlying model.

Table 4.2: Average accuracy for methods in paper I using multiple datasets and two different metrics. Higher value is better.

Method	**Average score**	
	f_{pa1}	f_{p1}
USAD [18]	0.789	0.512
OmniAnomaly [20]	0.779	0.506
LSTM-VAE [82]	0.721	0.501
DAGMM [83]	0.651	0.336
MAD-GAN [19]	-	0.570
OCSVM [84]	0.579	0.536
LOF [85]	0.601	0.302
IF [86]	0.789	0.516
PCA [87]	0.781	0.590
ICA [87]	0.692	0.496
$DCAE_a$	0.798	0.595
$DCAE_s$ (our)	**0.9181**	**0.794**

Furthermore, in paper II, where the user interaction threshold setting procedure called UVT was suggested, a significant increase in performance was shown compared to when using methods from state-of-the-art research, as can be seen in Figure 4.1. Interestingly, when using the same parameters for the state-of-the-art threshold setting methods across all tasks, which is the expected scenario in industrial applications, the performance is poor, highlighting the difficulty of applying these in practice.

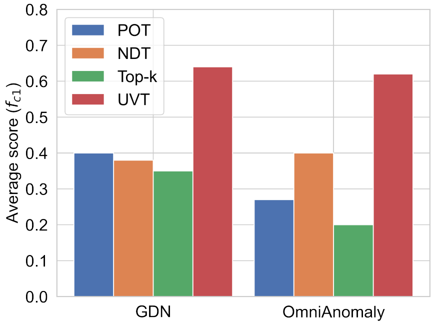


Figure 4.1: The average accuracy score for the different threshold setting methods from paper II. UVT is the suggested method with the highest score.

The result from could be achieved with a relatively small number of interactions from the simulated user seen in Figure 4.2a, which also displays the importance of the different parts of the method and the significant difference when not using . It is essential to evaluate the number of interactions relative to the accuracy, shown in 4.2b, which collectively demonstrates the superiority of the suggested approach. Nevertheless, it is also apparent that this procedure requires effort from the user for every new machine to be monitored, creating some practical challenges. In addition, the result shows that despite increased performance compared to state-of-the-art, applying solely anomaly detection as a general method is challenging, considering the selection of optimal threshold and the induced risk of false alarms or the inability to capture faulty events.

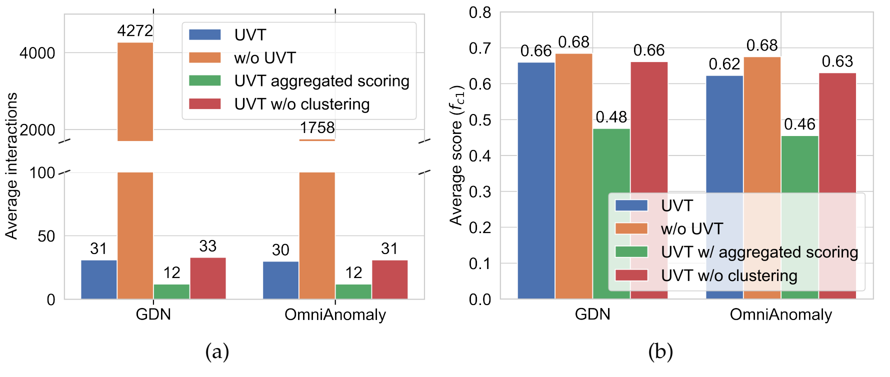


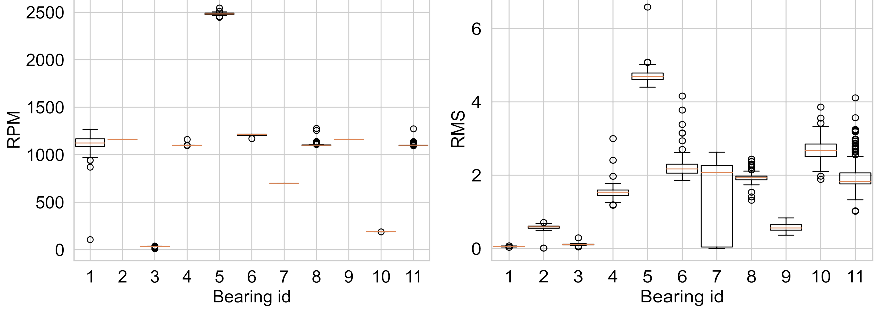
Figure 4.2: Ablation study of the proposed method from paper II considering different aspects. (a) Average number of user interactions. (b) Average anomaly detection accuracy after threshold optimization.

4.2 RQ2: Fault Diagnosis

This section describes the studies related to RQ2, including characterizing vibration measurements from an industrial environment from paper III and evaluating the method for bearing fault diagnosis in paper IV.

4.2.1 Description of a Bearing Dataset from an Industrial Environment

When examining the characteristics of the vibration measurements from an industrial environment, high discrepancies and a lack of historical data for each component were observed. Firstly, the rotational speed of the shaft varied significantly between the different cases, as seen in Figure 4.3a. Secondly, shown in Figure 4.3b, the vibration amplitude of the healthy data described using RMS also differed, indicating a difference in, for example, noise and load between the cases. Lastly, the development time from the apparent fault to a change of bearing also varied substantially 4.4. This suggests that methods for RUL prediction of bearings need to manage differences in degradation. Moreover, fault scenarios were sparse for the different cases, and all bearings had only a single fault available for the duration of the evaluation period of four years. In addition, it is worth considering that the overwhelming majority of bearings in the factory did not have any historical faults and were therefore not included in the evaluation. Overall, the result underlines the challenges C1 and C2 and shows that methods developed for components in rotating machines using vibration data must be robust and have a high level of generalizability across different machines to work in practical applications.



(a) (b)

Figure 4.3: Diagrams of the different cases from paper III. (a) The rotational speed of the shaft in RPM. (b) The RMS amplitude of the vibration measurements when all components are healthy.

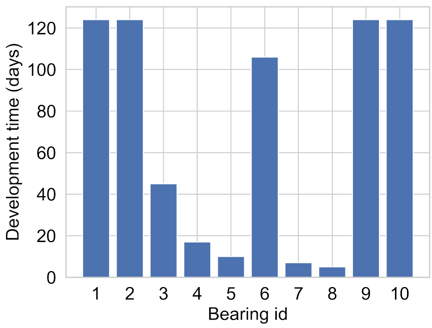


Figure 4.4: The development time from paper III between the faults showing and the bearings being changed.

4.2.2 Evaluation of Method for Domain Generalization for Bearing Fault Diagnosis

The result from the method based on domain generalization for bearing fault diagnosis is displayed in Figure 4.5, which shows the average performance for the three cases, each containing various tasks (seven or more each) for domain generalization evaluation. It can be seen that the suggested method, ConKIDDG, performed the best across all the different cases compared to the best-performing state-of-the-art alternative. The third case was the industrial dataset from paper III, where no results from the other methods were available, but where the method demonstrated excellent performance.

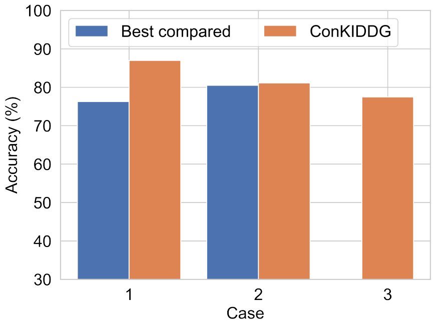


Figure 4.5: The average result from the different cases from paper IV comparing the suggested method to the best compared state-of-the-art method.

The importance of knowledge enrichment based on standardization was also apparent and is observable in Figure 4.6, compared to not using standardization, w/o standardization, as the most significant factor contributing to the method's high accuracy. This result, supported by previous research [52], suggests that using knowledge of fault characteristics is highly beneficial when constructing a general solution for machine fault diagnosis. In addition, it can be seen that the time-based context using model 2 and the transfer learning technique based on metric learning were highly influential to the model's performance.

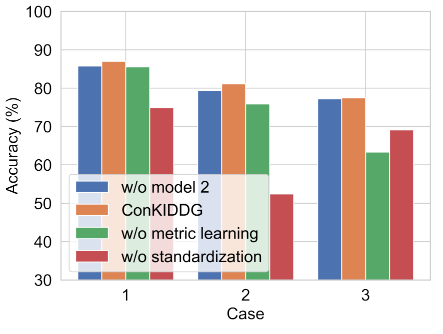
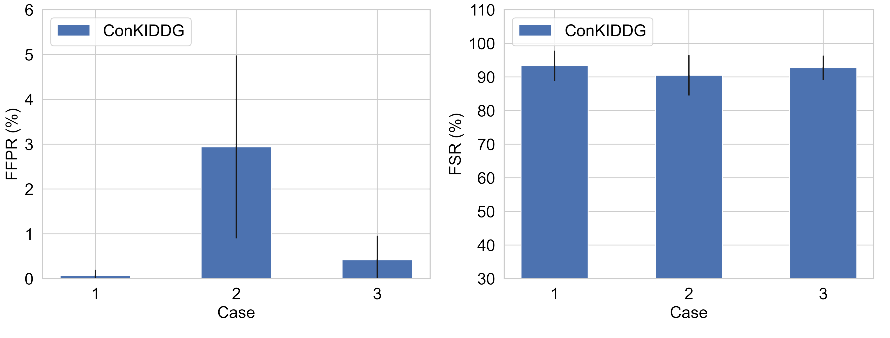


Figure 4.6: Ablation study from paper IV of the different parts of ConKIDDG.

Considering that most of the operations of a mechanical component in industrial environments are healthy, as supported by the description in paper III, the false positive rate was also evaluated for the method. It can be seen in Figure 4.7a that the suggested method is robust against noise and abnormal events in the healthy data. At the same time, it is easy to develop a method that does not generate false positives by ignoring faulty segments. However, it can be observed that the suggested method simultaneously had a high fault segment recall displayed in Figure 4.7b.



(a) (b)

Figure 4.7: Tables showing the average accuracy for ConKIDDG from paper IV using different metrics. (a) Fault false positive rate. (b) Fault segment recall.

Lastly, when applying a method as part of a semi-autonomaous or autonomous PdM system, it is crucial to consider robustness and trustworthiness from a user perspective. This is inherently difficult for DL-based methods to achieve due to their complexity, and is a significant reason for the potential value of Explainable AI (XAI) methods [88]. An advantage of using the standardization procedure based on the enveloped frequency spectrum as input is that it can be used to give user feedback about the model's prediction based on the known fault characteristics. An example is shown in Figure 4.8 using saliency mapping related to the models' predicted bearing health state using the implementation by [89] and highlights the features in the input data that had the highest importance for the classification. The red dotted lines are the corresponding fault frequencies of the bearing for the predicted fault (except for Figures 4.8c where they correspond to the ground truth). If the model correctly understands the fault characteristics, peaks related to the fault frequencies for the predicted fault should be highlighted, which occur for Figure 4.8a and 4.8b, where the model correctly predicts an outer ring fault and an inner ring fault, respectively. In those scenarios, it can contribute to an increase in the model's trustworthiness. In addition, if the model highlights segments that do not relate to the ground truth bearing health state, for example, Figures 4.8c and 4.8d, where the model misclassified the health state, more training examples can be used to increase the robustness of the model.

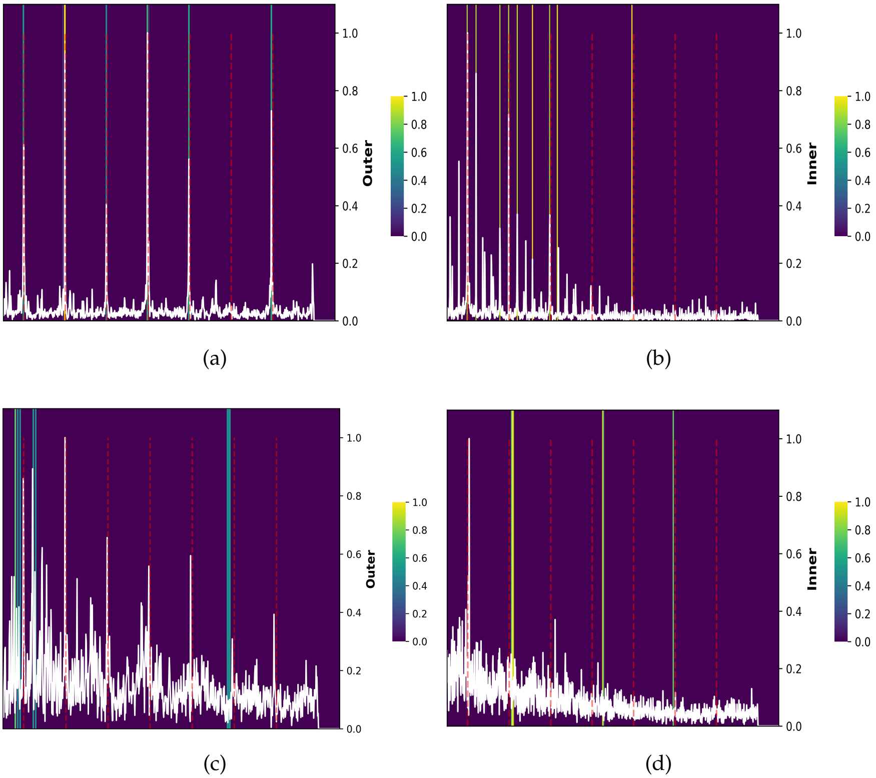


Figure 4.8: Examples of user feedback from paper IV. The red dotted line shows the fault frequency of the bearing, and the highlighted features display what is most important for the model's prediction. (a) Correct prediction of an outer ring fault. (b) Correct prediction of an inner ring fault. (c) An incorrect classification of a rolling element fault when the ground truth is an outer ring fault. (d) An incorrect classification of an inner ring fault when the ground truth is healthy.

4.3 RQ3: PdM System for Rotating Machines

The summary of the result targeting RQ3, including the evaluation of state-of-the-art transfer learning methodologies for bearing RUL prediction from paper V and the suggested framework for rotating machines, developed and submitted in paper VI, is described in the following sections.

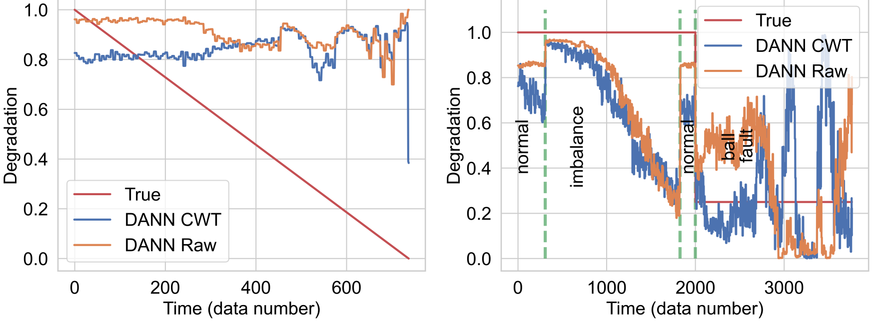
4.3.1 Evaluation of Transfer Learning for RUL Prediction

In the result for paper V, it became apparent that using transfer learning methods based on state-of-the-art methodologies is challenging when considering scenarios expected in industrial applications. Firstly, to demonstrate the relevance of the experiment, the implementation of DANN was compared against state-of-the-art on transfer between operational conditions seen in Table 4.3. It shows that the implemented method had a similar accuracy to the state-of-the-art when considering the commonly used Root Mean Squared Error (RMSE) metric, which measures how close the estimated RUL prediction is to the ground truth degradation.

Table 4.3: Average RMSE for the different tasks compared to [90] from paper V. Lower values mean a better result.

DDAN [90]	**CNN+TL [90]**	**RNN+TL [90]**	**DCNN [91]**	**LSTM [92]**	**BLSTM [93]**	**DANN Raw (our)**
0.191	0.244	0.287	0.251	0.268	0.250	0.225

However, when transfer between machines, rarely done in current research and expected in practice (see section 2.3), a substantial decrease in performance was observed, as exemplified in Figure 4.9a where the method is unable to detect the degradation of the bearing. In addition, Figure 4.9b shows an example from the performance when considering a non-bearing related scenario. In those cases, the model reacted as if there were bearing faults. This result strongly indicates that the model understands little about the bearing's fault characteristics and instead considers the amplitude in the vibration data. It also suggests that defining the degradation start is necessary to construct a prognostic method with high generalizability and robustness in industrial applications.



(a) (b)

Figure 4.9: The RUL prediction of the model with CWT and the raw data on different tasks from paper V. (a) Deployment on a new machine. (b) Deployment on a new machine with a non-bearing-related fault.

4.3.2 Evaluation of Framework for Multi-Component Systems

In paper VI, the framework RoMaP was evaluated using an instance based on the methods and knowledge obtained from previous studies, considering bearing fault scenarios for the component-specific models. Table 4.4 shows the average performance for the 27 tasks considering different parameters for the threshold selection that determines when the method detects a degradation in any of the machine's components. A lower value means the method is more sensitive, and a higher value means the method is less sensitive. In the table, RMSE measures the RUL prediction accuracy, MF the number of missed fault scenarios, FATD the delay from the actual degradation start to the predicted degradation start, and FP the number of false positives. As can be seen, the suggested method is, in contrast to the other methods, not dependent on an optimal threshold to generate high accuracy across different metrics. This is highly beneficial in practice because, as was shown in paper II, the thresholds are unknown in advance, differ between cases, and are challenging to set. In addition, the implemented method achieved the highest accuracy on all metrics except for FATD, meaning it is slightly slower to detect the fault than the other methods. This means that it is excellent at capturing fault scenarios and determining the degradation of the component without generating false alarms for the user.

Table 4.4: Average performance of the compared methods with different thresholds for anomaly detection from paper VI. Lower value is best for all metrics.

Metric	Method	1σ	2σ	3σ	4σ	5σ
FP	GIF_{CAE}	245.04	58.00	29.30	11.89	9.78
	SIF_{FRMS}	256.04	177.07	112.84	42.52	15.87
	SGIF_{FRMS}	87.44	44.00	18.30	13.15	9.52
	$\text{RoMaP}_{ConKIDDG}$ (our)	**0**	**0**	**0**	**0**	**0**
FATD	GIF_{CAE}	**4.19**	**8.65**	**15.13**	18.87	**21.22**
	SIF_{FRMS}	6.74	11.22	16.09	19.74	22.48
	SGIF_{FRMS}	12.17	13.13	16.78	**18.83**	22.91
	$\text{RoMaP}_{ConKIDDG}$ (our)	14.70	19.48	20.09	23.91	25.91
RMSE	GIF_{CAE}	0.24	**0.17**	0.19	0.22	0.17
	SIF_{FRMS}	0.27	0.23	0.24	0.23	0.22
	SGIF_{FRMS}	0.24	0.24	0.23	0.22	0.22
	$\text{RoMaP}_{ConKIDDG}$ (our)	**0.19**	0.20	**0.16**	**0.17**	**0.17**
MF	GIF_{CAE}	4	4	4	4	4
	SIF_{FRMS}	3	3	2	**0**	2
	SGIF_{FRMS}	3	3	2	2	3
	$\text{RoMaP}_{ConKIDDG}$ (our)	**0**	**0**	**0**	**0**	**1**

An example of the difference between the suggested method and one of the compared approaches considering sensitivity is depicted in Figure 4.10. It shows the bearing degradation prediction at the bottom and the bearing diagnosis at the top, with -1 being an anomaly, 1 a bearing inner ring fault, 2 a bearing rolling element fault, and 3 a bearing outer ring fault. It can be seen that when using a low threshold, in this case 2σ as the threshold-setting parameter, the compared method performs poorly, because it generates false alarms and has a low accuracy when predicting the degradation. In contrast, the suggested method gives accurate predictions with the same threshold.

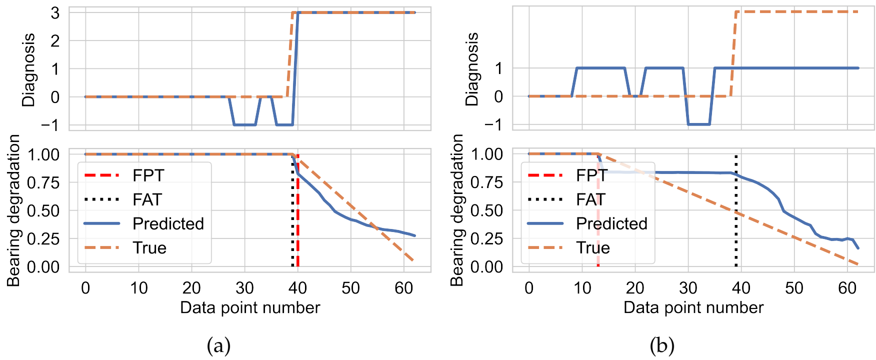
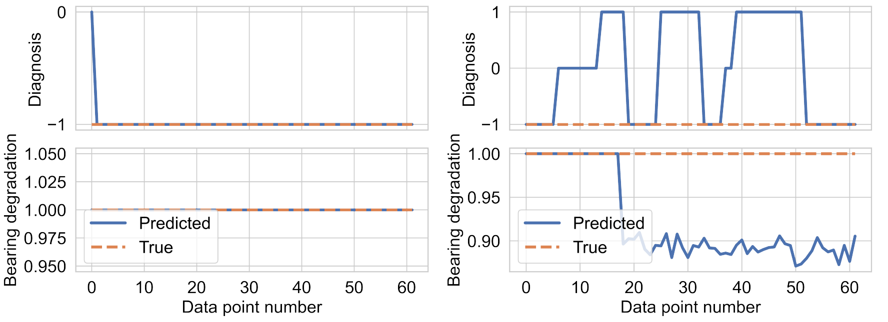


Figure 4.10: Performance considering the threshold-setting paremeter 2σ with different methods from paper VI. In the diagnosis, -1 means an anomaly, 1 a bearing inner ring fault, 2 a rolling element fault, and 3 a bearing outer ring fault. (a) The suggested method RoMaP$_{ConKIDDG}$. (b) SIF$_{FRMS}$.

When considering non-bearing related faults, the suggested method correctly identified them as anomalies, while the compared approaches regarded them as bearing faults, as exemplified in Figure 4.11. This means the user would be told to change a bearing based on an estimated RUL prediction when, in reality, no fault is occurring in the bearing. This is a significant issue primarily because it can lead to machine failure if the failing component has a shorter degradation time than the bearing, or should be maintained as soon as a fault occurs. With the suggested method, the user would instead be told to physically examine the machine and be notified that there is no bearing fault, which can be helpful when diagnosing and deciding on appropriate maintenance action.



(a) (b)

Figure 4.11: Performance when considering non-related faults for different methods from paper VI. In the diagnosis, -1 means an anomaly, 1 a bearing inner ring fault, 2 a rolling element fault, and 3 a bearing outer ring fault. (a) The suggested method RoMaP$_{ConKIDDG}$. (b) SIF$_{FRMS}$.

Lastly, increased validation in the diagnosis can be greatly beneficial because it can help reduce the number of false alarms, considering that no diagnosis method is entirely accurate. An example can be observed in Figure 4.12 where the diagnosis model detects a fault, but the user is not notified because no anomaly is observed.

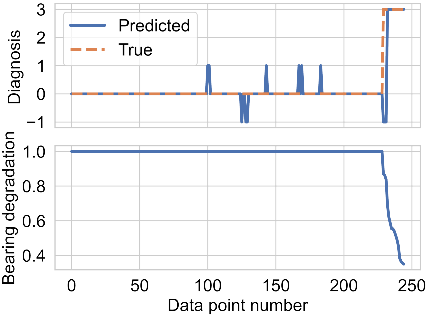


Figure 4.12: Performance of the suggested method RoMaP$_{ConKIDDG}$ from paper VI in a scenario where the diagnosis model incorrectly predicts a fault.

Discussion

This chapter summarizes the key findings of the research and focuses on its impact and implications in research on data-driven DL methods. It also discusses some of the limitations and suggests future research directions.

5.1 Impact and Implications

This work has contributed to the body of research by providing methods and valuable insights for deploying data-driven PdM solutions from a practical perspective. As described in Chapter 2, much research has been conducted concerning fault diagnosis, anomaly detection, and machine health prognosis. Still, most studies do not consider the extent to which developed methods work in industrial scenarios. This includes the performance perspective and the feasibility of implementation. To enable more research to target scenarios expected in industrial applications, multiple previous works [28], [29] have advocated for underlining the necessary characteristics of DL methods based on vibration data to work in practice and providing industrial data for validation. Considering this, the description and published dataset related to paper III have the potential to enable the development and evaluation of methods based on real scenarios. Based on the described characteristics, it is also of value to highlight limitations of state-of-the-art methodologies to motivate future research to find new approaches instead of relying on methods with little viability in practice. In this context, paper II demonstrates the issue with state-of-the-art methods for threshold selection and the need for a robust threshold-setting procedure for these to be practically usable. In addition, paper V illustrated the challenge of applying state-of-the-art methodologies for transfer learning for RUL prediction of bearings in industrial scenarios, suggesting that more research should consider the fault characteristics when developing these methods.

Based on the limitations of these approaches, this research also presents methods that have been tested in industrial scenarios, displaying a realistic possibility of successful deployment in practice. This includes: (1) the threshold-setting procedure in paper II that enables the creation of optimal thresholds for anomaly detection, (2) the method for domain generalization for bearing fault diagnosis in paper IV and the prognostics framework for rotating machines suggested in paper VI. These methods have the potential to work because they are rigorously evaluated in various tasks from different datasets, evaluated considering challenge C1-C3, and have consistent parameter selection throughout the experiment. Based on this evaluation methodology, this research has contributed to a research direction for PdM methods that, to a larger extent, focuses on the industrial deployment perspective. It is established on the foundation that it is easy to make a method perform well in some handpicked scenarios, but challenging to achieve accurate and robust prediction on a general basis in expected scenarios. This does not mean there is no room for some unrealistic evaluation procedures. For example, it can be beneficial to optimize the training parameters for a single dataset to compare its theoretical performance against other methods, even though this is not possible in practice because it is a consistent, fair, and straightforward approach. However, there needs to be a general shift towards methods developed and evaluated, considering, for example, challenges C1-C3 based on scenarios that need to be managed by these methods. Only then will we see a possibility for industries to adopt the research output, which ultimately needs to be the goal.

Some of the most significant insights from this research are the benefits of using knowledge enrichment in the training procedure and employing hybrid methods that utilize the advantages of each model type. This can enforce a method for PdM to learn the fault characteristics, reducing the impact of randomness and creating dynamic, robust, and highly generalizable systems. This research has targeted PdM, particularly rotating machines, but these insights can likely be of value to be incorporated in other fields.

One of the most significant enabling factors of this research has been the industry collaboration, which, for example, made it possible to publish data from an industrial environment, gain a greater understanding of the practical challenges, and validate methods in practice. This experience suggests that cooperation is imperative, mutually beneficial, and the most promising way to incorporate research into practice. Considering this, a hopeful implication of this work is that it can inspire future collaboration.

5.2 Sustainability and Ethical Aspects

Two predominant aspects of sustainability to consider for the methods suggested in this thesis are the energy and resource perspectives. Firstly, evaluating DL-based methods' energy demand is generally relevant due to having a higher computational cost than traditional methods. Because this research is only based on time series data, which is not as demanding as, for example, images, and the suggested methods processing time is relatively fast, as shown in paper IV, it is argued that the negative implication of the suggested method from an energy perspective is limited. Secondly, given that the proposed methods aim to reduce failures but simultaneously make machines and components run their whole lifespan and thereby reduce the use of resources and energy when constructing and changing these, the suggested solutions should have an overwhelmingly positive effect from a sustainability perspective.

In addition, as has been described in [94], AI-based solutions for PdM can have ethical implications that are worthy of consideration, including accountability, transparency, and social-economic aspects related to an increased level of automation. Particularly, it is essential to consider the interaction with the user, how the output of the models should be interpreted, and who should be accountable if a wrongful action is carried out. In this context, this thesis has made an effort to consider and describe the interaction with the users and the suggested methods, such as the user feedback method in paper IV and explaining the expected user actions based on the output of the method in paper VI. It is also vital to consider that AI has many positive ethical aspects for predictive maintenance, including fewer machine failures, especially considering dangerous equipment, and less time spent conducting maintenance actions with little planning, which can lead to reduced injuries.

5.3 Limitations

One of the significant challenges with the research has been to evaluate the proposed methods on industrial data, since it is often lacking. A limitation, for example, with the method in paper II, also acknowledged by [17], is that it is not evaluated in an industrial context with real users. Furthermore, when data can be collected from an industrial environment, achieving the same quality as in the laboratory is difficult. This became apparent in paper III describing the published industrial dataset of bearing faults. The limiting factors were, for example, the challenge of labelling events and a lower sampling rate than recommended by previous studies [95]. However, it is essential to consider that extracting a dataset consisting of, for example, vibration measurements of high quality from a laboratory environment is much easier than in an industrial environment, and that lower quality should be expected when deploying methods in practice. Lastly, to increase the generalizability, especially considering paper VI, it would have been of great value if another method for component diagnosis and RUL prediction had been suggested or adopted.

5.4 Future Research

The most significant issue with state-of-the-art research is that few studies consider the factors expected in industrial applications. There is a discrepancy between the research-defined scenarios for laboratory-based experiments and scenarios found in industry. Because of this, multiple future research areas will be beneficial in increasing the possibility of industrial adoption of the methods suggested in research. These are listed and described in the following subsections.

5.4.1 Evaluating Methodology Based on Industrial Scenarios

Future research should consider industrial online deployment, especially challenges C1-C3, when developing and evaluating methods. Based on the insights gained from the research, the following factors should be adopted when evaluating methods for rotating machines:

- Test the method on a scenario deployed on a different machine than the one on which the model was trained.
- Test the method on tasks with various operational settings.
- Use the same parameters for the method throughout the experiment.
- Test the method on fault scenarios from more than one component.

5.4.2 Explainability of Methods

When considering the adoption of automatic or semi-automatic monitoring systems based on DL methods, it is of great importance to be able to interpret and enable validation of their output. This is related to trustworthiness, which is likely required for these systems to be successful in practice, and has also been highlighted by multiple previous studies [5], [7], [9]. Therefore, future research should focus on the interaction between the user and these methods using XAI and related techniques, such as the suggestion in paper IV.

5.4.3 Multi-Component Fault Scenarios

As multiple previous studies have highlighted [1], [5], most research focuses on single-component fault scenarios. The research for this scenario will still be valuable in the future, provided they evaluate the method properly (see subsection 5.4.1). However, it is necessary to develop methods that handle multi-component fault scenarios because these will occur in industrial environments. Therefore, more research similar to the suggested framework in paper VI is suggested.

5.4.4 Usage of Hybrid Methods

This thesis has, in conjunction with previous work [1], [5], [7], demonstrated the value of utilizing multiple methods to limit the disadvantage of each method type and use their respective advantage. An example of this was shown in paper VI. More research should focus on using multiple model types and different data in fusion to increase the robustness of methods for industrial applications. For rotating machines, it is, for example, valuable to consider sensors such as temperature and current alongside vibration measurements to improve the robustness of methods. Regarding the selection of model type, DL-based methods have been the main approach used for developing models in this work, which can be particularly useful when considering fault diagnosis using vibration data, supported by the result in paper VI. However, due to their complexity and low interpretability, it is highly recommended to incorporate simpler methods in scenarios where they can complete the objective as well or better.

5.4.5 Adaptivity and Robustness

A challenge, also acknowledged in previous research [6], when deploying methods for PdM in autonomous or semi-autonomous scenarios is the continuous changes to the machine being monitored or the production process affecting the machine. A key aspect to handle these aspects in future work is models for domain generalization that target the fault characteristics of components and can disregard parts of the signal that change over time unrelated to any fault, for example, using a knowledge enrichment procedure and transfer learning, such as suggested paper IV. Despite this, it is expected that models' accuracy will degrade over time, and it is therefore valuable for future research to develop mechanisms to manage this, for example, by retraining on baseline healthy data.

Conclusion

Due to the limited attention in existing research to the real-world conditions under which PdM systems are deployed, this thesis proposes methods that are both practically applicable and robust across industrial settings. The overarching goal was to reduce the gap between controlled, research-defined scenarios and the variability and constraints found in practice. To this end, three research questions (RQ1–RQ3) were formulated—each motivated by challenges identified in industrial environments, including the lack of labeled fault data, variability across machines, and the need to handle multi-component systems.

In response to RQ1, a threshold-setting method based on user interaction was suggested in paper II utilizing separate scoring presented in paper I. The result showed that the method can optimize the performance of multivariate anomaly detection methods and simultaneously outperform alternative methods based on state-of-the-art.

For RQ2, a novel bearing fault diagnosis method for domain generalization using vibration data was proposed in paper IV based on knowledge enrichment, time-based contextual enrichment, and a transfer learning technique. It can be used on any new rotating machine without needing historical faults and was constructed considering industrial data published and described in paper III.

RQ3 was addressed in paper VI using insights gained from papers II-V, where a prognostic hybrid framework for multi-component rotating machines based on vibration data was proposed. It utilizes advancements in anomaly detection, fault diagnosis, and RUL prediction to enable robustness and accuracy across various scenarios expected in industrial applications.

In summary, two of this research's significant insights are the value of using prior knowledge to enrich the training procedure and using hybrid methods to utilize different models' advantages. These can contribute to creating robust, dynamic, and highly generalizable systems. Although these aspects are particularly valuable when developing PdM methods for rotating machines, they are likely also valuable when incorporated into other fields.

Apart from the suggested methods for industrial scenarios, this thesis has provided insights into valuable future research areas related to transferring methods developed in research into practice, including essential evaluation criteria such as cross-machine and multiple fault scenarios. It has also suggested using hybrid methods, research for XAI, and developing adaptive methods for multi-component scenarios.

This work has narrowed the gap between research-defined scenarios and those expected in practice by addressing the research questions and providing valuable insights. Hopefully, this can inspire future work to develop methods that challenge what is possible for AI-based methods for PdM. However, despite being challenging, a focus shift from minor improvements to scenarios rarely or never seen in practice towards practical scenarios is necessary in future research to develop methods that are also valuable outside the research community. This means it is essential that when conceptualizing, developing, and evaluating new methods for PdM, the primary attention should be on the feasibility and practicality of the solution.

Bibliography

[1] J. J. Montero Jimenez, S. Schwartz, R. Vingerhoeds, B. Grabot, and M. Salaün, "Towards multi-model approaches to predictive maintenance: A systematic literature survey on diagnostics and prognostics," *Journal of Manufacturing Systems*, vol. 56, pp. 539–557, Jul. 2020.

[2] Y. Wen, M. Fashiar Rahman, H. Xu, and T.-L. B. Tseng, "Recent advances and trends of predictive maintenance from data-driven machine prognostics perspective," en, *Measurement*, vol. 187, p. 110 276, Jan. 2022.

[3] S. Gupta, A. Kumar, and J. Maiti, "A critical review on system architecture, techniques, trends and challenges in intelligent predictive maintenance," *Safety Science*, vol. 177, p. 106 590, Sep. 2024.

[4] C. Huang, S. Bu, H. H. Lee, C. H. Chan, S. W. Kong, and W. K. C. Yung, "Prognostics and health management for predictive maintenance: A review," *Journal of Manufacturing Systems*, vol. 75, pp. 78–101, Aug. 2024.

[5] D. Neupane, M. R. Bouadjenek, R. Dazeley, and S. Aryal, "Data-driven machinery fault diagnosis: A comprehensive review," *Neurocomputing*, vol. 627, p. 129 588, Apr. 2025.

[6] K. T. P. Nguyen, K. Medjaher, and D. T. Tran, "A review of artificial intelligence methods for engineering prognostics and health management with implementation guidelines," en, *Artificial Intelligence Review*, vol. 56, no. 4, pp. 3659–3709, Apr. 2023.

[7] S. Gawde, S. Patil, S. Kumar, P. Kamat, K. Kotecha, and A. Abraham, "Multi-fault diagnosis of Industrial Rotating Machines using Data-driven approach : A review of two decades of research," *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106 139, Aug. 2023.

[8] O. Das, D. Bagci Das, and D. Birant, "Machine learning for fault analysis in rotating machinery: A comprehensive review," *Heliyon*, vol. 9, no. 6, e17584, Jun. 2023.

[9] O. Serradilla, E. Zugasti, J. Rodriguez, and U. Zurutuza, "Deep learning models for predictive maintenance: A survey, comparison, challenges and prospects," en, *Applied Intelligence*, vol. 52, no. 10, pp. 10 934–10 964, Aug. 2022.

[10] Y. Hu, X. Miao, Y. Si, E. Pan, and E. Zio, "Prognostics and health management: A review from the perspectives of design, development and decision," *Reliability Engineering & System Safety*, vol. 217, p. 108 063, Jan. 2022.

[11] Z. Zhao, Q. Zhang, X. Yu, C. Sun, S. Wang, R. Yan, and X. Chen, "Applications of Unsupervised Deep Transfer Learning to Intelligent Fault Diagnosis: A Survey and Comparative Study," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–28, 2021.

[12] Y. Shi, A. Deng, M. Deng, J. Li, M. Xu, S. Zhang, X. Ding, and S. Xu, "Domain Transferability-Based Deep Domain Generalization Method Towards Actual Fault Diagnosis Scenarios," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 6, pp. 7355–7366, Jun. 2023.

[13] X. Chen, R. Yang, Y. Xue, M. Huang, R. Ferrero, and Z. Wang, "Deep Transfer Learning for Bearing Fault Diagnosis: A Systematic Review Since 2016," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–21, 2023.

[14] A. H. Bonnett and C. Yung, "Increased Efficiency Versus Increased Reliability," *IEEE Industry Applications Magazine*, vol. 14, no. 1, pp. 29–36, Jan. 2008.

[15] K. Choi, J. Yi, C. Park, and S. Yoon, "Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines," *IEEE Access*, vol. 9, pp. 120 043–120 065, 2021.

[16] G. Li and J. J. Jung, "Deep learning for anomaly detection in multivariate time series: Approaches, applications, and challenges," *Information Fusion*, vol. 91, pp. 93–102, Mar. 2023.

[17] L. Correia, J.-C. Goos, P. Klein, T. Bäck, and A. V. Kononova, "Online model-based anomaly detection in multivariate time series: Taxonomy, survey, research challenges and future directions," *Engineering Applications of Artificial Intelligence*, vol. 138, p. 109 323, Dec. 2024.

[18] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, "USAD: UnSupervised Anomaly Detection on Multivariate Time Series," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '20, New York, NY, USA: Association for Computing Machinery, Aug. 2020, pp. 3395–3404.

[19] D. Li, D. Chen, L. Shi, B. Jin, J. Goh, and S.-K. Ng, "MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks," *arXiv:1901.04997 [cs, stat]*, Jan. 2019, arXiv: 1901.04997.

[20] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '19, New York, NY, USA: Association for Computing Machinery, Jul. 2019, pp. 2828–2837.

[21] A. Deng and B. Hooi, "Graph Neural Network-Based Anomaly Detection in Multivariate Time Series," in *AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

[22] H. Kang and P. Kang, "Transformer-based multivariate time series anomaly detection using inter-variable attention mechanism," *Knowledge-Based Systems*, vol. 290, p. 111 507, Apr. 2024.

[23] A. Garg, W. Zhang, J. Samaran, R. Savitha, and C.-S. Foo, "An Evaluation of Anomaly Detection and Diagnosis in Multivariate Time Series," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2021.

[24] M. Zhang, D. Wang, W. Lu, J. Yang, Z. Li, and B. Liang, "A Deep Transfer Model With Wasserstein Distance Guided Multi-Adversarial Networks for Bearing Fault Diagnosis Under Different Working Conditions," *IEEE Access*, vol. 7, pp. 65 303–65 318, 2019.

[25] K. Hundman, V. Constantinou, C. Laporte, I. Colwell, and T. Soderstrom, "Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '18, New York, NY, USA: Association for Computing Machinery, Jul. 2018, pp. 387–395.

[26] I. El-Thalji and E. Jantunen, "A summary of fault modelling and predictive health monitoring of rolling element bearings," en, *Mechanical Systems and Signal Processing*, vol. 60-61, pp. 252–272, Aug. 2015.

[27] D. Neupane and J. Seok, "Bearing Fault Detection and Diagnosis Using Case Western Reserve University Dataset With Deep Learning Approaches: A Review," *IEEE Access*, vol. 8, pp. 93 155–93 178, 2020.

[28] M. Hakim, A. A. B. Omran, A. N. Ahmed, M. Al-Waily, and A. Abdellatif, "A systematic review of rolling bearing fault diagnoses based on deep learning and transfer learning: Taxonomy, overview, application, open challenges, weaknesses and recommendations," en, *Ain Shams Engineering Journal*, p. 101 945, Sep. 2022.

[29] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep Learning Algorithms for Bearing Fault Diagnostics—A Comprehensive Review," *IEEE Access*, vol. 8, pp. 29 857–29 881, 2020.

[30] J. Lee, H. Qiu, J. Lin, and Technical Services, 2007, *Bearing Data Set*, Dataset, Dataset.

[31] *Case Western reserve University (CWRU) Bearing Data Center*, Dataset, Dataset, Aug. 2021.

[32] Patrick Nectoux, Rafael Gouriveau, Kamal Medjaher, Emmanuel Ramasso, Brigitte Chebel-Morello, N. Zerhouni, and C. Varnier, "PRONOSTIA : An experimental platform for bearings accelerated degradation tests," in *IEEE International Conference on Prognostics and Health Management*, Denver, Colorade, United States, Jun. 2012.

[33] C. Lessmeier, J. K. Kimotho, D. Zimmer, and W. Sextro, "Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: A Benchmark Data Set for Data-Driven Classification," en, *PHM Society European Conference*, vol. 3, no. 1, 2016.

[34] *Konstruktions- und Antriebstechnik (KAt) - Bearing DataCenter (Universität Paderborn)*, de, Dataset, Dataset.

[35] N. D. Thuan and H. S. Hong, "HUST bearing: A practical dataset for ball bearing fault diagnosis," *BMC Research Notes*, vol. 16, no. 1, p. 138, Jul. 2023.

[36] Hoang Si Hong and Nguyen Thuan, *HUST bearing*, Dataset, Dataset, 2023.

[37] Society for Machinery Failure Prevention Technology (MFPT), *Fault Data Set*, Dataset, Dataset.

[38] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly Accurate Machine Fault Diagnosis Using Deep Transfer Learning," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.

[39] S. Shao, S. McAleer, R. Yan, and P. Baldi, *Mechanical Dataset*, Dataset, Dataset.

[40] H. Huang and N. Baddour, "Bearing vibration data collected under time-varying rotational speed conditions," en, *Data in Brief*, vol. 21, pp. 1745–1749, Dec. 2018.

[41] H. Huang and N. Baddour, *Bearing Vibration Data under Time-varying Rotational Speed Conditions*, en, Dataset, Dataset, Feb. 2019.

[42] C. Zhao, E. Zio, and W. Shen, "Domain generalization for cross-domain fault diagnosis: An application-oriented perspective and a benchmark study," *Reliability Engineering & System Safety*, vol. 245, p. 109 964, May 2024.

[43] L. Wan, Y. Li, K. Chen, K. Gong, and C. Li, "A novel deep convolution multi-adversarial domain adaptation model for rolling bearing fault diagnosis," *Measurement*, vol. 191, p. 110 752, Mar. 2022.

[44] J. Jiao, M. Zhao, J. Lin, and K. Liang, "Residual joint adaptation adversarial network for intelligent transfer fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 145, p. 106 962, Nov. 2020.

[45] Y. Yang, J. Yin, H. Zheng, Y. Li, M. Xu, and Y. Chen, "Learn Generalization Feature via Convolutional Neural Network: A Fault Diagnosis Scheme Toward Unseen Operating Conditions," *IEEE Access*, vol. 8, pp. 91 103–91 115, 2020.

[46] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Domain generalization in rotating machinery fault diagnostics using deep neural networks," en, *Neurocomputing*, vol. 403, pp. 409–420, Aug. 2020.

[47] H. Zheng, R. Wang, Y. Yang, Y. Li, and M. Xu, "Intelligent Fault Identification Based on Multisource Domain Generalization Towards Actual Diagnosis Scenario," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 2, pp. 1293–1304, Feb. 2020.

[48] Y. Liao, R. Huang, J. Li, Z. Chen, and W. Li, "Deep Semisupervised Domain Generalization Network for Rotary Machinery Fault Diagnosis Under Variable Speed," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 8064–8075, Oct. 2020.

[49] Q. Zhang, Z. Zhao, X. Zhang, Y. Liu, C. Sun, M. Li, S. Wang, and X. Chen, "Conditional Adversarial Domain Generalization With a Single Discriminator for Bearing Fault Diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–15, 2021.

[50] T. Han, Y.-F. Li, and M. Qian, "A Hybrid Generalization Network for Intelligent Fault Diagnosis of Rotating Machinery Under Unseen Working Conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[51] Y. Xie, J. Shi, C. Gao, G. Yang, Z. Zhao, G. Guan, and D. Chen, "Rolling Bearing Fault Diagnosis Method Based On Dual Invariant Feature Domain Generalization," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–11, 2024.

[52] H. Zheng, Y. Yang, J. Yin, Y. Li, R. Wang, and M. Xu, "Deep Domain Generalization Combining A Priori Diagnosis Knowledge Toward Cross-Domain Fault Diagnosis of Rolling Bearing," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[53] L. Jia, T. W. S. Chow, Y. Wang, and J. Ma, "Dynamic Balanced Dual Prototypical Domain Generalization for Cross-Machine Fault Diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–10, 2024.

[54] C. Okoh, R. Roy, J. Mehnen, and L. Redding, "Overview of Remaining Useful Life Prediction Techniques in Through-life Engineering Services," *Procedia CIRP*, Product Services Systems and Value Creation. Proceedings of the 6th CIRP Conference on Industrial Product-Service Systems, vol. 16, pp. 158–163, Jan. 2014.

[55] Y. Ding, P. Ding, X. Zhao, Y. Cao, and M. Jia, "Transfer Learning for Remaining Useful Life Prediction Across Operating Conditions Based on Multisource Domain Adaptation," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 4143–4152, Oct. 2022.

[56] P. Xia, Y. Huang, P. Li, C. Liu, and L. Shi, "Fault Knowledge Transfer Assisted Ensemble Method for Remaining Useful Life Prediction," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1758–1769, Mar. 2022.

[57] Y. Deng, S. Du, D. Wang, Y. Shao, and D. Huang, "A Calibration-Based Hybrid Transfer Learning Framework for RUL Prediction of Rolling Bearing Across Different Machines," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–15, 2023.

[58] W. Mao, J. He, and M. J. Zuo, "Predicting Remaining Useful Life of Rolling Bearings Based on Deep Feature Representation and Transfer Learning," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 4, pp. 1594–1608, Apr. 2020.

[59] N. Li, Y. Lei, J. Lin, and S. X. Ding, "An Improved Exponential Model for Predicting Remaining Useful Life of Rolling Element Bearings," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7762–7773, Dec. 2015.

[60] Z. Xu, M. Bashir, Q. Liu, Z. Miao, X. Wang, J. Wang, and N. Ekere, "A novel health indicator for intelligent prediction of rolling bearing remaining useful life based on unsupervised learning model," *Computers & Industrial Engineering*, vol. 176, p. 108 999, Feb. 2023.

[61] Q. Ni, J. C. Ji, and K. Feng, "Data-Driven Prognostic Scheme for Bearings Based on a Novel Health Indicator and Gated Recurrent Unit Network," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1301–1311, Feb. 2023.

[62] L. Wen, G. Yang, L. Hu, C. Yang, and K. Feng, "A new unsupervised health index estimation method for bearings early fault detection based on Gaussian mixture model," *Engineering Applications of Artificial Intelligence*, vol. 128, p. 107 562, Feb. 2024.

[63] M. M. Manjurul Islam, A. E. Prosvirin, and J.-M. Kim, "Data-driven prognostic scheme for rolling-element bearings using a new health index and variants of least-square support vector machines," *Mechanical Systems and Signal Processing*, vol. 160, p. 107 853, Nov. 2021.

[64] X. Pei, L. Gao, and X. Li, “Remaining useful life prediction of machinery based on performance evaluation and online cross-domain health indicator under unknown working conditions,” *Journal of Manufacturing Systems*, vol. 75, pp. 213–227, Aug. 2024.

[65] Y. Han, M. Xu, X. Sun, X. Ding, X. Chen, and F. Gu, “Gear Health Monitoring and RUL Prediction Based on MSB Analysis,” *IEEE Sensors Journal*, vol. 22, no. 5, pp. 4400–4409, Mar. 2022.

[66] P. Kumar, I. Raouf, and H. S. Kim, “Review on prognostics and health management in smart factory: From conventional to deep learning perspectives,” *Engineering Applications of Artificial Intelligence*, vol. 126, p. 107 126, Nov. 2023.

[67] J. Chen, R. Huang, Z. Chen, W. Mao, and W. Li, “Transfer learning algorithms for bearing remaining useful life prediction: A comprehensive review from an industrial application perspective,” en, *Mechanical Systems and Signal Processing*, vol. 193, p. 110 239, Jun. 2023.

[68] J. Zhou, J. Yang, and Y. Qin, “A systematic overview of health indicator construction methods for rotating machinery,” *Engineering Applications of Artificial Intelligence*, vol. 138, p. 109 356, Dec. 2024.

[69] W. Mao, J. Liu, J. Chen, and X. Liang, “An Interpretable Deep Transfer Learning-Based Remaining Useful Life Prediction Approach for Bearings With Selective Degradation Knowledge Fusion,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–16, 2022.

[70] H. Cheng, X. Kong, Q. Wang, H. Ma, S. Yang, and G. Chen, “Deep transfer learning based on dynamic domain adaptation for remaining useful life prediction under different working conditions,” en, *Journal of Intelligent Manufacturing*, vol. 34, no. 2, pp. 587–613, Feb. 2023.

[71] N. Ding, H. Li, Z. Yin, and F. Jiang, “A novel method for journal bearing degradation evaluation and remaining useful life prediction under different working conditions,” *Measurement*, vol. 177, p. 109 273, Jun. 2021.

[72] S. Siahpour, X. Li, and J. Lee, “A Novel Transfer Learning Approach in Remaining Useful Life Prediction for Incomplete Dataset,” *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[73] J. Zhu, N. Chen, and C. Shen, “A new data-driven transferable remaining useful life prediction approach for bearing under different working conditions,” *Mechanical Systems and Signal Processing*, vol. 139, p. 106 602, May 2020.

[74] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, *Domain-Adversarial Training of Neural Networks*, arXiv:1505.07818 [cs, stat], May 2016.

[75] S. Dong, J. Xiao, X. Hu, N. Fang, L. Liu, and J. Yao, "Deep transfer learning based on Bi-LSTM and attention for remaining useful life prediction of rolling bearing," *Reliability Engineering & System Safety*, vol. 230, p. 108 914, Feb. 2023.

[76] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Transfer learning using deep representation regularization in remaining useful life prediction across operating conditions," *Reliability Engineering & System Safety*, vol. 211, p. 107 556, Jul. 2021.

[77] W. Mao, W. Zhang, K. Feng, M. Beer, and C. Yang, "Tensor representation-based transferability analytics and selective transfer learning of prognostic knowledge for remaining useful life prediction across machines," *Reliability Engineering & System Safety*, vol. 242, p. 109 695, Feb. 2024.

[78] T. Hu, Y. Guo, L. Gu, Y. Zhou, Z. Zhang, and Z. Zhou, "Remaining useful life estimation of bearings under different working conditions via Wasserstein distance-based weighted domain adaptation," *Reliability Engineering & System Safety*, vol. 224, p. 108 526, Aug. 2022.

[79] B. Wang, Y. Lei, N. Li, and N. Li, "A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings," *IEEE Transactions on Reliability*, vol. 69, no. 1, pp. 401–412, Mar. 2020.

[80] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, "PRONOSTIA : An experimental platform for bearings accelerated degradation tests.," in *IEEE International Conference on Prognostics and Health Management, PHM'12.*, vol. sur CD ROM, Denver, Colorado, United States: IEEE Catalog Number : CPF12PHM-CDR, Jun. 2012, pp. 1–8.

[81] W. Gousseau, J. Antoni, F. Girardin, and J. Griffaton, "Analysis of the Rolling Element Bearing data set of the Center for Intelligent Maintenance Systems of the University of Cincinnati," in *CM2016*, Charenton, France, Oct. 2016.

[82] D. Park, Y. Hoshi, and C. C. Kemp, "A Multimodal Anomaly Detector for Robot-Assisted Feeding Using an LSTM-Based Variational Autoencoder," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1544–1551, Jul. 2018.

[83] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly Detection," en, Feb. 2018.

[84] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the Support of a High-Dimensional Distribution," en, *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

[85] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, “LOF: Identifying density-based local outliers,” *ACM SIGMOD Record*, vol. 29, no. 2, pp. 93–104, 2000.

[86] F. T. Liu, K. M. Ting, and Z.-H. Zhou, “Isolation Forest,” in *2008 Eighth IEEE International Conference on Data Mining*, ISSN: 2374-8486, Dec. 2008, pp. 413–422.

[87] A. A. Patel, *Hands-On Unsupervised Learning Using Python*, en. O’Reilly Media, Inc., 2019.

[88] T. Rojat, R. Puget, D. Filliat, J. Del Ser, R. Gelin, and N. Díaz-Rodríguez, *Explainable Artificial Intelligence (XAI) on TimeSeries Data: A Survey*, arXiv:2104.00950 [cs], Apr. 2021.

[89] J. Höllig, C. Kulbach, and S. Thoma, “TSInterpret: A Python Package for the Interpretability of Time Series Classification,” en, *Journal of Open Source Software*, vol. 8, no. 85, p. 5220, May 2023.

[90] M. Miao and J. Yu, “A Deep Domain Adaptative Network for Remaining Useful Life Prediction of Machines Under Different Working Conditions and Fault Modes,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.

[91] X. Li, Q. Ding, and J.-Q. Sun, “Remaining useful life estimation in prognostics using deep convolution neural networks,” *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, Apr. 2018.

[92] Y. Zhang, P. Hutchinson, N. A. J. Lieven, and J. Nunez-Yanez, “Remaining Useful Life Estimation Using Long Short-Term Memory Neural Networks and Deep Fusion,” *IEEE Access*, vol. 8, pp. 19 033–19 045, 2020.

[93] A. Zhang, H. Wang, S. Li, Y. Cui, Z. Liu, G. Yang, and J. Hu, “Transfer Learning with Deep Recurrent Neural Networks for Remaining Useful Life Estimation,” en, *Applied Sciences*, vol. 8, no. 12, p. 2416, Dec. 2018.

[94] A. T. Keleko, B. Kamsu-Foguem, R. H. Ngouna, and A. Tongne, “Artificial intelligence and real-time predictive maintenance in industry 4.0: A bibliometric analysis,” en, *AI and Ethics*, vol. 2, no. 4, pp. 553–577, Nov. 2022.

[95] W. A. Smith and R. B. Randall, “Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study,” en, *Mechanical Systems and Signal Processing*, vol. 64-65, pp. 100–131, Dec. 2015.

[96] A. Lundström, M. O’nils, F. Z. Qureshi, and A. Jantsch, “Improving Deep Learning Based Anomaly Detection on Multivariate Time Series Through Separated Anomaly Scoring,” *IEEE Access*, vol. 10, pp. 108 194–108 204, 2022.

[97] A. Lundström, M. O'Nils, and F. Z. Qureshi, "An Interactive Threshold-Setting Procedure for Improved Multivariate Anomaly Detection in Time Series," *IEEE Access*, vol. 11, pp. 93 898–93 907, 2023.

[98] A. Lundström and M. O'Nils, "Factory-Based Vibration Data for Bearing-Fault Detection," en, *Data*, vol. 8, no. 7, p. 115, Jul. 2023.

[99] Adam Lundström, Mattias O'Nils, and Faisal Z. Qureshi, "Contextual Knowledge-Informed Deep Domain Generalization for Bearing Fault Diagnosis," *IEEE Access*, vol. 12, pp. 196 842–196 854, 2024.

[100] A. Lundström, M. O'Nils, and F. Z. Qureshi, "Towards Practically Applicable Transfer Learning Methods for Remaining Useful Life Prediction of Bearings," in *2024 IEEE 22nd International Conference on Industrial Informatics (INDIN)*, Beijing, China: IEEE, Aug. 2024, pp. 1–8.

[101] Adam Lundström, Mattias O'Nils, and Faisal Z. Qureshi, "A prognostic framework for rotating machines considering multi-component fault scenarios," *Submitted to IEEE Access*, 2025.